

Cours de Mathématiques

ISA BTP, 3^e année

12 février 2013

Table des matières

1	Systèmes linéaires	5
	Introduction	5
1.0	Calcul Matriciel	6
1.0.1	Définition	6
1.0.2	Opérations matricielles	6
1.0.3	Propriétés de la multiplication	7
1.0.4	Matrice identité, matrices inversibles	7
1.0.5	Transposée et matrices symétriques	8
1.0.6	Déterminant	9
1.0.7	Exercices	11
1.1	Généralités	12
1.1.1	Définitions	12
1.1.2	Équations linéaires et hyperplans	13
1.1.3	Équations et dépendance linéaire	16
1.1.4	Dimension de l'ensemble des solutions	18
1.1.5	Résolution d'un système linéaire	18
1.2	Le pivot de Gauss	22
1.2.1	Le principe du pivot	22
1.2.2	L'algorithme	24
1.2.3	Pivot de Gauss et opérations matricielles	30
1.2.4	Applications du Pivot de Gauss et de la décomposition LU	34
1.3	Systèmes sur déterminés	36
1.3.1	Un exemple	37
1.3.2	Interprétation géométrique	38
1.3.3	Les équations normales	39
1.3.4	Application : la régression	39
2	Fonctions spéciales	45
	Introduction	45
2.1	Intégrales généralisées	46
2.1.1	Intervalles $[a, +\infty[$	46
2.1.2	Intervalles $]a, b]$	47
2.1.3	Propriétés	48

2.1.4	Fonctions de référence	48
2.1.5	Critères de convergence	49
2.2	Fonctions spéciales	52
2.2.1	Intégrales à paramètres	52
2.2.2	La fonction Γ	53
2.2.3	La fonction Beta	56

Chapitre 1

Systemes linéaires

Introduction

De nombreux problèmes (physiques et mathématiques) peuvent se traduire sous forme d'équations linéaires ou de familles d'équations linéaires. Or les matrices et le calcul matriciel donnent un outil très efficace pour représenter ces types de problèmes et plus généralement tout problème ayant un caractère linéaire :

- des familles d'équations ou inéquations linéaires,
- des familles de vecteurs (exprimés dans une base),
- des applications linéaires,
- des équations différentielles...

L'objectif de ce cours est de présenter des méthodes algorithmiques de résolution de systèmes linéaires, rendues possibles par la représentation matricielle.

Après une étude théorique des systèmes linéaires et de leurs solutions, nous verrons comment, en traduisant la résolution d'un système en termes d'opérations matricielles, on peut élaborer ces méthodes algorithmiques.

Nous étudierons essentiellement une méthode : le pivot de Gauss ; dont l'idée générale est de construire, à partir du système de départ, un système équivalent (i.e. qui admet les mêmes solutions), plus simple à résoudre. Dans le cas du pivot de Gauss, l'objectif est d'obtenir un système *triangulaire*.

Dans un second temps, nous étudierons quelques exemples de problèmes pouvant se ramener à des systèmes linéaires. Nous verrons ainsi comment résoudre des systèmes linéaires comportant plus de contraintes que de degrés de liberté.

Nous laissons au lecteur la lecture du chapitre 0 portant sur le calcul matriciel.

1.0 Calcul matriciel

1.0.1 Définition

Une matrice est un tableau de chiffres. Les éléments d'une matrice sont appelés coefficients. C'est la position d'un coefficient qui détermine son rôle dans le système modélisé.

Soit M une matrice. Les caractéristiques fondamentales de M sont :

- son nombre de lignes (traditionnellement noté m),
- son nombre de colonnes (traditionnellement noté n).

La position d'un coefficient est donc donnée par le numéro de la ligne et celui de la colonne auxquelles il appartient. De façon générale, on le note a_{ij} , i étant le numéro de la ligne et j celui de la colonne.

Sous sa forme générale, une matrice A dont les coefficients sont a_{ij} se note

$$A = (a_{ij}).$$

D'autre part, la première ligne est décrite par

$$L_1 = \{a_{1j}, j = 1..n\}.$$

De même, la première colonne est donnée par

$$C_1 = \{a_{i1}, i = 1..m\}.$$

Dans le cas où $m = n$, on dit que la matrice est carrée. Elle possède alors une diagonale dont les coefficients sont données par

$$D = \{a_{jj}, j = 1..n\}.$$

De même, les éléments qui sont au dessus de la diagonale peuvent être représentés par

$$D^+ = \{a_{ij} / j \geq i\}.$$

1.0.2 Opérations matricielles

Si l'on considère les matrices dans leur ensemble, on peut définir des opérations sur les matrices comme on en connaît sur les nombres. On peut par exemple additionner deux matrices à condition qu'elles soient de même taille. Le résultat de cette addition est la matrice dont les coefficients sont donnés par la somme des coefficients de chacune des matrices : si $A = (a_{ij})$ et $B = (b_{ij})$ sont deux matrices de taille $m \times n$, la somme $A + B$ est la matrice $C = (c_{ij})$ de taille $m \times n$ telle que

$$c_{ij} = a_{ij} + b_{ij}.$$

On peut également définir une multiplication matricielle mais là encore on rencontre des problèmes de taille : si $A = (a_{ij})$ et $B = (b_{ij})$ sont deux matrices telles que le nombre de

colonnes de A est égal au nombre de lignes de B , le produit $A \times B$ est la matrice $C = (c_{ij})$ dont les coefficients sont

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.$$

C'est une matrice dont le nombre de lignes est celui de A et dont le nombre de colonnes est celui de B .

Exemple : la multiplication matricielle est l'opération qui permet par exemple de représenter un système linéaire sous forme matricielle.

Note : dans la plupart des exemples que l'on verra, on travaillera avec des matrices carrées. On n'aura donc pas à se soucier de ces problèmes de taille.

1.0.3 Propriétés de la multiplication

La multiplication des matrices est bien une multiplication. Elle est en particulier associative et distributive sur l'addition (des matrices). On va donc pouvoir faire des calculs dans $\mathcal{M}_n(\mathbb{R})$ comme on les fait dans \mathbb{R} . Cependant, la première spécificité du produit de matrices est son caractère non commutatif : pour A et B quelconque, on a en général

$$A \times B \neq B \times A.$$

Cela n'a l'air de rien, mais ça nous oblige à être attentifs lors de nos calculs. Par exemple, on peut élever une matrice A à une certaine puissance p en multipliant A par elle-même p fois, mais comme la multiplication n'est pas commutative, on n'a plus la formule $(ab)^p = a^p b^p$.

1.0.4 Matrice identité, matrices inversibles

Parmi toutes les matrices carrées d'ordre n , il en est une à distinguer : la matrice identité d'ordre n , notée I_n dont tous les éléments diagonaux valent 1 et dont tous les autres sont nuls. C'est ce que l'on appelle l'élément neutre de $\mathcal{M}_n(\mathbb{R})$ pour la multiplication. Autrement dit, pour toutes matrice $A \in \mathcal{M}_n(\mathbb{R})$, on a

$$A \times I_n = I_n \times A = A.$$

(Cela correspond au 1 pour la multiplication des réels).

Cette matrice inverse nous permet de définir la notion de matrice inversible. Précisément, une matrice A carrée d'ordre n sera dite inversible s'il existe une matrice A' (de même taille) telle que

$$A \times A' = I_n.$$

La matrice A' est alors appelée inverse de A et notée A^{-1} .

Remarques :

- Le produit de matrices n'est pas commutatif, mais dans le cas particulier de deux matrices inverses l'une de l'autre, on a

$$A \times A^{-1} = A^{-1} \times A = I_n.$$

- On peut montrer que si A et B sont deux matrices inversibles, alors le produit AB est encore inversible. Attention, comme le produit de matrices n'est pas commutatif, on a

$$(AB)^{-1} = B^{-1}A^{-1}.$$

En effet, si l'on fait $(AB).(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I_n$. C'est bien l'inverse de AB .

Attention : contrairement aux réels, toutes les matrices de $\mathcal{M}_n(\mathbb{R})$ ne sont pas inversibles.

Si l'on veut trouver l'inverse d'une matrice A , il faut donc s'assurer que cette matrice est bien inversible. Une première façon (naïve) de montrer que A est inversible est de trouver son inverse A^{-1} . En notant x_{ij} les coefficients de A' , on peut traduire l'équation

$$A \times A' = I_n$$

en un système matriciel. La résolution de ce système nous donne la réponse :

- si ce système n'a pas de solution, la matrice A n'est pas inversible,
- si ce système admet une solution, la matrice A est inversible et la solution du système nous donne l'inverse.

En pratique, cette méthode est très fastidieuse car le système à résoudre est un système $n^2 \times n^2$.

Une autre façon de calculer l'inverse d'une matrice (et donc de déterminer si elle est inversible) est de résoudre le système linéaire associé. Si on trouve une solution, c'est que la matrice est inversible et la solution que l'on a trouvé nous donne A^{-1} .

Enfin, on a un outil très efficace permettant de déterminer si une matrice donnée est inversible, sans calculer son inverse : le déterminant (que l'on verra plus loin en détails).

Dans le cas des matrices 2×2 , il est cependant possible de déterminer rapidement si une matrice est inversible ou non. En effet, une matrice 2×2 est inversible si et seulement si ses deux lignes (ou ses deux colonnes) ne sont pas proportionnelles.

Dans ce cas, il est alors possible de déterminer la matrice inverse en posant le problème sous la forme d'un système 4×4 . Cependant, cette méthode est à proscrire pour les matrices de tailles supérieures.

1.0.5 Transposée et matrices symétriques

Étant donnée une matrice A (de taille quelconque), on appelle *transposée de A* la matrice tA obtenue en échangeant les lignes et les colonnes de A .

Exemples :

- Si $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, on a ${}^tA = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$.
- Si $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, on a ${}^tA = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

Dans le cas particulier où A est symétrique, on voit que la transposée de A s'obtient en effectuant, sur les coefficients de A , une symétrie par rapport à la diagonale. On appelle alors matrice symétrique toute matrice carrée invariante par cette transformation. Autrement dit, une matrice A est symétrique si et seulement si elle égale à sa transposée.

1.0.6 Déterminant

L'outil principal permettant de déterminer si une matrice est inversible est le déterminant. Le déterminant d'une matrice A (noté $\det(A)$) est un nombre calculé à partir des coefficients de A . Il est calculé de telle sorte que l'on a

$$A \text{ est inversible} \iff \det(A) \neq 0.$$

Formellement, le déterminant d'une matrice permet (entre autre) de mesurer la dépendance linéaire des lignes (ou des colonnes) d'une matrice. Ainsi, si A représente une famille de vecteurs (dont les coordonnées sont mises en ligne ou en colonne), le déterminant de A permet de déterminer si ces vecteurs sont linéairement indépendants ou non. Cela donne en particulier une indication sur la taille de l'espace que cette famille engendre.

Intuitivement, le déterminant de A permet de savoir si chaque vecteur apporte une direction supplémentaire. Ainsi, dans \mathbb{R}^3 , si l'on se donne trois vecteurs $\vec{v}_1 = (x_1, y_1, z_1)$, $\vec{v}_2 = (x_2, y_2, z_2)$ et $\vec{v}_3 = (x_3, y_3, z_3)$ et si l'on note

$$A = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix}$$

on a

- si $\det(A) \neq 0$, alors les trois vecteurs sont indépendants et forment une base de l'espace,
- si $\det(A) = 0$, alors les trois vecteurs sont liés. Ils sont donc tous dans un même plan, voire sur une même droite.

Dans la suite de ce cours, on verra qu'il est toujours possible de représenter un système linéaire par une matrice dont les lignes correspondent aux équations. On verra alors que dans ce cas, le caractère libre ou non de la famille des lignes de A est une caractéristique importante du système que l'on étudie.

Question : Comment calculer le déterminant d'une matrice donnée ?

Dans le cas d'une matrice 2, on a

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Notons qu'ainsi définit, le déterminant d'une matrice 2×2 répond bien à la question de dépendance linéaire. En effet, deux vecteurs (ici, (a, b) et (c, d)) sont liés si et seulement si ils sont colinéaires. Autrement dit, ils sont liés si et seulement si

$$\frac{a}{c} = \frac{b}{d},$$

ou encore si et seulement si

$$ad - bc = 0.$$

Pour calculer un déterminant en dimension quelconque, on se ramène au cas $n = 2$ en développant par rapport à une ligne ou une colonne. On peut en effet montrer (pour $n = 3$ par exemple) que

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}$$

On se ramène ainsi au calcul de 3 déterminants 2×2 .

On peut ainsi développer par rapport à la ligne ou la colonne que l'on veut. Notons tout de même l'alternance des signes \pm . Ce signe est donné par la parité de la somme

$$n^\circ \text{ de ligne} + n^\circ \text{ de colonne}.$$

En particulier, on remarque que si A est une matrice triangulaire, son déterminant est donné par le produit des éléments diagonaux.

D'autre part, on peut montrer que le déterminant d'une matrice n'est pas modifié si l'on ajoute une ligne à une autre. Il est alors possible d'effectuer certaines de ces opérations pour remplacer certains coefficients par des 0, par exemple dans une colonne. Cela permet alors de simplifier les calculs lorsque l'on développe par rapport à cette colonne.

Enfin, notons que dans le cas $n = 3$, on a également la méthode dite de Sarrus qui donne

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_2 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - (a_3 b_2 c_1 + a_2 b_1 c_3 + a_1 b_3 c_2).$$

1.0.7 Exercices

Exercice 1 Soient

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 2 \\ -1 & -2 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$E = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad F = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

1. Calculer les produits suivants :

$$\begin{array}{ccccc} A \times B & B \times A & A \times E & E \times A & A^2 \\ B^2 & C \times D & D \times C & I \times E & F \times B \end{array}$$

2. Parmi les matrices ci-dessus, déterminer celles qui sont inversibles.

3. Calculer l'inverse des matrices inversible déterminées à la question précédente.

Exercice 2 Soient

$$A = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad E = \begin{pmatrix} 3 & 0 & -1 \\ 1 & 0 & -\frac{1}{3} \\ -1 & 1 & -1 \end{pmatrix} \quad I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

1. Calculer les produits suivants :

$$\begin{array}{cccccc} A \times E & B \times E & E \times B & C^2 & C^3 \\ C \times D & D \times C & I \times E & & \end{array}$$

2. Commenter les résultats obtenus.

Exercice 3 Reprendre les matrices de l'exercice 1 et déterminer celles qui sont inversibles à l'aide du déterminant.

- Exercice 4**
1. Reprendre les matrices de l'exercice 2 et déterminer celles qui sont inversibles à l'aide du déterminant.
 2. Calculer l'inverse des matrices inversibles déterminées à la question précédente, à l'aide de la méthode suivante :
 - Créer deux colonnes
 - Dans la première, placer la matrice à inverser.
 - Dans la seconde, placer la matrice identité.
 - Effectuer les combinaisons de lignes nécessaires sur la matrice à inverser pour la transformer en la matrice identité.
 - Effectuer en parallèles les mêmes opérations de lignes sur la matrice identité.
 - Vérifier par le calcul qu'à la fin du procédé, la seconde colonne contient la matrice inverse cherchée.

1.1 Généralités

1.1.1 Définitions

Une équation linéaire est une équation de la forme

$$(E) : a_1x_1 + a_2x_2 + \dots + a_nx_n = b$$

liant les inconnues x_1, \dots, x_n .

Les a_i sont les coefficients (connus) de l'équation et b est le second membre. Si b est nul, l'équation est dite *homogène*.

Un système linéaire est la donnée d'un ensemble d'équations linéaires portant sur les inconnues x_1, \dots, x_n :

$$(S) : \begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m \end{cases}$$

Ici, n est le nombre d'inconnues et m est le nombre d'équations.

Les coefficients a_{ij} sont les coefficients du système, les x_i sont les inconnues et les b_i forment le second membre. S'ils sont tous nuls, le système est dit homogène.

Une *solution* du système (S) est un n -uplet (x_1, \dots, x_n) satisfaisant l'ensemble des équations. C'est un élément de l'ensemble \mathbb{R}^n .

Résoudre (S) , c'est déterminer l'ensemble des n -uplets solutions. L'ensemble des solutions d'un système linéaire à n inconnues est donc un sous ensemble de \mathbb{R}^n . Dans les cas $n = 2$ et $n = 3$, on pourra représenter cet ensemble comme une partie du plan ou de l'espace (muni d'un repère).

D'autre part, tout système linéaire admet une représentation matricielle : en notant $A = (a_{ij})$ la matrice des coefficients, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ le vecteur inconnu et $B = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ le vecteur second membre, on a

$$(S) \iff \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \times \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \iff AX = B$$

Chaque ligne de A correspond alors à une équation du système et chaque colonne correspond à une variable.

Enfin, on pourra parfois ajouter à A le second membre B . La matrice

$$A = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right)$$

est alors *la matrice étendue* du système $AX = B$.

L'utilisation des matrices permet d'utiliser tous les outils de calcul matriciel pour étudier et résoudre les systèmes linéaires. La représentation sous forme de tableau permet en particulier de représenter et manipuler les systèmes linéaires à l'aide d'un ordinateur, toute manipulation dans (S) pouvant se traduire par une manipulation des lignes de la matrice A associée.

1.1.2 Équations linéaires et hyperplans

Droites du plan. Une équation linéaire en deux variables est une équation de la forme

$$ax + by = c \quad (*)$$

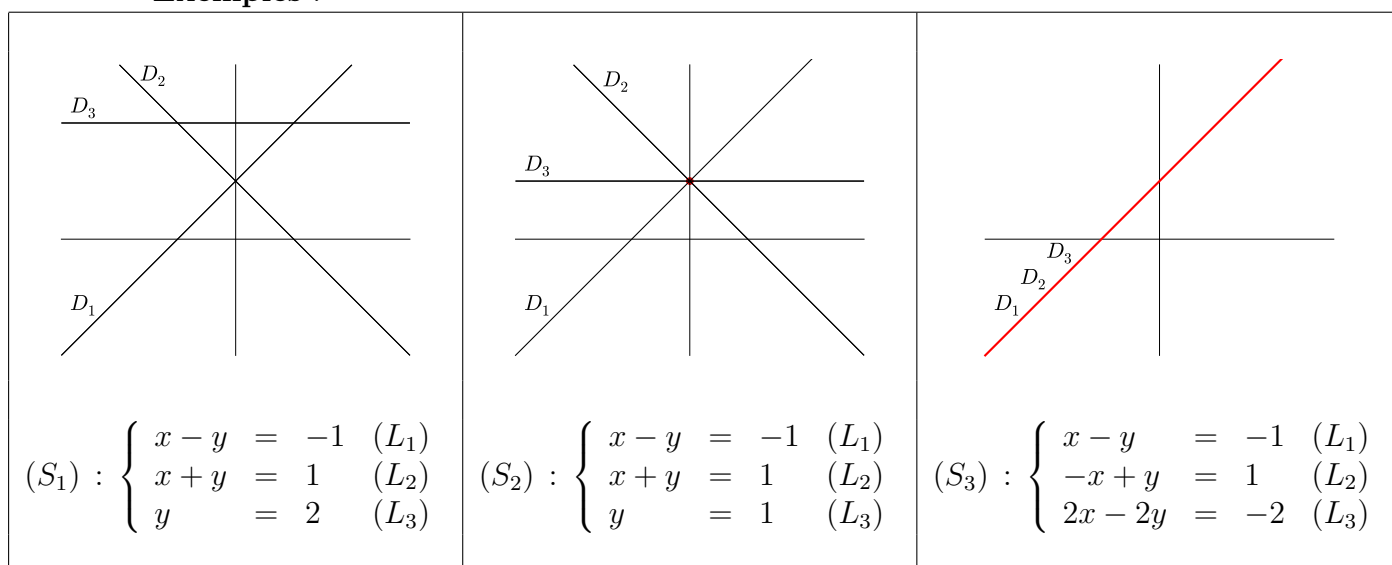
où a, b, c sont des réels fixés.

L'ensemble des solutions d'une telle équation est l'ensemble des couples (x, y) vérifiant cette équation.

Dans le plan muni d'un repère (orthonormé), l'ensemble de ces solutions correspond à *une droite*.

Un *système d'équations* à deux inconnues correspond à un ensemble d'équations de la forme (*). Dans le plan, cela correspond donc à une famille de droites (une par équation). L'ensemble des solutions d'un tel système est correspond alors à l'intersection de toutes ces droites.

Exemples :



Ainsi, en dimension 2, l'intersection de droites peut être

- l'ensemble vide (système (S_1)),
- un point (système (S_2)),
- une droite (système (S_3)).

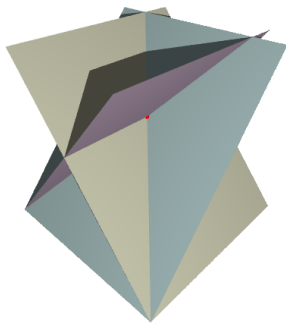
Note : une équation homogène en deux variables (i.e. de la forme $ax + by = 0$) correspond à une droite passant par l'origine $(0, 0)$ du repère. Ainsi, dans le cas d'un système homogène, chacune des droites passent par l'origine. Le cas du système (S_1) est alors exclu.

Plans de l'espace. En dimension 3, une équation linéaire est une équation de la forme

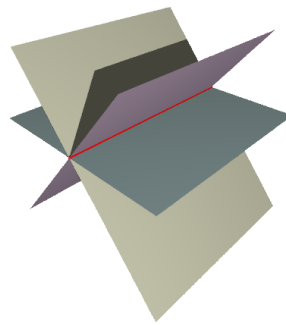
$$ax + by + cz = d$$

Dans l'espace muni d'un repère, l'ensemble des triplets (x, y, z) qui satisfont une telle équation correspondent à un plan.

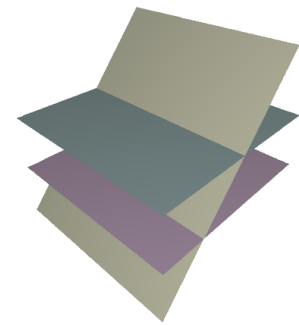
Un système linéaire à trois inconnues correspond donc à l'intersection de plans dans l'espace :



Une solution



Une droite de solutions



Pas de solution

Hyperplans de \mathbb{R}^n . De façon générale, l'ensemble \mathcal{S} des solutions d'une équation linéaire à n variables de la forme

$$(E) : a_1x_1 + \dots + a_nx_n = b$$

est un sous ensemble de \mathbb{R}^n . Si (E) est homogène (i.e. $b = 0$), \mathcal{S} est *un sous espace vectoriel de \mathbb{R}^n , de dimension $n - 1$* .

Si $n = 2$, il s'agit des droites passant par l'origine. Si $n = 3$, il s'agit des plans passant par l'origine.

Note : si $b \neq 0$, l'ensemble des solutions est *un espace affine*. En dimension 2 ou 3, on retrouve les droites et plans, mais ils ne passent plus par l'origine. Ils ont été translatés.

L'ensemble des solutions d'un système (S) d'équations linéaires homogène est donc une intersection d'hyperplans. On peut montrer (en s'inspirant notamment des différents cas vus plus haut) qu'une telle intersection est encore un sous espace vectoriel de \mathbb{R}^n . Selon les liens qu'il existe entre les différentes équations du système, la dimension de cette intersection peut varier.

Dans le cas général, si les équations sont totalement indépendantes, chacune équation du système correspond à un hyperplan indépendant.

Exemple : en dimension 2, un système

$$\begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases}$$

de deux équations indépendantes correspondent à deux droites non parallèles. L'intersection de ces deux droites est donc un point. Autrement dit, un système de deux équations à deux inconnues dont les équations sont indépendantes admet une et une seule solution.

Si l'on ajoute une troisième équation à deux variables,

$$\begin{cases} a_{11}x + a_{12}y = b_1 & (1) \\ a_{21}x + a_{22}y = b_2 & (2) \\ a_{31}x + a_{32}y = b_3 & (3) \end{cases}$$

l'intersection peut être vides (c.f. système (S_1)) ou réduite à un point (c.f. (S_2)). Si c'est un point, c'est le même que pour le système à deux équations.

Du point de vue des équations, cela signifie que, soit le système n'admet pas de solution (il est surdéterminé), soit l'ajout d'une troisième équation ne change pas l'ensemble des solutions. Dans ce cas, l'équation (3) n'est pas indépendante des deux équations précédentes. *Elle n'apporte pas de contrainte supplémentaire au système.*

Exercice :

1. Déterminer les différents types d'intersections d'hyperplans qu'il existe dans l'espace à 3 dimensions.
2. En trois variables, quel est le nombre minimal d'équations qu'il faut pour avoir une unique solution ?

1.1.3 Équations et dépendance linéaire

La notion de dépendance évoquée plus haut peut se traduire en termes de relations entre les d'équations.

Ainsi, si le système ne contient qu'une équation

$$(E) : a_1x_1 + \dots + a_nx_n = b$$

les équations liées à (E) sont les équations proportionnelles

$$\lambda.(E) : (\lambda.a_1)x_1 + \dots + (\lambda.a_n)x_n = \lambda.b, \quad \lambda \in \mathbb{R}.$$

Toutes ces équations correspondent au même hyperplan (une même droite dans le plan, un même plan dans l'espace, etc).

Si le système contient deux équations indépendantes (i.e. non proportionnelles),

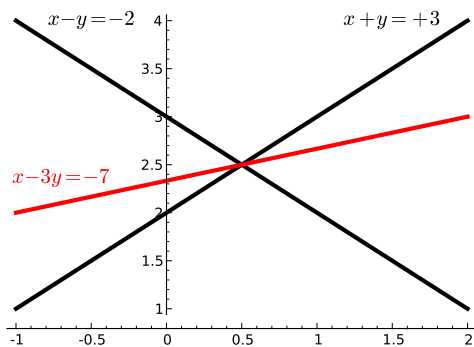
$$(S) : \begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 & (E_1) \\ a_{21}x_1 + \dots + a_{2n}x_n = b_2 & (E_2) \end{cases}$$

les équations qui sont proportionnelles à l'une ou l'autre de ces équations est encore liée au système. Mais il en existe d'autres : toutes les équations de la forme

$$(E) = \alpha.(E_1) + \beta.(E_2), \quad \alpha, \beta \in \mathbb{R}.$$

En dimension 2, le système (S) correspond à deux droites sécantes en un point P . Une équation de la forme (E) correspond alors à une droite du plan passant par P .

En dimension 3, le système (S) correspond à deux plans se coupant le long d'une droite D . Une équation de la forme (E) correspond alors à un plan contenant également la droite D .



$$(E_3) = 2.(E_1) - (E_2)$$

Enfin, de façon générale, si (S) est un système à m équations $\{(E_1), \dots, (E_m)\}$, les équations liées à (S) sont toutes les équations de la forme

$$(E) = \alpha_1.(E_1) + \dots + \alpha_n.(E_n) = \sum_{k=1}^n \alpha_k.(E_k), \quad \alpha_k \in \mathbb{R}.$$

Il s'agit de toutes les *combinaisons linéaires* que l'on peut faire à partir des équations (E_k) .

Ainsi, si une équation est *liée* à un système d'équations, le fait de l'ajouter au système ne change pas l'ensemble des solutions.

De façon générale, un système est dit *lié* si l'une de ses équations est liée aux autres (i.e. si elle peut s'écrire en fonction des autres). Elle est donc "inutile" et en la supprimant, on ne change pas l'ensemble des solutions.

En supprimant successivement les équations inutiles d'un système, on obtient un système dans lequel chaque équation correspond à un hyperplan indépendant et amène ainsi une contrainte supplémentaire. Un tel système est dit *libre*.

On appelle *rang d'un système linéaire* le nombre de contraintes qu'il contient. Il s'agit du nombre d'équations qu'il reste à la fin du tri évoqué plus haut.

Or le rang d'un système permet de déterminer la taille de l'ensemble de ses solutions.

1.1.4 Dimension de l'ensemble des solutions

Comme on l'a vu plus haut, l'ensemble des solutions d'un système est l'intersection des hyperplans correspondant à chacune des équations.

Dans le plan, il s'agit donc soit d'un point, soit d'une droite. Dans l'espace, il s'agit d'un point, d'une droite et d'un plan. (On laissera pour l'instant de côté les cas où le système n'admet pas de solutions).

De façon générale, l'ensemble des solutions est un sous espace (affine ou vectoriel) de \mathbb{R}^n .

Si cet ensemble est un point, le système admet une unique solution. Dans tout autre cas, le système admet une infinité de solutions. La taille de cet ensemble est alors mesuré par sa *dimension* (1 pour une droite, 2 pour un plan, 0 pour un point, $n - 1$ pour un hyperplan...).

On peut montrer que pour tout système linéaire en n variables contenant au plus n équations, le rang r et la dimension d de l'ensemble des solutions vérifient

$$d = n - r$$

Dans un système libre, le rang étant donné par le nombre d'équations, la dimension de l'ensemble des solutions s'obtient en soustrayant le nombre d'équations au nombre d'inconnues. Autrement dit, dans un système libre, chaque équation (i.e. chaque contrainte) fait perdre une dimension à l'ensemble des solutions (i.e. *un degré de liberté*).

Dans la suite, on considèrera donc uniquement des système pour lesquels le nombre d'équations est au plus égale au nombre d'inconnues.

1.1.5 Résolution d'un système linéaire

Le principe d'élimination

L'étude d'un système linéaire (S) donné a pour objectif la résolution de celui-ci, i.e. déterminer l'ensemble de ces solutions.

L'étude préalable de (S) au sens vu plus haut permet de déterminer le type de solutions que l'on cherche. On distinguera ainsi les systèmes admettant une unique solution (i.e. inversibles) et les systèmes admettant une infinité de solution (dits sous-déterminés).

Cependant, dans tous les cas, la résolution d'un système donné passe par une phase d'élimination. Le principe est de faire disparaître un maximum d'inconnues dans un maximum d'équations.

Il existe essentiellement deux méthodes d'élimination :

- La méthode de *substitution* consiste à utiliser l'une des équations du système pour exprimer une inconnue x_i en fonction des autres. En remplaçant dans les autres équations la variable x_i par l'expression obtenue, on obtient un sous-système dans lequel l'inconnue x_i a disparu. En répétant l'opération sur l'une des variables présente dans le sous système, on peut ainsi faire disparaître un maximum d'inconnues dans un maximum d'équations. Bien que théoriquement valable, la méthode de substitution est souvent la pire des solutions à cause de la complexité des calculs qu'elle implique.
- La méthode de *combinaison* consiste à exploiter la notion d'équations liées pour remplacer, dans un système donné, une équation par une autre, contenant une inconnue de moins, tout en conservant l'ensemble des solutions du système global. Précisément, il est clair d'après ce que l'on a vu, que si l'on remplace une équation (E_i) par une combinaison de la forme $(E_i) + \lambda(E_j)$, on ne change pas l'ensemble des solutions du système.

Exemple : soit

$$(S) : \begin{cases} x - y = 1 & (E_1) \\ 2x + y = -1 & (E_2) \end{cases}$$

En remplaçant (E_2) par $((E'_2) = E_2) - 2.(E_1)$, on élimine x de la seconde équation :

$$(S) \Leftrightarrow (S') : \begin{cases} x - y = 1 & (E_1) \\ 3y = -3 & (E'_2) \end{cases}$$

Les systèmes (S) et (S') ont donc pour unique solution le couple $(0, -1)$.

De façon générale, on utilise donc l'élimination par combinaison pour simplifier au maximum le système que l'on souhaite résoudre.

Johann Carl Friedrich Gauß (1777-1885) a étudié en détails l'élimination par combinaison. Il en a tiré une méthode systématique permettant de résoudre tout type de système linéaire : le Pivot de Gauss.

Le Pivot de Gauss permet en outre d'obtenir, au cours du calcul, de nombreuses informations sur le système étudié (notamment en terme de rang du système ou de dimension de l'ensemble des solutions).

D'autre part, les opérations du Pivot de Gauss, uniquement basées sur les différents coefficients des matrices du système étudié, peuvent être confiées à un ordinateur. Il est alors possible de construire un algorithme effectuant de façon automatique le pivot sur n'importe quel système donné.

Systèmes inversibles

Un système linéaire (S) est dit inversible s'il admet une unique solution. D'après la formule $n = r + d$, un système admet une unique solution si et seulement si $d = 0$. Autrement dit, les systèmes inversibles sont ceux pour lesquels le nombre de contraintes (i.e. le rang) est égal au nombre d'inconnues.

D'un point de vue matriciel, un système inversible est associé à une matrice A inversible. En multipliant l'équation $AX = B$ par A^{-1} à gauche, on a

$$AX = B \iff A^{-1}.AX = A^{-1}.B \iff I_n.X = A^{-1}.B$$

Autrement dit,

$$\boxed{AX = B \iff X = A^{-1}.B}$$

En calculant le déterminant de la matrice d'un système, on peut donc déterminer si ce système est inversible ou non sans le résoudre. Malheureusement, la résolution d'un système inversible (i.e. le calcul de A^{-1} est rarement simple, même à l'aide d'un ordinateur.

Les seuls systèmes inversibles pour lesquels le calcul de l'unique solution est simple sont les systèmes triangulaires :

$$\begin{cases} u_{11}x_1 + u_{12}x_2 + \dots + u_{1n}x_n = b_1 \\ \phantom{u_{11}x_1} + u_{22}x_2 + \dots + u_{2n}x_n = b_2 \\ \phantom{u_{11}x_1} \phantom{+ u_{22}x_2} + \ddots \phantom{+ u_{2n}x_n} = \vdots \\ \phantom{u_{11}x_1} \phantom{+ u_{22}x_2} + u_{nn}x_n = b_n \end{cases}$$

pour lesquels tous les coefficients u_{jj} sont non nuls.

Un tel système se résout alors en remontant puisque la dernière équation donne x_n , l'avant dernière donne x_{n-1} en fonction de x_n , etc. On verra en particulier que ce procédé peut s'automatiser.

Or on verra au paragraphe suivant que le pivot de Gauss permet précisément de transformer tout système inversible en un système triangulaire.

Systèmes non inversibles

Par opposition au cas précédent, un système carré est non inversible si sa matrice n'est pas inversible (i.e. si son déterminant est nul).

Dans ce cas, le rang de la matrice (et donc du système) est strictement inférieur au nombre d'inconnues. Si le système admet alors au moins une solution, la formule $n = r + d$ montre que l'ensemble des solutions est un sous espace de \mathbb{R}^n de dimension $d \geq 1$.

On note qu'alors, le système admet une infinité de solutions, et sa dimension permet de mesurer sa taille.

De plus, cette dimension donne également une indication sur la façon de décrire cet ensemble.

Précisément, la dimension d d'un sous espace de \mathbb{R}^n correspond au nombre de paramètres nécessaires à la description complète de l'ensemble.

Exemples :

– Soit

$$(S) : \begin{cases} x - 2y & = 0 & (E_1) \\ & y + z = 0 & (E_2) \\ x - y + z & = 0 & (E_3) \end{cases}$$

On peut rapidement vérifier que $(E_3) = (E_1) + (E_2)$. Autrement dit, l'équation (E_3) n'apporte pas de contrainte supplémentaire. On peut donc la supprimer du système :

$$(S) \iff (S') : \begin{cases} x - 2y & = 0 & (E_1) \\ & y + z = 0 & (E_2) \end{cases}$$

Les deux équations de (S') n'étant pas proportionnelles, elles sont indépendantes. Le rang du système (S') (et donc celui de (S)) est donc $r = 2$ et l'ensemble des solutions est de dimension $d = 3 - 2 = 1$.

Il est alors possible de décrire l'ensemble de ces solutions à l'aide d'un unique paramètre. Traditionnellement, on choisit l'une des variables et l'on se base sur les équations de (S') pour exprimer les autres variables en fonction de la variable choisie. Dans notre exemple, la variable y est la plus indiquée pour servir de paramètre puisque

$$(S') \iff \begin{cases} x = 2y \\ z = -y \end{cases}$$

Ainsi, l'ensemble des solutions est l'ensemble des triplets de la forme $(2y, y, -y)$:

$$D = \{(2y, y, -y), y \in \mathbb{R}\}.$$

Notes :

- Le choix de y est arbitraire. Il est possible de choisir la variable que l'on veut. Il est laissé au lecteur le soin d'exprimer l'ensemble D si dessus en fonction de x uniquement puis de z uniquement.
- Dans l'exemple ci dessus, le système (S) étant homogène, l'ensemble de ses solutions est un sous espace vectoriel de \mathbb{R}^3 . Puisqu'il est de dimension 1, il s'agit d'une droite passant par l'origine. Il est alors possible d'extraire de la description ci-dessus un vecteur directeur de cette droite. Il suffit pour cela de prendre pour y

n'importe quelle valeur non nulle. Il est laissé au lecteur le soin de vérifier que si l'on applique la même méthode aux autres représentations de D tous les vecteurs que l'on obtient sont colinéaires.

– Soit

$$(S) : \begin{cases} x - y - z = 0 & (E_1) \\ -x + y + z = 0 & (E_2) \\ 2x - 2y - 2z = 0 & (E_3) \end{cases}$$

Ici, les trois équations sont proportionnelles. Le système est donc de rang 1 et l'on peut ne garder qu'une seule équation, par exemple (E_1) . L'ensemble P des solutions est donc un sous-espace de \mathbb{R}^3 de dimension 2 (i.e. un plan). On peut décrire cet ensemble à l'aide de deux paramètres, par exemple y et z , l'équation (E_1) donnant x en fonction de y et z :

$$(E_1) \Leftrightarrow x = y + z$$

et

$$P = \{(y + z, y, z), y, z \in \mathbb{R}\}$$

Note : là encore, le système étant homogène, P est un plan vectoriel de l'espace. On peut extraire de la description ci-dessus une famille de deux vecteurs engendrant P :

$$\begin{aligned} \forall y, z \in \mathbb{R}, \quad (y + z, y, z) &= (y, y, 0) + (z, 0, z) \\ &= y \cdot (1, 1, 0) + z \cdot (1, 0, 1) \end{aligned}$$

P regroupe donc l'ensemble des combinaisons linéaires que l'on peut faire à partir des vecteurs $v_1 = (1, 1, 0)$ et $v_2 = (1, 0, 1)$.

Terminons cette partie par un mot sur les *systèmes incompatibles*. Un système est dit incompatible s'il contient des équations contradictoires, par exemple

$$\begin{cases} x + y = 1 \\ x + y = 0 \end{cases}$$

Un tel système n'admet aucune solution quel que soit le nombre de contraintes qu'il contient.

Ces systèmes incompatibles font partie des systèmes non inversibles car leur matrice est non inversible. L'incompatibilité vient alors du second membre.

1.2 Le pivot de Gauss

1.2.1 Le principe du pivot

Comme on l'a évoqué plus haut, le Pivot de Gauss est un procédé permettant de transformer un système linéaire carré en un système équivalent dont la famille des équations est

libre. Le principe du pivot est basée sur les deux principes suivants :

1. Un système d'équations n'ayant jamais le même nombre d'inconnues est toujours libre.
2. Si l'on ajoute une équation à une autre, on ne change pas l'ensemble des solutions du système.

Exercice : à démontrer.

L'idée de Gauss a été d'utiliser la propriété 2 pour éliminer des inconnues de façon à transformer le système de départ en un système vérifiant la propriété 1.

En appliquant cette idée avec méthode (i.e. en traitant les variables l'une après l'autre, dans l'ordre), Gauss a obtenu un procédé *algorithmique* au cours duquel les équations du système se simplifient de plus en plus. À la fin du procédé, le système de départ est remplacé par un système *triangulaire*.

D'autre part, si le système n'est pas de rang maximal, les équations liées sont remplacées par des lignes de 0 qui s'accumulent au bas du système.

Le nombre d'équations non nulles restant à la fin du procédé donne donc le rang de système de départ, et le nombre de lignes nulles donne la dimension de l'ensemble des solutions.

De plus, le système triangulaire se résout simplement en remontant les équations non nulles.

Exemples :

1. *Un système inversible.*

Le système

$$(S) : \begin{cases} x + y + z = 1 & (E_1) \\ x + 2y - z = 1 & (E_2) \\ 2x + y - z = -3 & (E_3) \end{cases}$$

$$\iff \begin{cases} x + y + z = 1 & (E_1) \\ + y - 2z = 0 & (E'_2) = (E_2) - (E_1) \\ - y - 3z = -5 & (E'_3) = (E_3) - 2.(E_1) \end{cases}$$

$$\iff \begin{cases} x + y + z = 1 & (E_1) \\ + y - 2z = 0 & (E'_2) \\ - 5z = -5 & (E''_3) = (E'_3) + (E'_2) \end{cases}$$

Le système obtenu est triangulaire. Chaque équation contient une inconnue de moins que la précédente. Il est donc libre, de rang 3 et la résolution se fait en remontant :

$$\begin{cases} (E_3'') & \Rightarrow z = 1 \\ (E_2') & \Rightarrow y = 2z = 2 \\ (E_1) & \Rightarrow x = 1 - y - z = -2 \end{cases}$$

L'unique solution de ce système (et du système (S) de départ) est le triplet $(-2, 2, 1)$.

2. Un système non inversible.

Dans le système

$$(S) : \begin{cases} x + y + z = 3 & (E_1) \\ +3y - 2z = 1 & (E_2) \\ 2x + 2y + 2z = 6 & (E_3) \end{cases}$$

on a $(E_3) = 2(E_1)$. Or pour éliminer x dans (E_3) , on la remplace par $(E_3') = (E_3) - 2(E_1)$:

$$(S) \iff \begin{cases} x + y + z = 3 & (E_1) \\ +3y - 2z = 1 & (E_2) \\ 0 = 0 & (E_3') \end{cases}$$

on obtient un système libre de rang 2. L'ensemble des solutions est donc une droite de l'espace (il est laissé au lecteur le soin de décrire cet ensemble à l'aide par exemple de la variable z).

Note : en pratique, on travaille sur la matrice étendue du système. L'élimination de variables revient alors à placer des 0 dans les colonnes de cette matrice.

1.2.2 L'algorithme

Comme on l'a vu plus haut, le Pivot de Gauss n'est une application organisée du principe d'élimination.

Un exemple

Considérons le système suivant :

$$\begin{cases} 2x_1 + x_2 - 2x_3 + x_4 = 0 \\ 3x_1 - x_2 + 2x_3 - 2x_4 = 6 \\ -x_1 + 2x_2 - 2x_3 - x_4 = -3 \\ x_1 - x_2 + 2x_3 - x_4 = 3 \end{cases}$$

dont la matrice étendue est

$$A = \left(\begin{array}{cccc|c} 2 & 1 & -2 & 1 & 0 \\ 3 & -1 & 2 & -2 & 6 \\ -1 & 2 & -2 & -1 & -3 \\ 1 & -1 & 2 & -1 & 3 \end{array} \right)$$

Hormis la dernière, chacune des colonnes de la matrice A correspond à une inconnue du système. Le Pivot de Gauss consiste alors à traiter chacune de ces colonnes à l'aide de combinaisons de lignes de la forme

$$L_i \leftarrow L_i - \lambda L_j$$

pour y placer un maximum de 0.

Dans notre exemple, on peut ainsi placer trois 0 dans la première colonne, en dessous du premier coefficient, à l'aide des opérations

$$L_2 \leftarrow L_2 - \frac{3}{2}L_1$$

$$L_3 \leftarrow L_3 + \frac{1}{2}L_1,$$

$$L_4 \leftarrow L_4 - \frac{1}{2}L_1.$$

La matrice A devient alors

$$A^{(2)} = \left(\begin{array}{cccc|c} \boxed{2} & -1 & -2 & 1 & 0 \\ 0 & -\frac{5}{2} & 5 & -\frac{7}{2} & 6 \\ 0 & \frac{5}{2} & -3 & -\frac{1}{2} & -3 \\ 0 & -\frac{3}{2} & 3 & -\frac{3}{2} & 3 \end{array} \right).$$

Ici, le pivot est le premier coefficient de la première ligne. C'est (entre autres) lui qui détermine les coefficients λ permettant de faire apparaître des 0.

La première étape est terminée. On a traité la première colonne à l'aide de la première ligne. On passe ensuite à la deuxième colonne que l'on va traiter à l'aide de la deuxième ligne. Le nouveau pivot est alors le deuxième coefficient de la deuxième ligne et les opérations

$$L_3 \leftarrow L_3 + L_2$$

$$L_4 \leftarrow L_4 - \frac{3}{5}L_2$$

transforment $A^{(2)}$ en

$$A^{(3)} = \left(\begin{array}{cccc|c} 2 & 1 & -2 & 1 & 0 \\ 0 & \boxed{-\frac{5}{2}} & 5 & -\frac{7}{2} & 6 \\ 0 & 0 & 2 & -4 & 3 \\ 0 & 0 & 0 & \frac{3}{5} & -\frac{3}{5} \end{array} \right).$$

On a terminé le traitement de la seconde colonne. En théorie, il nous reste la troisième colonne à traiter. Cependant, on remarque que lors des opérations précédentes, on a gagné un 0 supplémentaire sur la dernière ligne. Le pivot est donc terminé car la matrice $A^{(3)}$

correspond à un système triangulaire de rang trois. On le résout en remontant et on trouve rapidement l'unique solution du système de départ :

$$X = \begin{pmatrix} 1 \\ -2 \\ -\frac{1}{2} \\ -1 \end{pmatrix}$$

Traitement des colonnes

À partir de l'exemple que l'on vient de voir, on peut construire les premières fonctions de l'algorithme de trigonalisation.

Pour cela, il faut généraliser les opérations précédentes à une matrice générale A de la forme

$$A = (a_{ij}) = \begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix}$$

Le premier pivot est a_{11} . Pour construire $A^{(2)}$, on effectue les opérations

$$\begin{aligned} L_2 &\leftarrow L_2 - \frac{a_{21}}{a_{11}} L_1 \\ L_3 &\leftarrow L_3 - \frac{a_{31}}{a_{11}} L_1 \\ &\vdots \\ L_n &\leftarrow L_n - \frac{a_{n1}}{a_{11}} L_1 \end{aligned}$$

On obtient alors la matrice

$$A^{(2)} = (a_{ij}^{(2)}) = \begin{pmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & a_{n3}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix}$$

Remarque : ceci n'est valable que si le pivot $a_{11} \neq 0$. On supposera dans tout ce qui suit que c'est le pivot est toujours non nul. On verra plus tard comment modifier la

méthode pour les cas où l'on tombe sur un pivot nul.

On peut traduire cette opération à l'aide d'une boucle `for` :

```

for i in range(2,n):
    Li ← Li -  $\frac{a_{i1}}{a_{11}}$ L1

```

De façon générale, pour passer de

$$A^{(j)} = \begin{pmatrix} a_{11}^{(j)} & a_{12}^{(j)} & \cdots & a_{1j}^{(j)} & \cdots & a_{1n}^{(j)} \\ 0 & a_{22}^{(j)} & \ddots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \vdots & & \vdots \\ \vdots & & 0 & \boxed{a_{jj}^{(j)}} & & \vdots \\ \vdots & & \vdots & a_{j+1,j}^{(j)} & \ddots & \vdots \\ 0 & \cdots & 0 & \vdots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nj}^{(j)} & \cdots & \cdots & a_{nn}^{(j)} \end{pmatrix}$$

à une matrice $A^{(j+1)}$ qui contient une colonne de zéros supplémentaire, on effectue les opérations

$$\begin{aligned} L_{j+1} &\leftarrow L_{j+1} - \frac{a_{j+1,j}^{(j)}}{a_{jj}^{(j)}} L_j \\ &\vdots \\ L_n &\leftarrow L_n - \frac{a_{nj}^{(j)}}{a_{jj}^{(j)}} L_j \end{aligned}$$

On peut encore traduire cela à l'aide d'une boucle `for` :

```

for i in range(j+1,n):
    Li ← Li -  $\frac{a_{ij}}{a_{jj}}$ Lj

```

Le pivot de Gauss complet consiste alors à appliquer cette boucle $n - 1$ fois. On peut

donc l'écrire sous la forme de deux boucles :

```

for j in range(1,n):
    for i in range(j+1,n):
        Li ← Li -  $\frac{a_{ij}}{a_{jj}}$ Lj

```

Choix du Pivot

Comme beaucoup d'algorithmes, l'algorithme est simple tant que l'on ne tient pas compte des cas exceptionnels. Mais il ne marche pas (ou mal) dans certains cas. Cela peut par exemple être le cas si l'algorithme rencontre un pivot nul. Étudions encore un cas "à la main" avant d'en déduire la version algorithmique :

Exemple :

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

Après la première étape du pivot, on obtient la matrice

$$A_2 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & \boxed{0} & 1 \\ 0 & 3 & 1 \end{pmatrix}$$

Pour pouvoir continuer le Pivot, il faut échanger les deux dernières lignes de A_2 :

$$A'_2 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & \boxed{3} & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Une fois que l'on a échangé le pivot avec un pivot non nul, on peut continuer le procédé. (Ici, on a terminé car la matrice est petite).

Pour trouver un pivot non nul, on a donc parcouru la colonne se trouvant sous le pivot nul à la recherche d'un coefficient non nul. Une fois repéré, on a échangé la ligne correspondante avec la ligne du pivot nul.

On peut traduire cette recherche sous forme algorithmique :

```

for k in range(j+1,n):
    if akp != 0:
        Lj ↔ Lk
        break

```

En ajoutant ce test aux boucles vues plus haut, on obtient un algorithme qui marche bien, à condition que l'on puisse trouver un pivot non nul pour remplacer chacun des pivots nuls rencontrés. On peut montrer que c'est le cas si la matrice A est inversible.

On obtient donc un algorithme efficace pour trigonaliser les matrices inversibles. Or la plupart des matrices sont inversibles. Cet algorithme marche donc en théorie la plupart du temps. Il existe cependant une limite à son utilisation pour des matrices trop grosses. En effet, le fait d'utiliser une division peut entraîner une "explosion" des coefficients, notamment si l'on divise par une quantité trop petite, ou une "disparition" si l'on divise par un nombre trop grand. On peut, dans une certaine mesure, combattre cet inconvénient en améliorant le choix du pivot mais la manipulation des grosses matrices reste un problème difficile en algorithmique.

Un mot sur les matrices non inversibles

L'algorithme que l'on vient de voir traite parfaitement les matrices inversibles (appliqué à une matrice non inversible, il s'arrête car il rencontre une division par 0).

Il ne permet donc de résoudre que des problèmes admettant une et une seule solution. Dans de nombreux domaines (notamment en mécanique ou en modélisation numérique), cet algorithme suffit pour traiter les problèmes linéaires rencontrés.

En étudiant le problème de près, il cependant est possible de modifier le programme présenté plus haut pour qu'il traite également les systèmes non inversibles.

L'algorithme transforme alors une matrice carrée donnée en une matrice *échelonnée*, i.e. une matrice dans laquelle chaque ligne contient *au moins* un zéro de plus que la ligne précédente ; les dernières lignes étant remplacées par des lignes de 0.

Or dans un système échelonné, les lignes non nulles forment automatiquement un système indépendant.

Autrement dit, à la fin du pivot de Gauss, le nombre de ligne non nul est exactement le nombre de contraintes. En la relation $r + d = n$ se traduit par le fait que le nombre de ligne nulles correspond, lui, à la dimension de l'ensemble des solutions (si le système n'est pas incompatible).

Notes :

- Dans le cas des matrices inversibles, les résultats que l'on vient de voir restent valables : un système triangulaire est un système de rang n maximal. L'ensemble des solutions est donc de dimension 0. C'est un point.
- Si l'on applique ces méthodes à un système incompatible, on voit apparaître, parmi les lignes de 0, une ligne correspondant à une équation de la forme $0 = a$ avec $a \neq 0$. Il est alors clair que le système étudié n'admet pas de solution.

1.2.3 Pivot de Gauss et opérations matricielles

L'algorithme que l'on vient de voir (modifié ou non pour traiter les matrices non inversibles) est un algorithme très général. Or en pratique, si l'on a à utiliser ces outils pour traiter des problèmes linéaires, les systèmes que l'on a à étudier peuvent avoir des caractéristiques particulières (matrices symétriques en mécanique, matrices à bandes en discrétisation, matrices creuses...). Selon les cas, on peut donc encore améliorer l'algorithme pour qu'il soit plus efficace dans le cas particulier qui nous concerne.

Ces améliorations passent en particulier par le passage à la représentation matricielle du Pivot. En effet, en détaillant les actions de la multiplication matricielle, on va voir que l'on peut traduire les opérations du pivot de Gauss en opérations matricielles.

Ces notions portant sur les matrices, on élargie notre étude à l'ensemble des matrices (en oubliant pour un temps les éventuels systèmes qu'ils représentent). On travaille alors uniquement sur des matrices carrées (et non plus sur les matrices étendues d'un système donné).

Ainsi, si A est une matrice carrée, on peut, à partir des coefficients d'une matrice A , construire une matrice M telle que la matrice triangulaire U associée à A via le pivot soit donnée par le produit

$$U = M \times A.$$

En pratique, il est plus pratique et plus utile de construire la matrice $L = M^{-1}$. La matrice A de départ est alors remplacée par les matrices L et U telles que

$$A = L \times U.$$

On peut alors remplacer tout système linéaire de matrice A par des systèmes associés aux matrices L et U . Or ces matrices étant plus simple que A (on peut en effet montrer que, comme U , la matrice L est triangulaire), les systèmes associés à A (et de façon général tous les problèmes associés à la matrice A) sont plus faciles à traiter.

Matrices de combinaisons

Pour modéliser le Pivot de Gauss par un produit de matrices, il faut, dans un premier temps, traduire les opérations du type

$$L_i \leftarrow L_i + \lambda L_j.$$

Or soit

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \lambda & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Si l'on multiplie un vecteur $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$ par E_1 , on observe

$$E_1 \times \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} a \\ b + \lambda a \\ c \\ d \end{pmatrix}$$

De façon générale, si l'on multiplie une matrice A (à 4 lignes) par E_1 à gauche, on agit sur chaque colonne de A comme sur le vecteur précédent. Autrement dit, on effectue la combinaison

$$L_2 \leftarrow L_2 + \lambda L_1.$$

Note : les indices 2 et 1 de la combinaison correspondent respectivement à la ligne et la colonne du coefficient λ dans E_1 .

De même, la matrice

$$E_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \mu & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

effectue l'opération

$$L_3 \leftarrow L_3 + \mu L_1.$$

De façon générale, on peut définir, pour toute taille n la matrice $n \times n$

$$E(p, q, \lambda) = \begin{matrix} & & & & C_q & & & \\ & & & & \downarrow & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ L_p \rightarrow & \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \vdots \\ \vdots & & 0 & 1 & \ddots & \vdots \\ \vdots & & \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & \lambda & 0 & \ddots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} & & \end{matrix}$$

Ici, p et q sont des entiers entre 1 et n , avec $p < q$ pour être sur que le coefficient λ soit sous la diagonale.

On peut vérifier, en revenant à la formule du produit de deux matrices, que la multiplication à gauche de n'importe quelle matrice à n lignes par $E(p, q, \lambda)$ effectue l'opération

$$L_p \leftarrow L_p + \lambda L_q$$

dans A . On dit que $E(i, j, \lambda)$ est une matrice de combinaison élémentaire.

Ainsi, pour appliquer le pivot de Gauss à une matrice carrée A inversible, on peut la multiplier successivement à gauche par différentes matrices de combinaisons élémentaires :

$$A \longrightarrow E_1.A \longrightarrow E_2.E_1.A \longrightarrow \cdots \longrightarrow E_p \cdots E_1.A = U$$

En notant $M = E_p \cdots E_1$, on montre qu'il existe une matrice M telle que $MA = U$.

En pratique, il est plus simple de construire la matrice $L = M^{-1}$ à partir des coefficients de A . On peut en effet montrer que pour L est une matrice triangulaire inférieure dont tous les termes diagonaux valent 1. De plus, on peut montrer que pour remplir les colonnes de L (sous sa diagonale), il suffit de récupérer les coefficients $\frac{a_{ij}}{a_{jj}}$ des combinaisons linéaires que l'on souhaite appliquer à A

Exemple :

$$A = \begin{pmatrix} 1 & 1 & -2 \\ -2 & 0 & 1 \\ -1 & 3 & 3 \end{pmatrix} \implies \left\{ L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 3 & 1 \end{pmatrix}, U = \begin{pmatrix} 1 & 1 & -2 \\ 0 & -2 & 3 \\ 0 & 0 & -10 \end{pmatrix} \right\}$$

D'un point de vue algorithmique, on peut donc ajouter aux boucles vues plus haut une commande permettant de remplir la matrice L au fur et à mesure.

Matrices de permutation

Pour obtenir une représentation complète du Pivot de Gauss sous forme matricielle, il faut maintenant étudier la permutation de deux lignes. On va voir que là encore, on peut représenter la permutation de deux lignes d'une matrice A par la multiplication d'une matrice P à gauche.

Ainsi, en dimension 3, on peut vérifier rapidement que la matrice

$$P_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

échange les lignes 1 et 2 :

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} b \\ a \\ c \end{pmatrix}$$

En dimension n , les matrices P_{ij} définies par

$$P_{ij} = \begin{pmatrix} 1 & 0 & & \cdots & & 0 \\ \vdots & \ddots & & & & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & & \ddots & \vdots \\ 0 & & \cdots & & 0 & & 1 \end{pmatrix}$$

permutent les lignes L_i et L_j . Elles sont appelées matrices de permutation élémentaires (une matrice de permutation élémentaire est obtenue en permutant dans la matrice I_n les lignes correspondantes).

On appelle alors matrice de permutation un produit de matrices de permutations élémentaires. On peut montrer que le produit de matrices de permutation élémentaires est toujours une matrice contenant un 1 et que des zéros sur chaque ligne et sur chaque colonne. (Il y a donc n 1).

On peut en outre montrer que les matrices de permutation élémentaires sont toutes de déterminant -1 et sont égales à leur propre inverse :

$$P_{ij}^2 = I_n$$

Ainsi, si une matrice de permutation P s'écrit

$$P = P_k \times \dots \times P_1$$

on a $\det P = (-1)^k$ (on appelle parfois signature de la permutation ce déterminant) et

$$P^{-1} = P_1 \cdots P_k$$

Décomposition LU générale

On a maintenant tous les outils pour exprimer le Pivot de Gauss sur une matrice A sous la forme de produits de matrices sur A .

Si le passage de A à U ne nécessite pas de permutation de ligne, on a vu comment construire la matrice L , triangulaire inférieure telle que $A = LU$.

Si on doit faire des permutations au cours du Pivot de Gauss, chaque permutation correspond à une matrice de permutation élémentaire P_i . L'ensemble des permutations effectuées au cours du Pivot de Gauss sont alors résumées dans le produit $P = P_k \cdots P_1$ de toutes ces permutations élémentaires. On doit alors tenir compte de ces permutations

dans la décomposition LU mais le principe reste le même. Précisément, si P représente toutes les permutations que l'on doit effectuer au cours du pivot, c'est la matrice PA que l'on décompose sous la forme

$$PA = LU.$$

En pratique, on ne connaît les permutations à effectuer qu'en faisant effectivement le Pivot de Gauss. On construit donc la matrice P de permutations en même temps que les matrices L et U .

D'autre part, lorsqu'une permutation est nécessaire au cours du Pivot de Gauss, on a en général déjà commencé à remplir la matrice L . Il faut alors effectuer dans L des permutations permettant de tenir compte de ces permutations. Précisément, si l'on doit échanger deux lignes L_i et L_j dans $A^{(i)}$, on doit permuter, dans L tous les coefficients des lignes L_i et L_j qui se trouvent sous la diagonale.

Exemple :

$$A = \begin{pmatrix} 2 & 1 & 0 & 4 \\ -4 & -2 & 3 & 7 \\ 4 & 1 & -2 & 8 \\ 0 & -3 & -12 & -1 \end{pmatrix}$$

Note : la décomposition LU d'une matrice A donnée n'est pas unique. En effet, si l'on rencontre un pivot nul au cours du procédé, on peut choisir comme pivot non nul n'importe lequel des coefficients non nuls qui se trouvent dessous. Selon les choix que l'on fait, les matrices P , L et U que l'on obtient peuvent être totalement différentes, mais vérifient toujours la relation $PA = LU$.

1.2.4 Applications du Pivot de Gauss et de la décomposition LU

Résolution de système

La première application du Pivot de Gauss est évidemment la résolution de (gros) systèmes linéaires.

Si l'on n'a qu'un seul système à résoudre, on se contente d'appliquer le pivot à la matrice étendue du système, puis on résout le système triangulaire ainsi obtenu.

On verra en TP comment programmer un tel algorithme.

Cependant, il est également possible de se baser sur la décomposition LU de la matrice A du système étudié.

Précisément, si P , L et U sont les trois matrices issues de la décomposition LU de A ,

on peut alors remplacer tout système de la forme $AX = B$ par le couple d'équations

$$\begin{cases} LY &= PB \\ UX &= Y_0 \end{cases}$$

Autrement dit, on commence par résoudre le système $LY = PB$. On utilise alors l'unique solution Y_0 obtenue comme second membre pour le système $UX = Y_0$. On obtient alors une unique solution X_0 . Or puisque $A = P^{-1}.L.U$, on a

$$A.X_0 = P^{-1}.L.\underbrace{U.X_0}_{Y_0} = P^{-1}.\underbrace{L.Y_0}_{P.B} = \underbrace{P^{-1}.P}_{I_n}.B = B$$

Le vecteur X_0 correspond donc bien à l'unique solution du système $AX = B$.

Cette méthode permet donc de remplacer la résolution d'un système linéaire quelconque par la résolution de deux systèmes linéaires triangulaires.

Cette méthode est en particulier utilisée pour résoudre plusieurs systèmes linéaires de la forme $AX = B_i$ associés à une même matrice A mais avec des seconds membres différents. On commence par calculer la décomposition LU de A puis on remplace chacun des systèmes $AX = B_i$ par les couples de systèmes triangulaires

$$\begin{cases} LY &= PB_i \\ UX &= Y \end{cases}$$

Calcul de déterminant

Une autre application importante du Pivot de Gauss est le calcul de déterminant. On peut en effet montrer que

- si l'on effectue, dans une matrice carrée, des opérations de la forme $L_i \leftarrow L_i - \lambda L_j$ on ne change pas le déterminant de la matrice,
- si l'on échange deux lignes dans une matrice carrée, on multiplie son déterminant par -1 .

De plus, grâce à la méthode de développement par rapport à une colonne, il est clair que le déterminant d'une matrice triangulaire est donné par le produit de ses termes diagonaux.

Ainsi, si l'on souhaite connaître le déterminant d'une matrice A donnée, on lui applique le pivot de Gauss en comptant le nombre p de permutations de lignes que l'on effectue. Le déterminant cherché est alors obtenu en multipliant par $(-1)^p$ le produit des termes diagonaux de la matrice triangulaire issue du pivot.

Du point de vue de la décomposition LU , ce résultat se retrouve en notant que

- le déterminant étant multiplicatif, on a

$$PA = LU \Rightarrow \det(P). \det(A) = \det(L). \det(U),$$

- le déterminant de P est précisément $(-1)^p$,
- le déterminant de L est 1 puisque L est triangulaire et n'a que des 1 sur sa diagonale,
- le déterminant de U est égal au produit de ses termes diagonaux.

On a donc

$$\det(A) = \frac{1}{\det(P)} \det(U) = (-1)^p \det(U).$$

Inversion de matrice

Le Pivot de Gauss peut également servir au calcul de l'inverse d'une matrice $A = (a_{ij})$ donnée. En effet, une méthode de calcul est la suivante

1. On place côte à côte la matrice A que l'on souhaite inverser et la matrice identité de même taille :

$$\left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & \cdots & 0 & 1 \end{array} \right)$$

2. On effectue des opérations de lignes sur cette "double matrice" de façon à transformer la matrice A en la matrice identité,

À la fin du procédé, la partie de gauche de la double matrice contient la matrice identité, et la partie de droite contient la matrice A^{-1} :

$$(I_n \mid A^{-1})$$

Or pour transformer la matrice A en la matrice identité, on utilise les opérations du Pivot de Gauss. Précisément, on commence par appliquer le Pivot que l'on connaît à A pour la rendre triangulaire. Une fois la matrice triangulaire obtenue, on effectue un "pivot inversé" en prenant comme premier pivot le dernier coefficient diagonal et en l'utilisant pour placer des 0 au dessus, dans la dernière colonne. On traite ainsi toutes les colonnes de la dernière à la deuxième. On obtient alors une matrice diagonale. Pour obtenir la matrice identité, il suffit alors de diviser chacune des lignes par le terme diagonal correspondant (à noter que cette dernière opération n'est pas une opération du pivot de Gauss).

Exemple : calculer l'inverse de la matrice $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix}$.

1.3 Systèmes sur déterminés

Jusqu'ici, on a vu comment traiter des problèmes admettant une ou plusieurs solutions : des systèmes linéaires où le nombre de contraintes est au plus égal au nombre d'inconnues.

À l'inverse, on appellera système surdéterminé un système linéaire contenant plus de contraintes que de degrés de liberté : soient

- $A = (a_{ij})$ une matrice rectangulaire à m lignes et n colonnes, avec $m > n$,
- $B = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ un vecteur de constantes,
- $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ un vecteur d'inconnues.

Le système linéaire $AX = B$ n'admet pas de solution exacte. On doit alors se contenter de déterminer X tel que le vecteur AX soit "le plus proche possible" du vecteur B . Autrement dit, on cherche un vecteur X qui *minimise* la norme $\|AX - B\|$.

Rappel : la norme d'un vecteur $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ est

$$\|X\| = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{x_1^2 + \dots + x_n^2}.$$

On va voir comment interpréter cette recherche d'un point de vue géométrique, et on verra comment le calcul matriciel permet, là encore, de construire une méthode de résolution algorithmique.

1.3.1 Un exemple

Considérons le système

$$(S) : \begin{cases} x - y & = 1 \\ x + y & = 1 \\ -x + 2y & = 2 \end{cases}$$

ou

$$AX = B \quad \text{avec} \quad A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ -1 & 2 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

On peut rapidement montrer que ce système n'admet pas de solution. On cherche donc le vecteur $X = \begin{pmatrix} x \\ y \end{pmatrix}$ tel que le vecteur $AX - B$ soit le plus petit possible. À priori, la taille de ce vecteur est donné par la norme euclidienne

$$\|AX - B\| = \sqrt{(x - y - 1)^2 + (x + y - 1)^2 + (-x + 2y - 2)^2}.$$

Pour simplifier les calculs, on minimise en réalité le carré de cette norme :

$$\Phi(x, y) = \|AX - B\|^2 = (x - y - 1)^2 + (x + y - 1)^2 + (-x + 2y - 2)^2.$$

$\Phi(x, y)$ est alors un polynôme de degré 2 en x et y . On sait déterminer, s'il existe les extrema d'une fonction de deux variables :

- on détermine les points critiques : les couples (x, y) qui annulent les dérivées partielles de Φ ,
- on détermine la nature de chacun de ces points critiques en étudiant les dérivées secondes.

D'autre part, pour le cas particulier qui nous occupe, Φ étant un polynôme de degré 2, les dérivées partielles de Φ sont des formes linéaires en x et y . Rechercher les points critiques revient alors à résoudre un système linéaire de deux équations à deux inconnues. L'ensemble des points critiques est donc soit un unique point, soit une droite.

Exercice : à faire à l'aide des dérivées partielles de Φ .

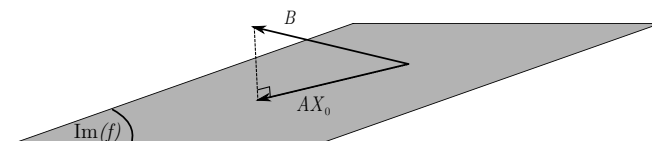
1.3.2 Interprétation géométrique

De façon générale, une matrice $A \in \mathcal{M}_{mn}(\mathbb{R})$ peut être vue comme la matrice d'une application linéaire $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (exprimée par exemple dans la base canonique). L'ensemble des vecteurs de la forme AX pour $X \in \mathbb{R}^n$ est alors l'image $\text{Im}(f)$. C'est un sous-espace vectoriel de \mathbb{R}^m dont la dimension est le rang de A . Précisément, le SEV $\text{Im}(f)$ est engendré par les colonnes de A .

Ainsi, dans notre exemple, A est la matrice d'une application linéaire $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ et $\text{Im}(f)$ est le SEV de \mathbb{R}^3 engendré par les vecteurs $(1, 1, -1)$ et $(-1, 1, 2)$. C'est un plan.

Un système de la forme $AX = B$ admet alors une solution si et seulement si $B \in \text{Im}(f)$. Si c'est le cas, le vecteur X cherché contient les coordonnées de B exprimé dans la base de $\text{Im}(f)$ donnée par les colonnes de A .

Si ça n'est pas le cas, chercher une solution optimale au sens où on l'a vu plus haut revient à chercher le vecteur de $\text{Im}(f)$ (de la forme AX) le plus proche de B . Dans notre exemple, la situation peut être représentée par le dessin ci-dessous.



On voit alors que la solution X_0 que l'on cherche est un vecteur tel que AX_0 soit le projeté orthogonal de B sur le plan $\text{Im}(f)$. Autrement dit, le vecteur $AX_0 - B$ doit être orthogonal au plan $\text{Im}(f)$.

En écrivant que le produit scalaire de $AX_0 - B$ avec chacune des lignes de A doit être nul, on obtient alors un système carré que doivent vérifier les coordonnées de X_0 .

Exercice : déterminer ce système dans le cas de l'exemple ci-dessus.

1.3.3 Les équations normales

Quelque soit l'approche choisie (analytique ou géométrique), on a vu que la recherche du "meilleur" X se réduisait à la résolution d'un système carré. Or on peut montrer que quelque soit l'approche choisie, ce système carré est toujours le même et dépend uniquement de la matrice A et du vecteur B du système de départ. Précisément, il s'agit du système

$${}^t A A X = {}^t A B.$$

appelé *système des équations normales*.

On peut alors utiliser les méthodes de résolution classiques pour résoudre ce système normal. D'autre part, la matrice ${}^t A A$ de ce système étant particulière (elle est symétrique, définie positive), on peut améliorer les méthodes de décomposition et donc la performance des algorithmes de résolution.

1.3.4 Application : la régression

Régression linéaire simple

Ce point de vue matriciel permet de retrouver les formules de régression linéaire simple et de les généraliser aux dimensions supérieures.

Rappel : étant données deux séries statistiques (x_i, y_i) associées à une population de n individus, on peut représenter ces séries par un nuage de points $M_i = (x_i, y_i)$ dans le plan. Si tous les points du nuage sont alignés, c'est qu'il existe un lien linéaire entre les séries x et y . Autrement dit, il existe deux coefficients a et b tels que

$$\forall i = 1..n, \quad y_i = ax_i + b.$$

On dit que la variable y est expliquée par la variable x .

En pratique, c'est rarement le cas. Cependant, si les points du nuage sont "globalement alignés", on cherche la droite qui colle le mieux à la forme du nuage. C'est la droite de régression.

Or on peut représenter ce problème à l'aide d'un système linéaire sur déterminé. Précisément, chercher une droite qui passe par tous les points du nuage revient à chercher deux coefficients a et b tels que pour tout $i = 1..n$, on ait

$$y_i = ax_i + b \iff ax_i + b = y_i$$

On a donc un système de n équations à 2 inconnues

$$AX = B$$

où

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \quad X = \begin{pmatrix} a \\ b \end{pmatrix}, \quad B = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Si les points ne sont pas alignés, ce système n'admet pas de solution. On cherche donc à minimiser la norme

$$\Phi(a, b) = \left\| A \begin{pmatrix} a \\ b \end{pmatrix} - B \right\|^2.$$

D'un point de vue géométrique, cela revient à minimiser la somme des carrés des distances de chacun des points à la droite d'équation $y = ax + b$.

D'après l'étude précédente, on doit donc résoudre le système

$${}^t A A X = {}^t A B.$$

En détaillant les calculs, on retrouve les formules de régression : la meilleure droite est donnée par les coefficients a_0 et b_0 définis par

$$a_0 = \frac{\text{Cov}(x, y)}{V(x)} \quad \text{et} \quad b_0 = \bar{y} - a_0 \bar{x}.$$

En effet, si les variables sont centrées (i.e. $\bar{x} = \bar{y} = 0$), on a

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$$

et

$$\begin{aligned} {}^t A A &= \begin{pmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \\ &= n \begin{pmatrix} V(x) & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

et

$$\begin{aligned} {}^tAB &= \begin{pmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix} \\ &= n \begin{pmatrix} \text{Cov}(x, y) \\ 0 \end{pmatrix} \end{aligned}$$

Le système ${}^tAAX = {}^tB$ s'écrit donc

$$\begin{cases} V(x)a = \text{Cov}(x, y) \\ b = 0 \end{cases}$$

Généralisations

Le point de vue matriciel permet de généraliser cela aux dimensions supérieures. Ainsi, de façon générale, si l'on a $m + 1$ variables statistiques x_1, \dots, x_m, y portant sur une population de n individus, on peut chercher à expliquer y par les x_i à l'aide d'une équation de la forme

$$y = a_1x_1 + a_2x_2 + \dots + a_mx_m + a_{m+1}$$

en résolvant le système sur déterminé $AX = B$ où

$$A = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} & 1 \end{pmatrix}, \quad X = \begin{pmatrix} a_1 \\ \vdots \\ a_{m+1} \end{pmatrix}, \quad B = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Si $m = 2$, on a 3 séries statistiques x_1, x_2 et y portant une sur une même population et on a encore une représentation géométrique de la population : on peut représenter chaque individu par un point $M_i = (x_{1i}, x_{2i}, y_i)$ de l'espace. Expliquer y à l'aide de x_1 et x_2 revient à trouver des coefficients a, b et c tels que

$$y \approx ax_1 + bx_2 + c$$

D'un point de vue géométrique, cela revient à trouver un plan de l'espace qui passe au plus près de chacun des points du nuage.

En posant le problème en terme de systèmes sur déterminés, on doit résoudre le système

$$AX = B$$

où

$$A = \begin{pmatrix} x_{11} & x_{21} & 1 \\ \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & 1 \end{pmatrix}, \quad X = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad B = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Les coefficients a_0 , b_0 et c_0 idéaux sont donnés par le système carré

$${}^tAA = {}^tAB.$$

Si la matrice tAA est inversible, ce système admet une unique solution et on trouve un unique plan.

En revenant à la dimension 1, on peut également généraliser la notion de régression en augmentant le degré de la courbe cherchée. Ainsi, si les points (x_i, y_i) semblent s'égrainer le long d'une parabole, on peut chercher, entre x et y un lien de degré 2 de la forme

$$y \approx ax^2 + bx + c$$

Là encore, la mise en équation du problème produit un système linéaire $AX = B$ sur-déterminer, où A est alors de la forme

$$A = \begin{pmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{pmatrix}$$

La résolution du système normal associé à cette matrice A donne alors la parabole passant au plus près de chacun des points (x_i, y_i) du nuage.

Mesure de l'erreur

La méthode décrite plus haut a un problème inattendu : elle donne un résultat dans tous les cas.

Il faut donc trouver un moyen de mesurer la pertinence du résultat produit. Les statisticiens ont élaboré plusieurs indices permettant de mesurer la qualité de leurs différentes régressions : les coefficients de corrélation.

Tous ces indices sont basés sur l'idée suivante : lorsque l'on prend pour solution de $AX = B$ l'unique solution X_0 du système normal ${}^tAAX = {}^tAB$, l'erreur commise est précisément la norme

$$\varepsilon = \|AX_0 - B\|.$$

Il s'agit là d'une erreur absolue. Pour avoir une idée de l'importance de cette erreur, on doit la comparer aux dimensions du problème. On peut ainsi envisager un indice de pertinence de la forme

$$\varepsilon_r = \frac{\|AX_0 - B\|}{\|B\|}.$$

L'approximation donnée par le système normal est alors d'autant meilleure que l'erreur ε_r est proche de 0.

En pratique, pour de la régression linéaire simple, les statisticiens utilisent le coefficient de corrélation

$$r(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{V(x)V(y)}}.$$

Avec nos notations, on peut montrer que

$$\|AX_0 - B\|^2 = V(y)(1 - r^2(x, y)).$$

Et puisque $B = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, on a $\|B\|^2 = \sum_{i=1}^n y_i^2 = V(y)$. Ainsi, l'erreur relative définie plus haut vérifie

$$\varepsilon_r^2 = 1 - r^2(x, y)$$

et plus le coefficient $r(x, y)$ est proche de ± 1 , meilleure est la régression.

Chapitre 2

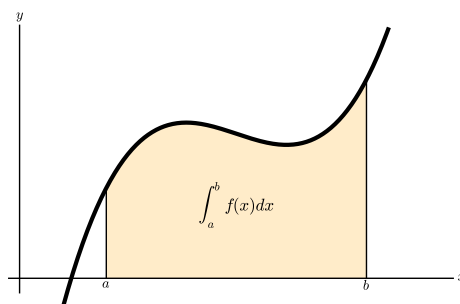
Fonctions spéciales

Introduction

Étant donnée une fonction réelle f définie et continue (par morceaux) sur un intervalle $[a, b]$, on connaît

$$\int_a^b f(t) dt.$$

On peut en particulier l'interpréter comme étant l'aire sous la courbe représentant f , coincée entre les droites d'équations $x = a$ et $x = b$:



cette représentation donnant une méthode de mesure dans le plan. Nous avons également vu dans les chapitres précédents comment une généralisation de cet outil aux dimensions supérieures, donnant des outils de mesure supplémentaires dans le plan et l'espace (et dans les espaces de dimensions supérieures)

Cependant, dans toutes les notions que l'on a vu jusqu'ici, on ne mesure que des domaines bornés.

Or il est possible de lever cette limite en étendant la définition d'intégrale aux domaines non bornés.

Dans ce cours, on va voir dans un premier temps qu'il est ainsi possible d'étendre la notion d'intégrale aux intervalles ouverts de la forme

$$[a, +\infty[,]-\infty, b], [a, b[,]a, b], \text{ etc}$$

(ce qui se généralise naturellement aux dimensions supérieures).

On verra ensuite deux fonctions particulière définies par des intégrales : la fonction Γ et la fonction Beta. Les intégrales intervenant dans la définition de ces fonctions sont bien entendues définies sur des intervalles ouverts. Ces deux fonctions sont souvent qualifiées de "des serpents de mer" car elles apparaissent dans de nombreux domaines des mathématiques (proba, fiabilité, arithmétique,...) et plus généralement lorsque l'on tombe sur une intégrale que l'on ne sait pas calculer (i.e. pour laquelle on ne connaît pas de primitive).

2.1 Intégrales généralisées

2.1.1 Intervalles $[a, +\infty[$

Soit f une fonction continue définie sur un intervalle de la forme $[a, +\infty[$. On dira que la fonction f est intégrable sur $[a, +\infty[$ si la limite

$$\lim_{X \rightarrow +\infty} \int_a^X f(t) dt$$

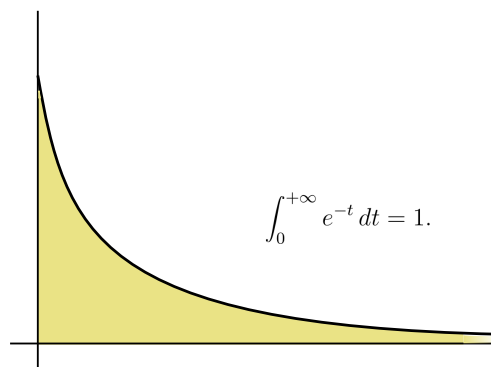
existe (et est finie). Cette limite est alors notée $\int_a^{+\infty} f(t) dt$.

Note : quelque soit $X \geq a$, l'intégrale $\int_a^X f(t) dt$ est de la forme que l'on connaît.

Exemples :

$$f : [0, +\infty[\longrightarrow \mathbb{R} \\ t \longmapsto e^{-t}$$

$$\int_0^X e^{-t} dt = [-e^{-t}]_0^X \\ = 1 - e^{-X} \\ \xrightarrow{X \rightarrow +\infty} 1$$



$$g : [1, +\infty[\longrightarrow \mathbb{R}$$

$$t \longmapsto \frac{1}{t}$$

$$\int_1^X \frac{1}{t} dt = [\ln t]_1^X$$

$$= \ln X$$

$$\xrightarrow{X \rightarrow +\infty} +\infty$$



Puisque la limite n'est pas fini, la fonction g n'est pas intégrable sur $[1, +\infty[$.

Note : on peut évidemment définir de mêmes des intégrales sur des intervalles de la forme $] -\infty, a]$.

2.1.2 Intervalles $]a, b]$

Pour les fonctions définies sur un intervalle de la forme $]a, b]$, on peut de même donner un sens à

$$\int_a^b f(t) dt$$

en termes de limite. Précisément, on notera

$$\int_a^b f(t) dt = \lim_{X \rightarrow a} \int_X^b f(t) dt$$

quand cette limite existe.

Là encore, on peut facilement étendre cette définition aux fonctions définies sur un intervalle de la forme $[a, b[$.

Exemples :

1. Soit $f : t \mapsto \frac{1}{\sqrt{t}}$ définie sur $]0, 1]$. Pour tout $X > 0$, on a

$$\int_X^1 \frac{dt}{\sqrt{t}} = [2\sqrt{t}]_X^1 = 2 - 2\sqrt{X} \xrightarrow{X \rightarrow 0} 2.$$

La fonction f est donc intégrable sur $]0, 1]$ et $\int_0^1 \frac{dt}{\sqrt{t}} = 2$.

2. Soit $g : t \mapsto \frac{1}{t}$ définie sur $]0, 1]$. Pour tout $X > 0$, on a

$$\int_X^1 \frac{1}{t} dt = [\ln t]_X^1 = -\ln X \xrightarrow{X \rightarrow 0} +\infty.$$

La fonction g n'est donc pas intégrable sur $]0, 1]$.

2.1.3 Propriétés

Les définitions ci-dessus permettent donc d'étendre la notion d'intégrale aux intervalles ouverts. On peut alors facilement montrer que ces nouvelles intégrales ont les mêmes propriétés analytiques que celles définies sur des intervalles fermés.

Précisément, on montre facilement que ces nouvelles intégrales sont toujours linéaires et qu'elles vérifient toujours la relation de Chasles. Notons au passage que cette dernière permet d'étendre encore la notion d'intégrales aux intervalles de la forme $]a, b[$ en étudiant séparément les limites

$$\lim_{X \rightarrow a} \int_X^c f(t) dt \quad \text{et} \quad \lim_{Y \rightarrow b} \int_c^Y f(t) dt$$

pour un $c \in]a, b[$ fixé.

Enfin, pour le calcul effectif de ces intégrales, il sera toujours possible d'utiliser les deux outils principaux du calcul intégral : l'intégration par partie et le changement de variable ; en remplaçant tous les calculs sur la borne "ouverte" par des limites.

Cependant, avant d'exploiter toutes ces propriétés pour le calcul effectif, l'étude des intégrales généralisées commence par une étude de convergence.

Or en se basant sur les exemples vus plus haut, il est possible de définir deux grandes classes de fonctions dont on connaît le comportement face à l'intégration sur des domaines ouverts.

2.1.4 Fonctions de référence

Les fonctions exponentielles. Soit

$$E_\lambda : t \mapsto e^{\lambda t}.$$

Ces fonctions sont définies sur \mathbb{R} . Les problèmes d'intégrabilités se posent donc en $\pm\infty$. Or on peut montrer que E_λ est intégrable sur tout intervalle de la forme $[a, +\infty[$ pour $a \in \mathbb{R}$ si et seulement si $\lambda < 0$ (on dit aussi que E_λ est intégrable en $+\infty$).

Les fonctions puissances (ou fonctions de Riemann). Soit

$$R_\alpha : t \mapsto \frac{1}{t^\alpha}.$$

Ces fonctions sont définies pour tout $t \neq 0$ (en particulier si $\alpha > 0$). Les problèmes d'intégrabilités se posent donc en $\pm\infty$ et en 0. Or on peut montrer que

- R_α est intégrable sur tout intervalle de la forme $[a, +\infty[$, $a > 0$ si et seulement si $\alpha > 1$.
- R_α est intégrable sur tout intervalle de la forme $]0, a]$, $a > 0$ si et seulement si $\alpha < 1$.

Note : si $\alpha < 0$, la fonction R_α est définie et continue en 0. Elle est donc intégrable en 0. Autrement dit, le résultat ci-dessus est surtout pertinent quand $0 < \alpha < 1$.

2.1.5 Critères de convergence

Hormis les fonctions de référence que l'on vient de voir, il existe peu de fonctions pour lesquelles on établit l'intégrabilité par le calcul (i.e. à l'aide d'un calcul de limite). Cependant, on va voir qu'à l'aide des outils d'analyse, il est possible de mettre en place des critères de convergence basés sur la comparaison entre les fonctions qui permettent la plupart du temps de comparer la fonction étudiée à l'une des fonctions de référence et d'en déduire le comportement face à l'intégration sur des intervalles ouverts.

Ces critères de comparaison sont dans un premier temps motivés par un argument géométrique : l'étude d'une intégrale généralisée correspond à la mesure d'un domaine ouvert du plan. Or les fonctions de référence vues plus haut produisent un ensemble de domaines ouverts du plan dont on connaît l'aire. Le principe de comparaison se base alors sur les principes (de bon sens) suivants :

- Si le domaine que l'on étudie est contenu dans un domaine d'aire finie alors son aire est finie.
- Si le domaine que l'on étudie contient un domaine d'aire infinie alors son aire est infinie.

Avant de détailler ces critères, notons que toutes les aires étudiées sont des aires *algébriques* dans le sens où les parties de surfaces situées sous l'axe des abscisses sont comptées négativement. En pratique, il peut donc apparaître des phénomènes de compensation. L'étude détaillée de ces compensations n'est jamais aisée, à tel point que sauf mention contraire, on supposera toujours que les fonctions que l'on étudie sont à valeurs positives.

Convergence en l'infini

On a vu plus haut que les fonctions de références qui sont intégrables en l'infini sont les fonctions exponentielles $t \mapsto e^{-\lambda t}$ de paramètre $\lambda > 0$ et les fonction de Riemann $t \mapsto \frac{1}{t^\alpha}$ de paramètre $\alpha > 1$. On constate en particulier que toutes ces fonctions tendent vers 0 en l'infini.

On conçoit aisément que cette dernière propriété est nécessaire à la convergence. Cependant, elle n'est pas suffisante puisqu'il existe des fonctions qui tendent vers 0 en l'infini

et qui ne sont pas intégrables en l'infini (précisément, toutes les fonctions de Riemann de paramètre $\alpha \in]0, 1[$).

Intuitivement, pour qu'une fonction soit intégrable en l'infini, il faut donc qu'elle tende vers 0 *assez vite* en l'infini, et les fonctions de référence permettent de quantifier cette vitesse de convergence.

D'autre part, il existe des outils analytiques permettant de comparer les fonctions que l'on étudie aux fonctions de référence.

Critère de majoration. Un premier critère de convergence est un critère de majoration : soit f une fonction positive définie sur un intervalle de la forme $[a, +\infty[$. S'il existe une fonction g définie sur le même intervalle, dont l'intégrale converge en l'infini et telle que

$$\forall t \geq a, \quad f(t) \leq g(t),$$

alors f est intégrable en $+\infty$ et $\int_a^\infty f(t)dt \leq \int_a^\infty g(t)dt$.

Réciproquement, s'il existe une fonction h dont l'intégrale diverge en l'infini et telle que

$$\forall t \geq a, \quad h(t) \leq f(t),$$

alors l'intégrale de f diverge en l'infini.

Exemples :

1. Soit $f : t \mapsto \frac{|\sin t|}{t^3 + 1}$. Pour tout $t \geq 1$, on a

$$0 \leq f(t) \leq \frac{1}{t^3}.$$

La fonction $g : t \mapsto \frac{1}{t^3}$ étant une fonction de Riemann de paramètre $\alpha = 3 > 1$, elle est intégrable en l'infini. La fonction f l'est donc également et

$$\int_1^\infty \frac{|\sin t|}{t^3 + 1} dt \leq \int_1^\infty \frac{1}{t^3} dt = \frac{1}{2}.$$

2. Soit $f : t \mapsto \frac{2t}{2t^2 - 1}$. Pour tout $t \geq 1$, on a

$$f(t) \geq \frac{1}{t}.$$

La fonction $g : t \mapsto \frac{1}{t}$ étant une fonction de Riemann de paramètre $\alpha = 1 \leq 1$, elle n'est pas intégrable en l'infini. La fonction f ne l'est donc pas non plus.

Critère de domination Il est possible d'étendre le critère vu plus haut en étudiant la limite de quotients du type $\frac{f(t)}{g(t)}$.

Précisément, soit f une fonction positive définie sur un intervalle de la forme $[a, +\infty[$. S'il existe une fonction g dont l'intégrale converge en $+\infty$ et telle que la limite

$$\lim_{t \rightarrow +\infty} \frac{f(t)}{g(t)}$$

soit finie, alors f est dite *dominée* par g et elle est aussi intégrable en $+\infty$.

Réciproquement, si f domine une fonction h dont l'intégrale diverge en l'infini, alors l'intégrale de f diverge également en l'infini.

Exemple : soit $f : t \mapsto \frac{1}{t^2 - 1}$. On ne peut pas majorer f par $g : t \mapsto \frac{1}{t^2}$, mais :

$$\frac{f(t)}{g(t)} = \frac{\frac{1}{t^2-1}}{\frac{1}{t^2}} = \frac{t^2}{t^2-1} \xrightarrow{t \rightarrow +\infty} 1$$

Autrement dit, la fonction f tend vers 0 à la même vitesse que g . Puisque g est intégrable en ∞ , il en est de même pour f .

De façon générale, ce critère de domination permet d'élargir les classes de fonctions de référence :

- Par comparaison aux fonctions de Riemann, on peut montrer que toute fraction rationnelle $f(t) = \frac{P(t)}{Q(t)}$ est intégrable en $+\infty$ si et seulement si on a

$$\deg Q \geq \deg P + 2.$$

- Par comparaison aux fonctions exponentielles, on peut montrer que toute fonction de la forme $f : t \mapsto P(t)e^{-\lambda t}$ (où P est un polynôme) est intégrable en $+\infty$ si $\lambda > 0$.

Note : les critères de comparaison permettent de déterminer l'intégrabilité d'une fonction donnée. Cependant, ils n'apportent aucune information sur la *valeur* des intégrales étudiées.

Convergence en 0

À l'image de ce que l'on a vu en l'infini, il est possible d'établir un critère (partiel) de convergence basé sur le bon sens. Cependant, là où le bon sens donnait une condition nécessaire en l'infini, il nous donne une condition *suffisante* en 0. Précisément, pour qu'une fonction f définie sur un intervalle de la forme $]0, a]$ soit intégrable sur cet intervalle, il suffit que la limite $\lim_{t \rightarrow 0} f(t)$ soit finie (c'est en particulier le cas des fonctions de Riemann

de paramètre $\alpha \leq 0$). Géométriquement, cela correspond en effet à la mesure d'une partie bornée du plan. Cependant, cette condition n'est pas nécessaire puisqu'il existe des fonctions qui tendent vers l'infini en 0 mais dont l'intégrale sur $]0, a]$ est finie (par exemple les fonctions de Riemann de paramètre $\alpha \in]0, 1[$).

De plus, comme pour le cas infini, il est possible de mettre en place des critères de comparaison basés sur la géométrie (majoration) et sur l'étude d'une limite de la forme $\lim_{t \rightarrow 0} \frac{f(t)}{g(t)}$ (domination), les conclusions étant les mêmes que dans le cas infini.

Notons pour finir que la notion d'intégrale généralisée s'applique à tout intervalle de la forme $]a, b]$. Cependant, les fonctions de référence ne sont étudiées qu'en 0 (pour les bornes finies). En pratique, on ramène toute étude sur $]a, b]$ en une étude en 0 en effectuant le changement de variable $t = a + u$ dans les intégrales étudiées.

Exemple : montrer que l'intégrale $\int_1^2 \frac{t}{\sqrt{t-1}} dt$ est convergente.

2.2 Fonctions spéciales

2.2.1 Intégrales à paramètres

Les fonctions spéciales sont des fonctions particulières qui apparaissent dans de nombreux domaines des mathématiques (arithmétique, statistiques, fiabilité, etc). Elles ont la particularité d'être des fonctions définies par des intégrales (à paramètres).

Pour définir une fonction à l'aide d'une intégrale, il suffit en effet d'introduire un paramètre supplémentaire dans la fonction que l'on intègre. Ainsi, si f est une fonction de deux variables t et x , définie en particulier pour $t \in [a, b)$, on peut définir la fonction F par

$$F(x) = \int_a^b f(t, x) dt.$$

Exemple : $F(x) = \int_0^{+\infty} e^{-xt} dt$.

Les outils exposés au paragraphe précédent permettent alors de déterminer le domaine de définition de ce type de fonctions : à quelles conditions sur x la fonction $t \mapsto f(t, x)$ est-elle intégrable sur l'intervalle $[a, b)$?

Les critères de comparaison permettent notamment de répondre à ce type de question sans avoir à calculer l'intégrale (ce que l'on ne sait en général pas faire).

Dans notre exemple, le calcul direct permet ainsi d'établir que F est définie sur $]0, +\infty[$. Par ailleurs, le calcul direct permet également d'obtenir ici une forme explicite pour $F(x)$:

$$\forall x > 0, \quad F(x) = \frac{1}{x}.$$

En règle générale, on ne sait pas obtenir une forme explicite pour les fonctions définies par une intégrale (c'est en particulier le cas pour les fonctions spéciales que l'on verra plus loin). Cependant, les outils de calcul intégral (IPP, changements de variables, etc) permettent d'établir certaines propriétés importantes. On peut notamment montrer que sous certaines conditions sur la fonction f , la fonction F est continue, dérivable, etc et que

$$F'(x) = \int_a^b \frac{\partial f}{\partial x}(t, x) dt.$$

Dans ce chapitre, on va s'intéresser à deux fonctions spéciales : la fonction Beta :

$$\text{Beta}(p, q) = \int_0^1 t^p (1-t)^q dt$$

et la fonction Γ :

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

Ce sont deux fonctions fondamentales en mathématiques dans le sens où elles apparaissent naturellement dans une multitude de problèmes mathématiques. Ce sont, comme prévu deux fonctions définies par des intégrales que l'on ne sait pas calculer (sauf pour quelques valeurs particulières des paramètres). Cependant, l'intérêt que portent les mathématiciens à ces fonctions a permis d'établir de nombreuses propriétés les concernant ; ce qui permet de travailler avec bien que l'on n'ait que très peu de valeurs exactes.

2.2.2 La fonction Γ

Définition

La fonction Γ est la fonction définie par

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$$

Hormis certaines valeurs de x particulières, on ne sait pas calculer $\Gamma(x)$ de façon explicite. Cependant, grâce aux techniques vues plus haut, on peut montrer que l'intégrale $\Gamma(x)$ converge pour tout $x > 0$.

Note : à cause du facteur t^{x-1} (dont la puissance peut être négative), il faut étudier la convergence de $\Gamma(x)$ en 0 et en l'infini.

Parmi les valeurs de $\Gamma(x)$ que l'on peut calculer de façon exacte, il y a en premier lieu $\Gamma(1)$. En effet :

$$\Gamma(1) = \int_0^{+\infty} e^{-t} dt = 1.$$

On verra en T.D. qu'au prix de certains efforts, on peut également montrer que

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

D'autre part, les outils d'analyse évoqués plus hauts permettent de montrer que Γ est dérivable sur son domaine de définition et que

$$\Gamma'(x) = \int_0^{+\infty} \ln t \cdot t^{x-1} e^{-t} dt$$

De façon générale, on peut même montrer que Γ est infiniment dérivable et que

$$\forall k \in \mathbb{N}, \quad \Gamma^{(k)}(x) = \int_0^{+\infty} (\ln t)^k t^{x-1} e^{-t} dt.$$

Cela permet en particulier que la fonction Γ est convexe sur $]0, +\infty[$ et que

$$\lim_{x \rightarrow 0^+} \Gamma(x) = +\infty.$$

L'équation fonctionnelle

Les valeurs $\Gamma(1)$ et $\Gamma\left(\frac{1}{2}\right)$ sont en réalité les seules valeurs de Γ que l'on sait calculer de façon directe. Cependant, à l'aide d'une intégration par parties, on peut montrer que

$$\forall x > 0, \quad \Gamma(x+1) = x\Gamma(x).$$

(Exercice : à faire en dérivant t^x dans $\Gamma(x+1)$).

Cette équation fonctionnelle nous permet dans un premier temps de déterminer la valeur de $\Gamma(n)$ pour tout $n \in \mathbb{N}^*$. En effet, puisque $\Gamma(1) = 1$, d'après l'équation fonctionnelle, on a

$$\begin{aligned} \Gamma(2) &= \Gamma(1+1) = 1 \cdot \Gamma(1) = 1 \\ \Gamma(3) &= \Gamma(2+1) = 2 \cdot \Gamma(2) = 2 \\ \Gamma(4) &= 3 \cdot \Gamma(3) = 6 \\ &\dots \end{aligned}$$

De façon générale, on peut montrer que

$$\forall n \in \mathbb{N}^*, \quad \Gamma(n) = (n-1)!$$

(Exercice : à faire par récurrence).

La fonction Γ peut donc être vue comme une généralisation continue de la notion de factoriel. On constate en particulier que

$$\lim_{x \rightarrow +\infty} \Gamma(x) = +\infty$$

et que cette croissance est extrêmement rapide puisque qu'elle se fait à la vitesse du factoriel.

Un premier champs d'application de la fonction Γ est donc par exemple l'étude des coefficients binomiaux

$$\binom{n}{p} = \frac{n!}{p!(n-p)!} = \frac{\Gamma(n+1)}{\Gamma(p+1)\Gamma(n-p+1)}$$

L'équation fonctionnelle de Γ permet en particulier de démontrer la formule de base de construction du triangle de Pascal :

$$\binom{n+1}{p} = \binom{n}{p} + \binom{n}{p-1}.$$

(Exercice : à faire).

De même, puisque l'on connaît la valeur de $\Gamma(\frac{1}{2})$, on peut déterminer $\Gamma(n + \frac{1}{2})$ pour tout $n \in \mathbb{N}$. On a par exemple

$$\Gamma\left(\frac{3}{2}\right) = \Gamma\left(1 + \frac{1}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

De façon générale, on montre par récurrence que

$$\forall n \in \mathbb{N}, \quad \Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!\sqrt{\pi}}{4^n n!}.$$

En résumé, on connaît une valeur exacte de la fonction Γ pour tout x de la forme $\frac{n}{2}$, $n \in \mathbb{N}^*$.

Prolongement

L'équation fonctionnelle permet de donner un sens à $\Gamma(x)$ pour certaines valeurs de $x < 0$. Précisément, en écrivant l'équation fonctionnelle pour $x - 1$, on obtient

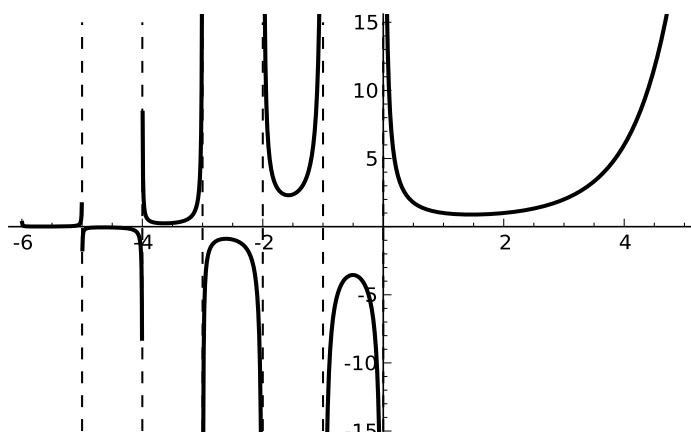
$$\Gamma(x-1) = \frac{\Gamma(x)}{x-1}.$$

Ainsi, pour $x = \frac{1}{2}$, on a

$$\Gamma\left(-\frac{1}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)}{-\frac{1}{2}} = -\sqrt{\pi}.$$

On peut alors montrer que $\Gamma(x)$ pour tout $x < 0$, $x \notin \mathbb{Z}$.

On obtient alors la courbe suivante :



Note : une étude plus poussée de la fonction Γ montre que l'on peut également donner un sens à $\Gamma(z)$ pour $z \in \mathbb{C}$. On peut en particulier montrer que $\Gamma(z)$ est définie pour tout $z \in \mathbb{C}$ hors des valeurs réelles entières négatives ou nulles.

2.2.3 La fonction Beta

La fonction Beta est la fonction de deux variables p et q définie par

$$\text{Beta}(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt.$$

L'étude des intégrales généralisées permet de montrer que Beta est définie pour $p, q > 0$.

Par ailleurs, hormis pour des valeurs entières de p et q (où l'on se retrouve à intégrer des polynômes), il n'existe pas de méthode simple pour calculer $\text{Beta}(p, q)$.

Historiquement, cette fonction est apparue pour la première fois dans les travaux de Wallis lorsqu'il s'intéressait aux intégrales

$$W_n = \int_0^{\frac{\pi}{2}} \sin^n(t) dt.$$

Wallis s'est rapidement rendu compte qu'hormis pour quelques valeurs particulières de n , il n'était pas simple de calculer ces intégrales (pourtant naturelles). À l'aide de différents changements de variables, il a donc déplacé le problème en espérant trouver une forme plus simple pour ces intégrales.

Ainsi, en notant que

$$\sin^n(t) = (1 - \cos^2 t)^{\frac{n-1}{2}} \sin t,$$

le changement de variable $u = \cos t$ dans W_n donne

$$W_n = \int_0^1 (1 - u^2)^{\frac{n-1}{2}} du.$$

Un nouveau changement de variables $t = u^2$ donne

$$\begin{aligned} W_n &= \frac{1}{2} \int_0^1 \frac{(1-u^2)^{\frac{n-1}{2}}}{u} 2u du \\ &= \frac{1}{2} \int_0^1 t^{-\frac{1}{2}} (1-t)^{\frac{n-1}{2}} dt \\ &= \frac{1}{2} \text{Beta} \left(\frac{1}{2}, \frac{n+1}{2} \right) \end{aligned}$$

De nombreux mathématiciens se sont alors penchés sur cette fonction (notamment Newton et Euler).

À l'aide du calcul intégral, on peut montrer de nombreuses propriétés de cette fonction Beta. À l'aide du changement de variable $u = 1 - t$, on peut par exemple montrer qu'elle est symétrique :

$$\text{Beta}(p, q) = \text{Beta}(q, p).$$

D'autre part, l'une des propriétés fondamentales de la fonction Beta est son lien avec la fonction Γ :

$$\text{Beta}(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

Là encore, cette propriété se démontre à l'aide d'un changement de variables

$$\begin{cases} u &= x + y \\ v &= \frac{x}{x+y} \end{cases}$$

dans l'intégrale double

$$\begin{aligned} \Gamma(p)\Gamma(q) &= \left(\int_0^{+\infty} x^{p-1} e^{-x} dx \right) \left(\int_0^{+\infty} y^{q-1} e^{-y} dy \right) \\ &= \iint_{1^{\circ} \text{cadrant}} x^{p-1} y^{q-1} e^{-x-y} dx dy \end{aligned}$$

Ainsi, les éléments du changement de variable sont les suivants :

$$\begin{cases} x &= uv \\ y &= u - uv \end{cases}$$

et

$$\text{Jac}(u, v) = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u$$

D'autre part, si $x \geq 0$ et $y \geq 0$, on a $0 \leq u < +\infty$ et $0 \leq v \leq 1$. D'où

$$\begin{aligned} \Gamma(p)\Gamma(q) &= \iint_{1^{\circ} \text{cadrant}} x^{p-1} y^{q-1} e^{-x-y} dx dy \\ &= \int_0^{+\infty} \int_0^1 (uv)^{p-1} (u-uv)^{q-1} e^{-u} (ududv) \\ &= \left(\int_0^{+\infty} u^{p+q-1} e^{-u} du \right) \left(\int_0^1 v^{p-1} (1-v)^{q-1} dv \right) \\ &= \Gamma(p+q) \text{Beta}(p, q). \end{aligned}$$

Cette formule permet en particulier de calculer les intégrales de Wallis. En effet :

$$W_n = \text{Beta}\left(\frac{1}{2}, \frac{n+1}{2}\right) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2}+1)}$$

Puisque l'on connaît la fonction Γ aux valeurs entières et demi-entières, on obtient

$$\begin{cases} W_{2p} = \frac{(2p)!}{2^{2p}(p!)^2} \frac{\pi}{2} \\ W_{2p+1} = \frac{2^{2p}(p!)^2}{(2p+1)!} \end{cases}$$

Il existe ainsi de nombreuses expressions de la fonction Beta, obtenues via des changements de variables. Une forme célèbre est la forme trigonométrique, obtenue en posant $t = \sin^2 u$:

$$\text{Beta}(p, q) = 2 \int_0^{\frac{\pi}{2}} (\sin u)^{2p-1} (\cos u)^{2q-1} du.$$

Cette forme permet en particulier de retrouver la valeur de $\Gamma(\frac{1}{2})$. En effet, pour $p = q = \frac{1}{2}$, on a

$$\text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) = \pi$$

Or le lien entre Γ et Beta donne

$$\text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(1)} = \Gamma\left(\frac{1}{2}\right)^2$$

$$\text{D'où } \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

La forme trigonométrique de Beta offre en particulier une forme simple pour les valeurs centrales de Beta :

$$\begin{aligned} \text{Beta}(p, p) &= 2 \int_0^{\frac{\pi}{2}} (\sin u \cos u)^{2p-1} du \\ &= \frac{1}{2^{2p-1}} \int_0^{\pi} (\sin t)^{2p-1} dt \end{aligned}$$

Enfin, le changement de variable $u = \frac{t}{1+t}$ dans la forme originale de la fonction Beta donne la forme suivante :

$$\text{Beta}(p, q) = \int_0^{+\infty} \frac{u^{p-1}}{(1+u)^{p+q}} du$$

Cette formule est particulièrement utilisée pour les valeurs complémentaires de la fonction Beta :

$$\forall p \in]0, 1[, \quad \text{Beta}(p, 1-p) = \int_0^{+\infty} \frac{u^{p-1}}{1+u} du.$$

L'étude détaillée de la fonction Beta (ainsi que de la fonction Γ liée au problème par la relation $\text{Beta}(p, 1 - p) = \Gamma(p)\Gamma(1 - p)$) permet en particulier d'établir la formule du complément :

$$\forall p \in]0, 1[, \quad \Gamma(p)\Gamma(1 - p) = \frac{\pi}{\sin(p\pi)}.$$

Pour conclure, les fonctions spéciales Γ et Beta ont occupé de nombreux mathématiciens, tant par le fait qu'elles apparaissent dans de nombreux domaines des mathématiques que par le fait que l'on n'en connaît pas grand chose.