

Cours de Mathématiques

J.R LICOIS

1995-1996

Première partie

ALGEBRE

Chapitre 1

Systèmes linéaires

1.1 Définitions

On note $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \dots \times \mathbb{R}$ l'ensemble des n-uplets de réels (couples pour $n = 2$, triplets pour $n = 3$, quadruplets pour $n = 4, \dots$). Un élément $X \in \mathbb{R}^n$ sera écrit sous la forme $(x_1, \dots, x_n)$ généralement en colonne.

Définition 1 On appelle système linéaire de p équations à n inconnues $x_1, \dots, x_n$ un système de la forme

$$(S) \quad \sum_{j=1}^n a_{ij}x_j = b_i \quad 1 \leq i \leq p$$

où les réels a_{ij} et b_i sont donnés.

Lorsque $n = 2$ ou $n = 3$ on préfère noter x, y, z les inconnues plutôt que x_1, x_2, x_3 .

Exemple :

$$(S_0) \quad \begin{cases} x + 3y - z + 2t = 7 \\ 3x + 4y + 2z + t = 11 \\ 2x + 3y + z + t = 8 \end{cases}$$

Nous allons définir des opérations sur les éléments de $\mathbb{R}^n$.

Définition 2 Soient $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$ deux éléments de $\mathbb{R}^n$, $\lambda \in \mathbb{R}$; on définit la somme $X + Y$ par

$$X + Y = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

et le produit λX par

$$\lambda X = (\lambda x_1, \lambda x_2, \dots, \lambda x_n).$$

Définition 3 Soit A une matrice à n colonnes et p lignes et $X \in \mathbb{R}^n$. On appelle produit de X par la matrice A , noté AX l'élément $Y \in \mathbb{R}^p$ défini par

$$y_i = \sum_{j=1}^n a_{ij}x_j$$

appelé produit lignes par colonnes.

Exemple :

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad AX = \begin{pmatrix} x + 2y + 3z \\ 4x + 5y + 6z \end{pmatrix}$$

En revenant à la définition d'un système linéaire, on voit immédiatement qu'il s'écrit encore sous la forme

$$(S) \quad AX = B$$

avec $X \in \mathbb{R}^n$ l'inconnue, $A = (a_{ij})$ la matrice du système et $B \in \mathbb{R}^p$ le second membre.

Définition 4 Un système $AX = B$ est dit homogène si le second membre est identiquement nul, c'est-à-dire si $b_i = 0 \quad \forall i$. On le note alors sous la forme $AX = \tilde{0}$.

Proposition 1.1.1 Soient $(x_1, \dots, x_n)$ et $(y_1, \dots, y_n)$ deux solutions d'un système $AX = B$. La "différence" $(x_1 - y_1, x_2 - y_2, \dots, x_n - y_n)$ est une solution du système homogène associé $AX = \tilde{0}$.

DÉMONSTRATION : Immédiate en écrivant les relations satisfaites par les x_i et les y_i puis leur différence. $\square$

Corollaire 1.1.2 L'ensemble des solutions d'un système $AX = B$ s'obtient en ajoutant à une solution quelconque de $AX = B$ n'importe quelle solution du système homogène associé $AX = \tilde{0}$.

REMARQUE : Il faut connaître au moins une solution du système "complet" pour appliquer le corollaire : en effet le système $AX = B$ peut très bien ne pas avoir de solutions, alors que le système homogène correspondant en possède une ou plusieurs.

Définition 5 Deux systèmes linéaires à n inconnues sont dits équivalents si et seulement si ils ont le même ensemble de solutions.

1.2 Méthode du pivot

1.2.1 Exemples

La méthode du pivot, introduite par le mathématicien allemand Gauss, est une des nombreuses méthodes permettant de transformer un système linéaire en un système équivalent plus facile à résoudre. La méthode est basée sur la proposition suivante dont la démonstration est immédiate

Proposition 1.2.1 Si on ajoute à une ligne L_k une autre ligne éventuellement multipliée par un coefficient λ , le nouveau système obtenu est équivalent à l'ancien.

Exemple 1 :

Soit le système

$$(S_1) \quad \begin{cases} x + y + 2z = 5 \\ x + 2y + z = 4 \\ 2x + y + z = 3 \end{cases}$$

Effectuons les opérations $L_2 - L_1$ puis $L_3 - 2L_1$, on obtient successivement

$$\begin{cases} x + y + 2z = 5 \\ y - z = -1 \\ 2x + y + z = 3 \end{cases} \quad \begin{cases} x + y + 2z = 5 \\ y - z = -1 \\ -y - 3z = -7 \end{cases}$$

On s'est donc servi du coefficient de x dans la première ligne (pivot) pour faire disparaître la variable x dans les autres relations. Continuons en utilisant le coefficient de y dans la deuxième ligne pour éliminer y dans la troisième relation en remplaçant L_3 par $L_3 + L_2$, on obtient

$$(S'_1) \quad \begin{cases} x + y + 2z = 5 \\ y - z = -1 \\ -4z = -8 \end{cases}$$

Il est facile de résoudre ce dernier système ; la dernière ligne donne $z = 2$, en reportant dans la deuxième on trouve $y = 1$ et en reportant dans la première $x = 0$.

Théorème 1 (Pivot de Gauss) Tout système linéaire $AX = B$ est équivalent à un système $\tilde{A}X = \tilde{B}$ les coefficients $\tilde{a}_{ij}$ vérifiant $\tilde{a}_{ij} = 0$ pour $i > j$.

DÉMONSTRATION : Si $a_{11} \neq 0$, on utilise ce coefficient comme pivot pour éliminer la variable x_1 dans les lignes $L_2, \dots, L_p$ en remplaçant L_i par $L_i - a_{i1}/a_{11}L_1$ pour $2 \leq i \leq p$. Si $a_{11} = 0$, en échangeant la ligne L_1 avec une autre ligne où le coefficient de x est non nul, on se ramène au cas précédent. On recommence ensuite la même opération avec le coefficient a'_{22} de y dans la deuxième ligne du nouveau système, et ainsi de suite.

Si une variable, mettons x_k "disparaît" lors d'une opération, c'est-à-dire que pour le nouveau système le coefficient a_{kk} est nul ainsi que tous les

autres coefficients a_{ik} ($i > k$) situés en dessous, on applique les opérations de pivot à la variable x_{k+1} , ou x_{k+2} , ... quitte à échanger des lignes.

Exemple 2 :

$$(S_2) \quad \begin{cases} x + y + 2z = a \\ 2x + 2y + 5z = b \\ x + 2y + z = c \\ x + 3y + 4z = d \end{cases}$$

Effectuons $L_2 - 2L_1$, $L_3 - L_1$ et $L_4 - L_1$; on obtient

$$\begin{cases} x + y + 2z = a \\ z = b - 2a \\ - z = c - a \\ 2y + 2z = d - 2a \end{cases}$$

On constate que le nouveau coefficient a_{22} est nul; par contre le coefficient a_{32} (le coefficient de y dans la troisième ligne) est non nul : en échangeant les lignes L_2 et L_3 on obtient un système équivalent

$$\begin{cases} x + y + 2z = a \\ - z = c - a \\ z = b - 2a \\ 2y + 2z = d - 2a \end{cases}$$

On peut maintenant éliminer la variable y , en effectuant $L_4 - 2L_2$

$$\begin{cases} x + y + 2z = a \\ - z = c - a \\ z = b - 2a \\ 4z = d - 2c \end{cases}$$

On peut enfin utiliser le coefficient a_{33} pour éliminer la variable z dans la dernière ligne en effectuant $L_4 - 4L_3$.

$$(S'_2) \quad \begin{cases} x + y + 2z = a \\ - z = c - a \\ z = b - 2a \\ 0 = d - 2c - 4b + 8a \end{cases}$$

Exemple 3 :

$$(S_3) \quad \begin{cases} x + y + 2z + 3t = 7 \\ 2x + 2y + 2z + 3t = 9 \\ -x - y + 5t = 3 \end{cases}$$

Effectuons $L_2 - 2L_1$ et $L_3 + L_1$; on obtient

$$\begin{cases} x + y + 2z + 3t = 7 \\ - 2z - 3t = -5 \\ 2z + 8t = 10 \end{cases}$$

La variable y a disparue, on cherche donc à éliminer la variable z en effectuant $L_2 + L_3$, ce qui donne

$$(S'_3) \quad \begin{cases} x + y + 2z + 3t = 7 \\ - 2z - 3t = -5 \\ 5t = 5 \end{cases}$$

1.2.2 Rang d'un système

Définition 6 Soit (S) un système tel que $a_{ij} = 0$ pour $i > j$. On appelle pivot le premier (à partir de la gauche) coefficient non nul, si cela existe, situé sur une ligne. Les inconnues associées aux pivots sont appelées inconnues principales.

En général les pivots sont situés sur la diagonale, sauf si une des inconnues a disparu : quitte à changer l'ordre des inconnues on peut toujours se ramener à des pivots situés sur la diagonale.

Définition 7 On appelle rang du système (S) le nombre de pivots (non nuls!). Nous admettrons provisoirement (c.f Théorème 7) qu'il ne dépend pas de la façon dont on "pivote" le système.

Exemple :

Pour (S_1) , $r = 3$, pivots : $a_{11} = 1$, $a_{22} = -1$, $a_{33} = 4$.

Pour (S_2) , $r = 3$, pivots : $a_{11} = 1$, $a_{22} = 1$, $a_{33} = 1$.

Pour (S_3) , $r = 3$, pivots : $a_{11} = 1$, $a_{23} = -3$, $a_{34} = 5$.

Le système (S_0) du début est équivalent à

$$(S'_0) \quad \begin{cases} x + 3y - z + 2t = 7 \\ -y + z - t = -2 \\ 0 = 0 \end{cases}$$

le rang est donc $r = 2$.

Proposition 1.2.2 Soit (S) un système de p équations à n inconnues de rang r . On a :

$$r \leq n \quad \text{et} \quad r \leq p$$

DÉMONSTRATION : Il ne peut y avoir évidemment plus de pivots que de lignes, ni plus de pivots que d'inconnues (i.e. de colonnes)! $\square$

1.2.3 Solutions

Proposition 1.2.3 Soit un système (S) de p équations à n inconnues de rang r . Il y a seulement 3 situations à envisager, compte-tenu de la propriété précédente :

- $r = n = p$. Le système (S) admet une unique solution. cf (S_1) .
- $r = p$ et $r < n$. Le système (S) possède une infinité de solutions dépendant de $(n - r)$ paramètres. cf (S_0) et (S_3) .
- $r < p$ et $r \leq n$. Le système possède au moins une solution si et seulement si les coefficients b_i du second membre vérifient $(p - r)$ relations de compatibilité. Pour $r = n$ la solution est unique. cf (S_2) .

DÉMONSTRATION : Dans le cas $\mathbf{r} = \mathbf{n} = \mathbf{p}$, il y a exactement un pivot par ligne et par colonne : la dernière relation détermine explicitement la dernière inconnue, et de proche en proche on obtient toutes les autres.

Dans le cas $\mathbf{r} = \mathbf{p}$ et $\mathbf{r} < \mathbf{n}$, il y a un pivot par ligne, mais il y a plus d'inconnues que de pivots. On peut réécrire le système en faisant "passer" au second membre les inconnues $x_{r+1}, \dots, x_n$. Pour chacune des valeurs (arbitraires) assignées aux inconnues $x_{r+1}, \dots, x_n$, dites inconnues auxiliaires, on trouve alors un unique p -uplet $x_1, \dots, x_r$ solution.

Autrement dit : les inconnues principales $x_1, \dots, x_r$ s'expriment comme combinaisons linéaires des inconnues auxiliaires.

Dans le cas $\mathbf{r} < \mathbf{p}$ et $\mathbf{r} \leq \mathbf{n}$, les $p - r$ dernières lignes se présentent sous la forme $0 = ??$. Soit les seconds membres correspondant sont non nuls et le système est impossible (ou incompatible) soit les seconds membres sont nuls : les $p - r$ dernières lignes sont de la forme $0 = 0$ et disparaissent d'elles-mêmes : on dit que le système est compatible. $\square$

1.3 Systèmes Homogènes

REMARQUE : Un système homogène est par définition compatible, on peut donc toujours supposer que $r = n$. En outre si $r = p$ l'unique solution est la solution nulle. Le seul cas à étudier est donc $r = n < p$.

Proposition 1.3.1 Soient $(x_1, \dots, x_n)$ et $(y_1, \dots, y_n)$ deux solutions d'un système homogène $AX = \tilde{0}$, $k \in \mathbb{R}$. Alors le n -uplet $(x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$ est encore une solution de $AX = \tilde{0}$; de même $(kx_1, kx_2, \dots, kx_n)$ est aussi une solution.

DÉMONSTRATION : Immédiate en écrivant les relations satisfaites par les x_i et les y_i et en les ajoutant. $\square$

Cherchons à déterminer la structure de l'ensemble des solutions d'un système homogène.

Exemple :

Considérons le système, déjà mis sous forme triangulaire :

$$\begin{cases} x + y - 2z - 3t = 0 \\ y - z - 2t = 0 \end{cases}$$

On peut encore l'écrire, en ne gardant que les inconnues principales,

$$\begin{cases} x + y = 2z + 3t \\ y = z + 2t \end{cases}$$

ce qui est équivalent à

$$\begin{cases} x = z + t \\ y = z + 2t \end{cases}$$

L'ensemble des solutions dépend donc de deux paramètres, ici notés z et t ; c'est l'ensemble des quadruplets de la forme $(z+t, z+2t, z, t)$. Une solution X peut encore s'écrire

$$X = (z, z, z, 0) + (t, 2t, 0, t) = z(1, 1, 1, 0) + t(1, 2, 0, 1)$$

Remarquons que les deux quadruplets $(1, 1, 1, 0)$ et $(1, 2, 0, 1)$ sont eux-mêmes deux solutions particulières de (S) correspondant aux choix particuliers des paramètres $z = 1, t = 0$ et $z = 0, t = 1$.

Théorème 2 *Soit (S) un système homogène de p équations à n inconnues de rang r ; on suppose $r = n$ et $r < p$. Il existe $p - r$ éléments $Y_{r+1}, \dots, Y_p$ fixés tels que toute solution X de (S) s'écrive sous la forme*

$$X = \lambda_{r+1}Y_{r+1} + \dots + \lambda_pY_p$$

où les λ_i sont des réels dépendant de X .

DÉMONSTRATION : Ecrivons le système vérifié par les inconnues principales, qui est donc un système de r équations, à r inconnues, de rang r : si on fixe arbitrairement les valeurs des inconnues auxiliaires $x_{r+1}, \dots, x_p$ il possède alors une unique solution. Soit Y_{r+1} la solution correspondant à $x_{r+1} = 1, x_{r+2} = 0, \dots, x_p = 0$, et plus généralement, pour $r \leq k \leq p$, soit Y_k la solution correspondant à $x_k = 1$ et $x_j = 0$ pour $r \leq j \leq p$ et $j \neq k$. Soit $X = (x_1, \dots, x_n)$ une solution de (S) donnée et soit $Y = x_{r+1}Y_{r+1} + \dots + x_pY_p$.

Il est clair que, les Y_i étant solutions de (S), Y est aussi une solution de (S). En outre Y est un p -uplet de la forme $(y_1, \dots, y_p)$ avec $y_{r+1} = x_{r+1}, \dots, y_p = x_p$. Compte tenu de l'unicité de la solution de (S) correspondant à une valeur fixée des inconnues auxiliaires, on peut affirmer que $X = Y$. $\square$

Chapitre 2

Espaces vectoriels

2.1 Sous-espaces vectoriels de $\mathbb{R}^n$

2.1.1 Rappels

On a défini dans $\mathbb{R}^n$ l'addition de deux éléments et la multiplication par un réel. L'élément $(0, 0, \dots, 0)$ est appelé vecteur nul et noté $\tilde{0}$. De même $(-1)X$ appelé opposé de X est noté $-X$. On vérifie aisément les propriétés suivantes :

Pour tout $X, Y, Z \in \mathbb{R}^n$, pour tout $\lambda, \mu \in \mathbb{R}$,

- $(X + Y) + Z = X + (Y + Z)$ (associativité)
- $X + Y = Y + X$ (commutativité)
- $X + \tilde{0} = \tilde{0} + X = X$
- $X + -X = -X + X = \tilde{0}$
- $\lambda(X + Y) = \lambda X + \lambda Y$
- $(\lambda + \mu)X = \lambda X + \mu X$
- $\lambda(\mu X) = (\lambda\mu)X$

On vérifie enfin les deux propriétés immédiates :

$$\forall \lambda \in \mathbb{R} \quad \lambda \tilde{0} = \tilde{0} \qquad \forall X \in \mathbb{R}^n \quad 0X = \tilde{0}.$$

2.1.2 Définitions

Définition 8 Une partie E de $\mathbb{R}^n$ qui vérifie

$$\left\{ \begin{array}{lll} \forall X \in E \quad \forall Y \in E & & X + Y \in E \\ \forall X \in E \quad \forall k \in \mathbb{R} & & kX \in E \end{array} \right.$$

est appelée sous-espace vectoriel de $\mathbb{R}^n$.

Exemple :

- L'ensemble $E = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$ est un sous-espace vectoriel de $\mathbb{R}^3$. En effet si on ajoute deux triplets $(x_1, y_1, 0)$ et $(x_2, y_2, 0)$ on obtient bien un élément de E , de même pour la multiplication par un réel.
- L'ensemble $\{(x, y, 1) \mid x, y \in \mathbb{R}\}$ n'est pas un sous-espace vectoriel de $\mathbb{R}^3$. En effet $X = (0, 1, 1)$ et $Y = (1, 0, 1)$ appartiennent à l'ensemble, mais $X + Y = (1, 1, 2)$ n'y appartient pas.
- L'ensemble des solutions d'un système linéaire homogène est un sous-espace vectoriel.
- $\mathbb{R}^n$ est un sous-espace vectoriel.
- L'ensemble $\{\tilde{0}\}$ réduit au vecteur nul est un sous-espace.

REMARQUE : En prenant $k = 0$ dans la deuxième propriété de linéarité, on voit que tout sous-espace contient $\tilde{0}$.

Définition 9 Soient $X_1, \dots, X_k$ des éléments de $\mathbb{R}^n$, $\lambda_1, \dots, \lambda_k$ des réels. La quantité

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k$$

est appelée combinaison linéaire des X_k avec comme coefficients les λ_k .

REMARQUE : Par définition, si E est un sous-espace, il contient n'importe quelle combinaison linéaire formée à partir de vecteurs de E . Réciproquement on a :

Définition 10 Soient $X_1, \dots, X_k$ des éléments de $\mathbb{R}^n$ donnés. L'ensemble des combinaisons linéaires des X_i , les coefficients λ_i prenant toutes les valeurs possibles, est un sous-espace vectoriel de $\mathbb{R}^n$, appelé sous-espace vectoriel engendré par les X_i . On le notera $\text{Vect}(X_1, \dots, X_k)$.

DÉMONSTRATION : Vérifions que le sous-espace engendré est bien un sous-espace. Soient X et Y deux combinaisons linéaires des X_i .

$$\begin{aligned} X &= \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k \\ Y &= \mu_1 X_1 + \mu_2 X_2 + \dots + \mu_k X_k \end{aligned}$$

D'après les propriétés rappelées on a, en regroupant les termes,

$$X + Y = (\lambda_1 + \mu_1)X_1 + (\lambda_2 + \mu_2)X_2 + \dots + (\lambda_k + \mu_k)X_k$$

donc $X + Y$ est une combinaison linéaire des X_i . De même

$$\mu X = \mu\lambda_1 X_1 + \mu\lambda_2 X_2 + \dots + \mu\lambda_k X_k$$

est aussi combinaison linéaire des X_i . $\square$

Proposition 2.1.1 $\text{Vect}(X_1, \dots, X_k)$ est le plus petit (au sens de l'inclusion) sous-espace vectoriel contenant les $X_1, \dots, X_k$.

DÉMONSTRATION : Immédiate car tout sous-espace contient les combinaisons linéaires des X_i . $\square$

Exemple type :

dans $\mathbb{R}^3$ on considère $X_1 = (1, 2, 3)$, $X_2 = (2, 3, 4)$, $X_3 = (3, 4, 5)$. Soit $Z = (6, 6, 6)$; on a immédiatement $Z = -6X_1 + 6X_2$, donc Z est combinaison linéaire de X_1, X_2, X_3 .

Soit $E_1 = \text{Vect}(X_1, X_2, X_3)$. Cherchons à caractériser E_1 . Un élément $Y = (a, b, c)$ appartient à E_1 si et seulement si il existe des réels x, y, z tels que $Y = xX_1 + yX_2 + zX_3$ autrement dit s'il existe au moins une solution au système

$$\begin{cases} x + 2y + 3z = a \\ 2x + 3y + 4z = b \\ 3x + 4y + 5z = c \end{cases}$$

Après transformation on obtient le système équivalent

$$\begin{cases} x + 2y + 3z = a \\ -y - 2z = b - 2a \\ 0 = c - 2b + a \end{cases}$$

Donc $E_1 = \{(a, b, c) \mid a + c - 2b = 0\}$.

2.1.3 Bases

Définition 11 Soient E un sous-espace de $\mathbb{R}^n$, $X_1, \dots, X_k$ des éléments de E . On dira que $(X_1, \dots, X_k)$ est une famille génératrice de E si et seulement si E est exactement égal au sous-espace engendré par les $X_1, \dots, X_k$.

Proposition 2.1.2 Soient E un sous-espace de $\mathbb{R}^n$, $X_1, \dots, X_k$ des éléments de E . $(X_1, \dots, X_k)$ est une famille génératrice de E si et seulement si pour tout $X \in E$ il existe des réels $\lambda_1, \dots, \lambda_k$ (dépendant évidemment de X) tels que

$$X = \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k$$

Exemple type :

(X_1, X_2, X_3) est une partie génératrice de E_1 . En outre $Z = (6, 6, 6)$ peut s'écrire $-6X_1 + 6X_2$ mais aussi $-3X_1 + 3X_3$. Lorsqu'il y a unicité de l'écriture sous forme d'une combinaison linéaire, on a la définition suivante

Définition 12 Soient E un sous-espace de $\mathbb{R}^n$, $X_1, \dots, X_k$ des éléments de E . On dira que $(X_1, \dots, X_k)$ est une base de E , si tout élément X de E s'écrit de manière unique comme combinaison linéaire des X_i : les coefficients de la combinaison linéaire sont appelés coordonnées de X dans la base $(X_1, \dots, X_k)$.

Exemple :

: Dans $\mathbb{R}^2$, $e_1 = (1, 0)$ et $e_2 = (0, 1)$ forment une base, appelée base canonique; en effet tout couple (x, y) s'écrit d'une seule façon

$$(x, y) = (x, 0) + (0, y) = x(1, 0) + y(0, 1) = xe_1 + ye_2$$

De même la base canonique de $\mathbb{R}^3$ est $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ et $e_3 = (0, 0, 1)$. Plus généralement la base canonique de $\mathbb{R}^n$ est formée des e_i avec $e_i = (0, 0, \dots, 1, \dots, 0)$ où seul le terme en i -ème position vaut 1.

Définition 13 Soient E un sous-espace de $\mathbb{R}^n$, $X_1, \dots, X_k$ des éléments de E . On dira que $(X_1, \dots, X_k)$ est une famille liée, s'il existe une combinaison linéaire des $X_1, \dots, X_k$ égale à $(0, \dots, 0)$ avec au moins un des coefficients λ_i non nul.

Exemple :

une famille contenant le vecteur nul $\tilde{0}$ est toujours liée : il suffit de mettre le coefficient du vecteur nul à 1 et les autres à 0.

Exemple type :

La famille X_1, X_2, X_3 est liée car on a

$$X_1 - 2X_2 + X_3 = \tilde{0}$$

Il est clair qu'une famille liée ne peut être une base car le vecteur nul aurait deux écritures.

Définition 14 Une famille qui n'est pas liée est dite libre.

REMARQUE : au lieu de libre (resp. liée) on dit aussi linéairement indépendant (resp. linéairement dépendant).

Proposition 2.1.3 $(X_1, \dots, X_k)$ est une famille libre si et seulement si la seule solution du système

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = \tilde{0}$$

est $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$.

Exemple type :

Soit $X_4 = (0, 0, 1)$. La famille (X_1, X_2, X_4) est libre. En effet écrivons une combinaison linéaire nulle de ces trois vecteurs

$$xX_1 + yX_2 + zX_4 = \tilde{0}$$

ce qui correspond au système

$$\begin{cases} x + 2y & = & 0 \\ 2x + 3y & = & 0 \\ 3x + 4y + z & = & 0 \end{cases}$$

Après transformation, on obtient

$$\begin{cases} x + 2y & = & 0 \\ -y & = & 0 \\ z & = & 0 \end{cases}$$

La seule solution est bien : $x = y = z = 0$.

Proposition 2.1.4 Une famille $(X_1, \dots, X_k)$ est une base si et seulement si elle est libre et génératrice.

DÉMONSTRATION : Si la famille est une base elle est génératrice par définition ; de plus le vecteur nul n'a qu'une seule écriture : la seule solution de $\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = \tilde{0}$ est $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$ ce qui signifie donc que le système est libre.

Réciproquement, la famille étant génératrice, il suffit de vérifier que l'écriture sous forme de combinaison linéaire est unique. Supposons qu'un vecteur X possède deux décompositions :

$$\begin{aligned} X &= \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k \\ X &= \mu_1 X_1 + \mu_2 X_2 + \dots + \mu_k X_k \end{aligned}$$

Par différence, on a alors

$$(\lambda_1 - \mu_1)X_1 + (\lambda_2 - \mu_2)X_2 + \dots + (\lambda_k - \mu_k)X_k = \tilde{0}$$

Comme le système est libre, les coefficients $\lambda_i - \mu_i$, sont tous nuls : on a donc bien unicité de la décomposition.

Exemple type :

Puisque (X_1, X_2, X_4) est libre, il en est de même de (X_1, X_2) . De plus tout vecteur X de E_2 s'écrit comme combinaison linéaire de X_1, X_2, X_3 puisque c'est une partie génératrice.

$$X = xX_1 + yX_2 + zX_3$$

Comme $X_3 = X_1 - 2X_2$, le vecteur X s'écrit encore

$$X = xX_1 + yX_2 + z(X_1 - 2X_2) = (x + z)X_1 + (y - 2z)X_2$$

ce qui montre que (X_1, X_2) est une partie génératrice de E_2 . Conclusion (X_1, X_2) est une base de E_2 .

2.1.4 Dimension

Théorème 3 Toute partie ayant strictement plus de vecteurs qu'une partie génératrice est liée.

DÉMONSTRATION : Soient $(G_1, \dots, G_p)$ une famille génératrice ayant p vecteurs, et n vecteurs $X_1, \dots, X_n$. On peut donc écrire chacun des X_i comme combinaison linéaire des $G_1, \dots, G_p$:

$$\begin{aligned} X_1 &= a_{11}G_1 + a_{12}G_2 + \dots + a_{1p}G_p \\ &\vdots \\ X_n &= a_{n1}G_1 + a_{n2}G_2 + \dots + a_{np}G_p \end{aligned}$$

Soient $\lambda_1, \dots, \lambda_n$ des réels tels que

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_n X_n = \tilde{0}$$

ce qui équivaut après regroupement à

$$(\lambda_1 a_{11} + \lambda_2 a_{21} + \dots + \lambda_n a_{n1})G_1 + \dots + (\lambda_1 a_{1p} + \lambda_2 a_{2p} + \dots + \lambda_n a_{np})G_p = \tilde{0}$$

Pour montrer que $X_1, \dots, X_n$ est liée, il suffit alors de trouver des λ_i , non tous nuls, solutions de

$$\begin{cases} \lambda_1 a_{11} + \lambda_2 a_{21} + \dots + \lambda_n a_{n1} &= 0 \\ &\vdots \\ \lambda_1 a_{1p} + \lambda_2 a_{2p} + \dots + \lambda_n a_{np} &= 0 \end{cases}$$

Le rang r de ce système homogène est tel que $r \leq p$. Par hypothèse $p < n$, le système possède donc une infinité de solutions. On peut trouver une combinaison linéaire nulle des X_i avec des coefficients non tous nuls : le système est lié. $\square$

Corollaire 2.1.5 Tout sous-espace E différent de $\{0\}$ possède (au moins) une base.

DÉMONSTRATION : Admise. $\square$

Corollaire 2.1.6 Toutes les bases de E ont le même nombre de vecteurs.

DÉMONSTRATION : Soient $\mathcal{B}$ et $\mathcal{B}'$ deux bases de E . Comme $\mathcal{B}$ est libre et $\mathcal{B}'$ est génératrice, $\mathcal{B}$ a moins de vecteurs que $\mathcal{B}'$; en échangeant les rôles de $\mathcal{B}$ et $\mathcal{B}'$ on obtient l'égalité.

Définition 15 Soit E un sous-espace vectoriel. On appelle dimension de E , le nombre d'éléments d'une base de E . Par convention $\{0\}$ est de dimension 0.

Exemple :

$\mathbb{R}^n$ est de dimension n , car la base canonique possède n vecteurs.

Exemple type :

E est de dimension 2, car une base est (X_1, X_2) .

Corollaire 2.1.7 Soit E un sous-espace vectoriel de dimension n .

- Toute partie libre formée de n vecteurs est une base.
- Toute partie génératrice formée de n vecteurs est une base.

DÉMONSTRATION : : La démonstration, qui se fait par l'absurde, est fondée sur les deux idées suivantes :

Si une partie libre n'est pas génératrice, on peut lui rajouter (au moins) un élément de façon qu'elle reste libre.

Si une partie génératrice est liée, on peut lui retirer (au moins) un vecteur de façon qu'elle reste génératrice. $\square$

Exemple type :

(X_1, X_2, X_4) est une partie libre formée de 3 vecteurs dans $\mathbb{R}^3$: c'est donc une base de $\mathbb{R}^3$.

Chapitre 3

Applications linéaires

3.1 Définitions

3.1.1 Linéarité

Définition 16 Une application $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est dite linéaire si et seulement si elle vérifie :

$$\begin{cases} \forall X \in \mathbb{R}^n & \forall Y \in \mathbb{R}^n & f(X+Y) &= f(X) + f(Y) \\ \forall X \in \mathbb{R}^n & \forall k \in \mathbb{R} & f(kX) &= kf(X) \end{cases}$$

Exemple :

- $f : (x, y) \mapsto (x^2 + y^2, xy)$ n'est pas linéaire. En effet pour $X = (1, 1)$ et $Y = (0, 1)$ on a

$$f(X) = f(1, 1) = (2, 1) \quad f(Y) = f(0, 1) = (1, 0)$$

- mais $f(X+Y) = f(1, 2) = (5, 2)$ n'est pas égal à $f(X) + f(Y) = (3, 1)$.
- $f(x, y) \mapsto x+y+5$ n'est pas une application linéaire. En effet $f(1, 0) = 6$, $f(0, 1) = 6$ alors que $f(1, 1) = 7$ est différent de $6 + 6 = 12$.
- Si A est une matrice à p lignes et n colonnes, l'application $f : X \mapsto AX$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$.

Exemple type :

L'application $f_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par $f_1(x, y) = (2x+3y, x+4y)$ est linéaire.

DÉMONSTRATION : Soit $X = (x, y)$ et $Y = (x', y')$, avec $X+Y = (x+x', y+y')$

$y')$, d'où

$$\begin{aligned} f_1(X) &= (2x + 3y, x + 4y) \\ f_1(Y) &= (2x' + 3y', x' + 4y') \\ f_1(X) + f_1(Y) &= (2x + 3y + 2x' + 3y', x + 4y + x' + 4y') \\ f_1(X+Y) &= (2(x+x') + 3(y+y'), (x+x') + 4(y+y')) \end{aligned}$$

De même

$$\begin{aligned} f_1(kX) &= f_1(kx, ky) = (2kx + 3ky, kx + 4ky) \\ kf_1(X) &= k(2x + 3y, x + 4y) = (2kx + 3ky, kx + 4ky) \end{aligned}$$

ce qui montre bien que f_1 est linéaire. $\square$

Proposition 3.1.1 L'image d'une combinaison linéaire par une application linéaire est la combinaison linéaire des images (avec les mêmes coefficients).

DÉMONSTRATION : Immédiate. $\square$

Théorème 4 Une application linéaire f est parfaitement déterminée lorsqu'on connaît l'image par f d'une base.

DÉMONSTRATION : Soit $(e_1, \dots, e_n)$ une base de $\mathbb{R}^n$, $X \in \mathbb{R}^n$. On peut décomposer X sur cette base

$$X = \lambda_1 e_1 + \dots + \lambda_n e_n$$

Comme f est linéaire, on a

$$f(X) = \lambda_1 f(e_1) + \dots + \lambda_n f(e_n)$$

on connaît donc $f(X)$ dès qu'on connaît X et les $f(e_i)$. $\square$

Exemple D :

éterminer une application linéaire f de $\mathbb{R}^2$, rapporté à la base canonique, dans $\mathbb{R}^2$ telle que $f(e_1) = f(1, 0) = (2, 1)$ et $f(e_2) = f(0, 1) = (3, 4)$. On a

$$\begin{aligned} f(x, y) &= f(xe_1 + ye_2) = xf(e_1) + yf(e_2) \\ &= x(2, 1) + y(3, 4) = (2x + 3y, x + 4y). \end{aligned}$$

3.1.2 Matrice

Définition 17 Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. On suppose que $\mathbb{R}^n$ est rapporté à une base $\mathcal{B}$ et $\mathbb{R}^p$ rapporté à une base $\mathcal{B}'$. On appelle matrice de f relativement à $\mathcal{B}$ et $\mathcal{B}'$ le tableau à p lignes et n colonnes dont les colonnes représentent les coordonnées dans $\mathcal{B}'$ des images des vecteurs de $\mathcal{B}$.

REMARQUE IMPORTANTE : Lorsque f va de $\mathbb{R}^n$ dans lui-même (i.e $p=n$), on supposera toujours que $\mathcal{B}' = \mathcal{B}$. La matrice de f correspond, si $\mathcal{B} = (e_1, \dots, e_n)$, au tableau suivant :

$$\begin{pmatrix} f(e_1) & \dots & f(e_n) \\ a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{matrix} e_1 \\ \vdots \\ e_n \end{matrix}$$

Exemple type :

On a donc pour f_1 , $f_1(e_1) = f_1(1, 0) = (2, 1)$ et $f_1(e_2) = f_1(0, 1) = (3, 4)$. La matrice de f_1 sera donc

$$A_1 = \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix}$$

Proposition 3.1.2 On suppose que $\mathbb{R}^n$ est rapporté à une base $\mathcal{B}$ et $\mathbb{R}^p$ rapporté à une base $\mathcal{B}'$. Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, de

matrice A relativement à $\mathcal{B}$ et $\mathcal{B}'$. Soit $X \in \mathbb{R}^n$ de coordonnées $(x_1, \dots, x_n)$ dans $\mathcal{B}$, les coordonnées de $f(X)$ dans $\mathcal{B}'$ sont données par le produit

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

DÉMONSTRATION : On a d'après ce qui précède : $f(X) = x_1 f(e_1) + \dots + x_n f(e_n)$. Par définition de la matrice A , on a :

$$f(e_i) = a_{1i}e_1 + a_{2i}e_2 + \dots + a_{ni}e_n \quad 1 \leq i \leq n$$

En regroupant les termes on obtient :

$$f(X) = y_1 e_1 + \dots + y_n e_n$$

avec

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n$$

ce qui correspond exactement au produit de la matrice A par la colonne $(x_1, \dots, x_n)$. $\square$

3.2 Isomorphismes

3.2.1 Rappels

Rappelons les définitions suivantes.

Définition 18 Soit f une application (pas nécessairement linéaire) d'un ensemble E vers un ensemble F .

- f est surjective si et seulement si : $\forall y \in F, \exists x \in E, f(x) = y$. Autrement dit tout élément $y \in F$ possède au moins un antécédent x dans E .
- f est injective si et seulement si deux éléments distincts de E ont des images distinctes dans F , ou encore $f(x) = f(y) \implies x = y$.
- f est bijective si et seulement si elle est injective et surjective : tout élément $y \in F$ possède exactement un antécédent $x \in E$ avec $f(x) = y$.
- Si f est bijective, l'application qui à $y \in F$ associe son antécédent $x \in E$ est appelée bijection réciproque de f , notée f^{-1} .

3.2.2 Matrice Inverse

Proposition 3.2.1 *Si f est linéaire bijective, alors f^{-1} est aussi linéaire.*

DÉMONSTRATION : Laissez en exercice : revenir à la définition de la linéarité et appliquer f à cette relation.

Définition 19 *Si f est linéaire bijective de matrice A , alors la matrice de f^{-1} est appelée matrice inverse de A , notée A^{-1} . En particulier, si X et Y sont deux éléments de $\mathbb{R}^n$ on a l'équivalence*

$$Y = AX \iff X = A^{-1}Y$$

Exemple type :

Soit $X = (x, y)$ et $Y = (a, b)$. On doit donc résoudre le système $Y = AX$ d'inconnues x et y et exprimer ces dernières en fonction de a et b sous forme matricielle.

$$Y = AX \iff \begin{cases} 2x + 3y = a \\ x + 4y = b \end{cases}$$

Après résolution, on obtient

$$\begin{cases} 5x = 4a - 3b \\ 5y = -a + 2b \end{cases} \iff X = \begin{pmatrix} \frac{4}{5} & \frac{-3}{5} \\ \frac{-1}{5} & \frac{2}{5} \end{pmatrix} Y.$$

La matrice inverse A^{-1} est donc

$$A^{-1} = \begin{pmatrix} \frac{4}{5} & \frac{-3}{5} \\ \frac{-1}{5} & \frac{2}{5} \end{pmatrix}.$$

Théorème 5 *Soit f linéaire de matrice A . f est bijective (et A inversible) si et seulement si $\det(A) \neq 0$*

DÉMONSTRATION : Admis (provisoirement) cf. Théorème 6. $\square$

Définition 20 *On appelle matrice identité d'ordre n , notée I_n ou I , la matrice de l'application identique de $\mathbb{R}^n$ dans lui-même. C'est une matrice ayant des 1 sur la diagonale et des 0 partout ailleurs.*

On appelle matrice nulle, notée aussi $\tilde{0}$, la matrice de l'application nulle, dont tous les coefficients sont égaux à 0.

MÉTHODE DU PIVOT Il s'agit d'une méthode permettant de calculer l'inverse d'une matrice. On choisit a priori de travailler sur les lignes ou sur les colonnes de la matrice, mais ce choix est définitif pour toute la suite. On écrit côte à côte la matrice à inverser A et la matrice I ; on effectue exactement les mêmes opérations sur les deux matrices de façon à transformer A en I , la matrice correspondant à I est devenue A^{-1} ; les opérations autorisées sont : on multiplie une ligne (resp. colonne) par un réel non nul, on ajoute à une ligne (resp. colonne) une combinaison linéaire des autres. S'il est impossible avec ces opérations de transformer A en l'identité, en particulier s'il apparaît un zéro sur la diagonale, A n'est pas inversible.

Exemple :

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I$$

Travaillons sur les lignes. $L_2 \leftarrow L_2 - L_1$

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Puis $L_3 \leftarrow L_3 - L_2$

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

Divisons L_3 par 2

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0.5 & -0.5 & 0.5 \end{pmatrix}$$

Effectuons enfin $L_2 \leftarrow L_2 + L_3$ et $L_1 \leftarrow L_1 - L_3$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0.5 & 0.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0.5 \end{pmatrix} = A^{-1}$$

ce qui donne A^{-1} .

3.3 Groupe linéaire

3.3.1 Opérations

Définition 21 Soient f et g deux applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^p$, k un réel. On définit les applications $f + g$ et kf de $\mathbb{R}^n$ dans $\mathbb{R}^p$ par

$$f + g : X \mapsto f(X) + g(X) \quad kf : X \mapsto k f(X)$$

et ces applications sont linéaires.

Définition 22 Soient f et g deux applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^p$ de matrices respectives A et B , k un réel. La matrice de $f + g$ est appelée somme des matrices A et B , notée $A + B$, la matrice de kf est notée kA .

Proposition 3.3.1 Le terme général de $C = A + B$ est donné par $c_{ij} = a_{ij} + b_{ij}$, de même le terme général de $D = kA$ est donné par $d_{ij} = ka_{ij}$.

DÉMONSTRATION : Soit $(e_1, \dots, e_n)$ la base de $\mathbb{R}^n$. Les colonnes de C correspondent aux images des e_i dans la base $(e'_1, \dots, e'_p)$. La jème colonne de C est donnée par

$$\begin{aligned} (f + g)(e_j) &= f(e_j) + g(e_j) \\ &= a_{1j}e'_1 + \dots + a_{pj}e'_p + b_{1j}e'_1 + \dots + b_{pj}e'_p \\ &= (a_{1j} + b_{1j})e'_1 + \dots + (a_{pj} + b_{pj})e'_p \\ &= c_{1j}e'_1 + \dots + c_{pj}e'_p \end{aligned}$$

De même pour les d_{ij} . $\square$

Proposition 3.3.2 Soient g une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^q$, f une application linéaire de $\mathbb{R}^q$ dans $\mathbb{R}^p$. L'application $f \circ g$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$.

DÉMONSTRATION : Calculons simplement

$$\begin{aligned} (f \circ g)(X + Y) &= f(g(X + Y)) = f(g(X) + g(Y)) && g \text{ linéaire} \\ &= f(g(X)) + f(g(Y)) && f \text{ linéaire} \\ &= (f \circ g)(X) + (f \circ g)(Y) \end{aligned}$$

De même

$$(f \circ g)(kX) = f(g(kX)) = f(kg(X)) = kf(g(X)) = k(f \circ g)(X).$$

$\square$

Définition 23 Soient g une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^q$, f une application linéaire de $\mathbb{R}^q$ dans $\mathbb{R}^p$, de matrices respectives B et A dans des bases données. La matrice de $f \circ g$ est appelée matrice produit de A par B , notée AB .

REMARQUE IMPORTANTE Ce produit n'est pas commutatif : en général AB est différent de BA , même pour des matrices carrées. De plus, contrairement à ce qui se passe dans $\mathbb{R}$, on ne peut pas effectuer de simplifications $AB = AC$ n'entraîne pas $B = C$; on peut même trouver facilement des matrices A et B non nulles telles que le produit AB soit égal à la matrice nulle.

On a les propriétés suivantes dont les démonstrations sont immédiates

Proposition 3.3.3

$$\begin{aligned} A(BC) &= (AB)C \\ A(B + C) &= AB + AC \end{aligned}$$

Proposition 3.3.4 Soit A une matrice inversible. La matrice A^{-1} est caractérisée par

$$A A^{-1} = A^{-1} A = I$$

DÉMONSTRATION : Soit f l'application linéaire associée à A , A^{-1} est par définition associée à f^{-1} ; $A A^{-1}$ est associée au composé $f \circ f^{-1}$ qui est l'identité. De même pour $A^{-1} A$ associée à $f^{-1} \circ f = Id$. $\square$

Proposition 3.3.5 Soit $A = (a_{ij})$ et $B = (b_{ij})$. Le terme général c_{ij} de la matrice $C = AB$ est donné par

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{iq}b_{qj}$$

DÉMONSTRATION : Il s'agit donc de déterminer les images des vecteurs $(e_1, \dots, e_n)$ par $f \circ g$. La j ème colonne de C est donnée par $(f \circ g)(e_j) = f(g(e_j))$ qui s'obtient en faisant le produit de la matrice A de f par la colonne $g(e_j)$ qui est exactement la j ème colonne de B . $\square$

DISPOSITION PRATIQUE On place A à gauche, B à droite et au dessus : le terme c_{ij} correspond au produit de la i ème ligne de A et de la j ème colonne de B .

$$A = \begin{pmatrix} \dots & i & \dots \end{pmatrix} \begin{pmatrix} \vdots \\ j \\ \vdots \\ \vdots \\ \dots & c_{ij} \end{pmatrix} = C$$

selon la formule du produit ligne par colonne.

3.3.2 Noyau-Image

Définition 24 Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. On appelle noyau de f , noté $\ker(f)$, l'image réciproque du vecteur nul

$$\ker(f) = \{X \in \mathbb{R}^n \mid f(X) = \tilde{0}\}$$

Proposition 3.3.6 Le noyau de f est un sous-espace de $\mathbb{R}^n$.

DÉMONSTRATION : Soient $X, Y \in \ker(f)$, $\lambda \in \mathbb{R}$. Vérifions que $X + Y$ et λX sont aussi dans $\ker(f)$. On a

$$f(X + Y) = f(X) + f(Y) = \tilde{0} + \tilde{0} = \tilde{0}$$

$$f(\lambda X) = \lambda f(X) = \lambda \tilde{0} = \tilde{0}$$

$\square$

Définition 25 Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. On appelle image de f , noté $\text{Im}(f)$, l'ensemble des images des éléments de $\mathbb{R}^n$

$$\text{Im}(f) = \{Y \in \mathbb{R}^p \mid \exists X \in \mathbb{R}^n \ f(X) = Y\}$$

Proposition 3.3.7 L'image de f est un sous-espace de $\mathbb{R}^p$.

DÉMONSTRATION : Immédiate

Définition 26 On appelle rang d'une application linéaire la dimension du sous-espace vectoriel $\text{Im}(f)$.

Proposition 3.3.8 Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, $\mathcal{B}$ une base de $\mathbb{R}^n$. Les images des éléments de $\mathcal{B}$ forment un système générateur de $\text{Im}(f)$.

DÉMONSTRATION : Soit $Y \in \text{Im}(f)$; il existe $X \in \mathbb{R}^n$ tel que $Y = f(X)$, en décomposant X sur la base $\mathcal{B}$, on obtient

$$\begin{aligned} Y &= f(X) \\ &= f(\lambda_1 e_1 + \dots + \lambda_n e_n) \\ &= \lambda_1 f(e_1) + \dots + \lambda_n f(e_n) \end{aligned}$$

ce qui montre que Y est combinaison linéaire des $f(e_i)$. $\square$

Proposition 3.3.9 Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, $\mathcal{B}$ une base de $\mathbb{R}^n$. On a :

- f est injective si et seulement si l'image de $\mathcal{B}$ est une famille libre
- f est surjective si et seulement si l'image de $\mathcal{B}$ est une famille génératrice de $\mathbb{R}^p$.

DÉMONSTRATION : Laisée en exercice pour la première partie ; la deuxième découle directement de la Proposition 3.3.8. $\square$

Corollaire 3.3.10 Si f linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est bijective, alors $n = p$.

DÉMONSTRATION : Si f est bijective, elle est injective et surjective, l'image de $\mathcal{B}$ est donc libre et génératrice, c'est donc une base comportant n vecteurs comme $\mathcal{B}$, donc $n = p$ car toutes les bases de $\mathbb{R}^p$ ont le même nombre de vecteurs. $\square$

Théorème 6 Soit f une application linéaire de $\mathbb{R}^n$ dans lui-même, de matrice A . Il y a équivalence entre

1. f est injective.
2. f est surjective.
3. f est bijective.
4. $\ker(f) = \{\tilde{0}\}$.
5. $\dim(\text{Im}(f)) = n$.
6. L'image d'une base $\mathcal{B}$ de $\mathbb{R}^n$ est une base $\mathcal{B}'$ de $\mathbb{R}^n$.
7. $\det(A) \neq 0$.

DÉMONSTRATION : On montre facilement que chacune des propriétés implique (6) ; par exemple pour (1) : si f est injective, l'image de $\mathcal{B}$ est un système libre de n vecteurs dans $\mathbb{R}^n$ c'est donc une base. Réciproquement (6) entraîne f bijective (donc les autres propriétés) : soit $Y \in \mathbb{R}^n$ de coordonnées $(\lambda_1, \dots, \lambda_n)$ dans $\mathcal{B}'$, on montre immédiatement qu'il existe un unique X tel que $f(X) = Y$ qui a pour coordonnées $(\lambda_1, \dots, \lambda_n)$ dans $\mathcal{B}$. $\square$

REMARQUE : Ce théorème ne se généralise pas à des espaces vectoriels quelconques, il ne s'applique qu'à des espaces vectoriels de dimension finie.

3.3.3 Rang

Théorème 7 (du rang) Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. On a :

$$\dim(\ker(f)) + \dim(\text{Im}(f)) = n$$

DÉMONSTRATION : Admise. $\square$

Soit A une matrice. On lui associe une application linéaire f , et le système homogène (S) : $AX = \tilde{0}$. Soit r le rang du système. L'ensemble des solutions de (S) est un sous-espace de dimension $n - r$; ce sous espace correspond au noyau de f qui a pour dimension $n - \dim(\text{Im}(f))$: on en déduit immédiatement que r est égal au rang de f .

Conclusion : Le rang d'un système ne dépend que de la matrice du système, et non pas de la façon dont on "pivote" celui-ci.

Définition 27 Soit A une matrice. On appelle rang de A , le rang du système homogène associé $AX = \tilde{0}$ qui est aussi le rang d'une application linéaire f ayant pour matrice A dans une base convenable.

3.3.4 Changement de bases

On suppose données dans $\mathbb{R}^n$ deux bases $\mathcal{B}$ et $\mathcal{B}'$.

Définition 28 On appelle matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$, la matrice P dont les colonnes correspondent aux coordonnées des e'_i exprimées dans la base $\mathcal{B}$.

Proposition 3.3.11 Avec les notations précédentes, la matrice P est inversible et la matrice inverse P^{-1} est la matrice de passage de $\mathcal{B}'$ à $\mathcal{B}$.

DÉMONSTRATION : Admise. $\square$

Théorème 8 Soit f une application linéaire de $\mathbb{R}^n$ dans lui-même. On suppose données dans $\mathbb{R}^n$ deux bases $\mathcal{B}$ et $\mathcal{B}'$. Soient A la matrice de f rapportée à la base $\mathcal{B}$, A' la matrice de f rapportée à la base $\mathcal{B}'$ et P la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$. On a

$$A' = P^{-1}AP$$

DÉMONSTRATION : Admise. $\square$

Exemple S :

oit $\mathbb{R}^2$ rapporté à la base canonique $\mathcal{B} = (e_1, e_2)$ avec $e_1 = (1, 0)$ et $e_2 = (0, 1)$ et à la base $\mathcal{B}' = (e'_1, e'_2)$ avec $e'_1 = (3, 1)$ et $e'_2 = (1, -1)$.

Soit

$$f : (x, y) \mapsto (x + 3y, x - y)$$

Déterminons la matrice de f relativement à $\mathcal{B}$; on doit exprimer $f(e_1)$ et $f(e_2)$ en fonction de e_1 et e_2 .

$$f(e_1) = f(1, 0) = (1, 1) = 1.e_1 + 1.e_2$$

$$f(e_2) = f(0, 1) = (3, -1) = 3.e_1 + (-1).e_2$$

La matrice cherchée est donc

$$A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix}$$

Déterminons la matrice de f relativement à $\mathcal{B}'$; on doit exprimer $f(e'_1)$ et $f(e'_2)$ en fonction de e'_1 et e'_2 .

$$f(e'_1) = f(3, 1) = (6, 2) = 2.e'_1 + 0.e'_2$$

$$f(e'_2) = f(1, -1) = (-2, 2) = 0.e'_1 + (-2).e'_2$$

Il est essentiel d'exprimer les $f(e'_i)$ dans la base $\mathcal{B}'$. La matrice cherchée est donc

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

D'autre part on a :

$$e'_1 = (3, 1) = 3.e_1 + 1.e_2 \quad e'_2 = (1, -1) = 1.e_1 + (-1).e_2$$

la matrice de passage P est donc égale à

$$P = \begin{pmatrix} 3 & 1 \\ 1 & -1 \end{pmatrix}$$

Un calcul direct montre que

$$P^{-1} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{3}{4} \end{pmatrix}$$

Calculons

$$AP = \begin{pmatrix} 6 & -2 \\ 2 & 2 \end{pmatrix} \quad P^{-1}AP = P^{-1} \begin{pmatrix} 6 & -2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

ce qui montre bien que la formule est vérifiée.

REMARQUE : Cette formule présente surtout un intérêt théorique; en pratique, lorsque cela est possible, il vaut mieux exprimer directement les $f(e'_i)$ dans la base $\mathcal{B}'$ que de calculer P , P_1 et $P^{-1}AP$.

Néanmoins, voici un exemple où cette formule prend tout son intérêt. Dans l'exemple précédent cherchons à calculer la matrice A^{11} , c'est-à-dire le produit de la matrice A par elle-même 11 fois. Un calcul direct est possible,

mais fastidieux. Par contre le recours à la matrice D simplifie les choses : en effet

$$\begin{aligned} A^{11} &= A.A \dots A \quad (11 \text{ fois}) \\ &= PDP^{-1}.PDP^{-1} \dots PDP^{-1} \quad (11 \text{ fois}) \end{aligned}$$

Compte tenu des simplifications $PP^{-1} = I$, sachant que $I.D = D.I = D$ on trouve finalement

$$A^{11} = PD^{11}P^{-1}$$

On peut retrouver ceci directement car, par définition du produit matriciel, A^{11} et D^{11} représentent les matrices dans les bases convenables de l'application linéaire $f^{11} = f \circ f \circ \dots \circ f$ (11 fois).

Déterminons alors D^{11} . On peut procéder de deux manières : Une récurrence naturelle montre que

$$D^n = \begin{pmatrix} 2^n & 0 \\ 0 & (-2)^n \end{pmatrix}$$

Sinon on peut calculer déterminer D^{11} comme la matrice associée à f^{11} , obtenue en déterminant $f^{11}(e'_1)$ et $f^{11}(e'_2)$:

$$\begin{aligned} f^{11}(e'_1) &= f(f(\dots f(e'_1))) \\ &= f(f(\dots 2e'_1)) \\ &= 2.2 \dots 2e'_1 \\ &= 2^{11}e'_1 \end{aligned}$$

De même $f^{11}(e'_2) = (-2)^{11}e'_2$.

Finalement

$$A^{11} = \begin{pmatrix} 3 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2048 & 0 \\ 0 & -2048 \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{3}{4} \end{pmatrix} = \begin{pmatrix} 1024 & 3072 \\ 1024 & -1024 \end{pmatrix}$$

Chapitre 4

Déterminants et Applications

4.1 Définitions et Propriétés

Parmi les nombreuses propriétés du déterminant on peut en choisir quelques unes comme définition ou propriétés caractéristiques et en déduire avec plus ou moins de difficultés les autres. Celles utilisées ici sont les définitions de base, c'est-à-dire celles qui mettent en valeur les propriétés fondamentales, mais aussi celles qui exigent des démonstrations plus élaborées.

Soit $\mathcal{M}_n$ l'ensemble des matrices carrées à n lignes et n colonnes. Soit A une matrice appartenant à $\mathcal{M}_n$, on désignera par c_i ou A_i la i ème colonne de A .

4.1.1 Définition

On définit le déterminant comme l'unique application de $\mathcal{M}_n$ dans $\mathbb{R}$:

1. qui est linéaire par rapport aux colonnes.
Soit $A \in \mathcal{M}_n$ de colonnes c_k pour $1 \leq k \leq n$; on suppose que la i ème colonne de A s'écrit $c_i = \lambda c'_i + \mu c''_i$ et soient A' et A'' les matrices obtenues en remplaçant dans A la i ème colonne c_i (et celle-là seulement !) respectivement par c'_i et c''_i . On a alors

$$\det(A) = \lambda \det(A') + \mu \det(A'')$$

2. telle que $\det(M) = 0$ si M possède deux colonnes identiques.
3. vérifiant $\det(I_n) = 1$.

Notation : le déterminant d'une matrice se note aussi en écrivant les coefficients de la matrice entre deux barres.

REMARQUE : Il faut, bien entendu, d'une façon ou d'une autre, justifier l'existence et l'unicité. En particulier les propriétés 1 et 2 définissent le déterminant à une constante multiplicative près, la troisième assure l'unicité.

Proposition 4.1.1 *Le déterminant d'une matrice M :*

- est nul si M possède une colonne identiquement nulle.
- change de signe si on permute deux colonnes adjacentes de M .
- est multiplié par k si on multiplie tous les coefficients d'une colonne (ou d'une ligne) par k .
- est multiplié par k^n si on multiplie la matrice M (i.e. tous les coefficients) par k .

DÉMONSTRATION : Pour la première, on peut prendre $\lambda = \mu = 0$ dans la propriété de linéarité.

Pour la deuxième, c'est un excellent exercice ... Les deux dernières sont immédiates. $\square$

Voyons maintenant la propriété la plus utile pour les calculs pratiques.

Proposition 4.1.2 *$\det(M)$ reste inchangé si on ajoute à une colonne de M une combinaison linéaire des autres colonnes.*

DÉMONSTRATION : Evidente, en combinant les deux propriétés caractéristiques. $\square$

Cette dernière propriété évoque la résolution d'un système linéaire : effectivement les déterminants peuvent aider à la résolution d'un tel système.

Proposition 4.1.3 (Cramer) *Soit le système de n équations à n inconnues*

$$A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

On dira que le système est un système de Cramer¹ si $\det(A) \neq 0$; il admet donc une unique solution, qui est donnée par les "formules de Cramer"

$$x_i = \frac{\det(M_i)}{\det(A)}$$

où M_i est la matrice obtenue en remplaçant dans A la i ème colonne par le second membre.

DÉMONSTRATION : Notons $A_1, A_2, \dots, A_n$ les colonnes de A . On peut encore écrire le système sous la forme

$$x_1 A_1 + x_2 A_2 + \dots + x_n A_n = B$$

On aura donc compte tenu de la linéarité par rapport aux colonnes

$$\det(M_i) = x_1 \det(B_1) + \dots + x_i \det(B_i) + \dots + x_n \det(B_n)$$

où on désigne par B_j la matrice obtenue en remplaçant dans A la i ème colonne par A_j ; pour $j \neq i$, la matrice B_j possède deux colonnes identiques et son déterminant est nul. D'où

$$\det(M_i) = x_i \det(B_i)$$

mais par construction $B_i = A$: on obtient bien la formule annoncée. $\square$

¹**Cramer** Gabriel (1704-1752) Mathématicien suisse.

4.1.2 Propriétés

Définition 29 *Soit $M \in \mathcal{M}_n$ de terme général m_{ij} . On appelle transposée de M , notée tM , la matrice dont le terme général est m_{ji} ; autrement dit les colonnes de tM sont les lignes de M et inversement.*

Proposition 4.1.4 *Pour toute matrice $M \in \mathcal{M}_n$, on a : $\det({}^tM) = \det(M)$*

DÉMONSTRATION : Admise. $\square$

REMARQUE Toute propriété des déterminants vérifiée par les colonnes sera aussi vérifiée par les lignes. Par exemple

Proposition 4.1.5 *On a*

$$\begin{aligned} \det(A) &= \sum_{j=1}^n (-1)^{i+j} a_{ij} D_{ij} & i \text{ fixé} \\ &= \sum_{i=1}^n (-1)^{i+j} a_{ij} D_{ij} & j \text{ fixé} \end{aligned}$$

DÉMONSTRATION : Admise. $\square$

Définition 30 *Soit $A \in \mathcal{M}_n$. Avec les notations précédentes, on appelle comatrice de A , ou encore matrice des cofacteurs, la matrice de terme général*

$$m_{ij} = (-1)^{i+j} D_{ij}$$

notée $\text{com}(A)$

Proposition 4.1.6 *Soit $A \in \mathcal{M}_n$ inversible (donc $\det(A) \neq 0$). On a :*

$$A^{-1} = \frac{1}{\det(A)} {}^t\text{com}(A)$$

DÉMONSTRATION : Effectuons le produit de A par ${}^t\text{com}(A)$: pour les termes situés sur la diagonale ($i = j$), on retrouve exactement la formule de développement de Lagrange : le coefficient obtenu est donc égal à $\det(A)$. Pour les autres termes, on montre que l'expression obtenue correspond à un déterminant ayant deux colonnes identiques : le coefficient est donc nul. $\square$

4.1.3 Déterminant d'une application

Proposition 4.1.7 Soient M et N deux matrices dans $\mathcal{M}_n$. On a

$$\det(MN) = \det(M) \det(N)$$

DÉMONSTRATION : Admise. L'idée est la suivante : on fixe N et on considère l'application $M \mapsto \det(MN)$. On montre qu'elle vérifie les propriétés 1 et 2 de la définition des déterminants, elle est donc proportionnelle à $\det(M)$; d'où la conclusion. $\square$

On retrouve ainsi que le déterminant de M^{-1} est égal à $1/\det(M)$ en écrivant que $M M^{-1} = I$.

Définition 31 Soit f une application linéaire de $\mathbb{R}^n$ dans lui-même de matrice A par rapport à une base $\mathcal{B}$. Le réel $\det(A)$ ne dépend que de f et non de la base choisie pour calculer la matrice. Par définition $\det(f) = \det(A)$.

DÉMONSTRATION : Si on change de base, la nouvelle matrice de f notée A' vérifie d'après la formule de changement de base

$$A' = P A P^{-1}$$

où P est une matrice de passage inversible. On a alors

$$\det(A') = \det(P A P^{-1}) = \det(P) \det(A) \det(P^{-1}) = \det(A)$$

compte tenu de $\det(P^{-1}) = 1/\det(P)$. $\square$

Théorème 9 Soit $M \in \mathcal{M}_n$. On a équivalence entre

$$\det(M) \neq 0 \iff M \text{ inversible} \iff \text{rang}(M) = n$$

DÉMONSTRATION : La dernière équivalence a été justifiée au chapitre précédent.

Si M est inversible, la formule $\det(M) \det(M^{-1}) = \det(I) = 1$ montre que $\det(M)$ est non nul.

Si M n'est pas inversible, son rang est strictement inférieur à n . Une des

colonnes C_k est donc combinaison linéaire des autres : en retranchant à C_k cette combinaison linéaire des autres colonnes, le déterminant reste inchangé, mais la matrice obtenue, qui possède une colonne nulle, a un déterminant nul. $\square$

Corollaire 4.1.8 Soient n vecteurs $X_1, \dots, X_n$ de $\mathbb{R}^n$ rapporté à une base $\mathcal{B}$ et M la matrice dont les colonnes sont les coordonnées des X_i dans $\mathcal{B}$. La famille $(X_1, \dots, X_n)$ est une base de $\mathbb{R}^n$ si et seulement si $\det(M) \neq 0$.

Corollaire 4.1.9 On a équivalence entre

$$\det(f) \neq 0 \iff f \text{ bijective} \iff \ker(f) = \{\tilde{0}\}$$

4.2 Diagonalisation

4.2.1 Éléments propres

Définition 32 Soit f une application linéaire de $\mathbb{R}^n$ dans lui-même de matrice A par rapport à une base $\mathcal{B}$. Un vecteur X non nul est dit vecteur propre pour f (ou pour A) s'il existe un réel λ , appelé valeur propre associée tel que

$$f(X) = \lambda X$$

Réciproquement

Définition 33 Un réel λ est dit valeur propre de f (ou de A) s'il existe un vecteur X non nul tel que $f(X) = \lambda X$; l'ensemble des X tels que $f(X) = \lambda X$ auxquels on rajoute $\tilde{0}$ est appelé sous-espace propre associé à λ .

EXERCICE : Justifier que le sous-espace propre est effectivement un sous-espace vectoriel de $\mathbb{R}^n$.

Proposition 4.2.1 λ est valeur propre de f (ou de A) si et seulement si

$$\det(A - \lambda I) = 0$$

DÉMONSTRATION : Soit X un vecteur propre associé à λ . On a donc les relations équivalentes suivantes

$$\begin{aligned} f(X) &= \lambda X \\ f(X) - \lambda X &= \tilde{0} \\ (f - \lambda Id)X &= \tilde{0} \\ X &\in \ker(f - \lambda Id) \end{aligned}$$

Puisque X est propre, il est non nul. Or le noyau d'une application linéaire est non réduit à $\{\tilde{0}\}$ si et seulement si le déterminant $\det(A - \lambda I)$ est nul. $\square$

REMARQUE : La quantité $\det(A - \lambda I)$ peut être considérée comme un polynôme de degré n dont les coefficients sont déterminés à partir de ceux de A . Ce polynôme est appelé polynôme caractéristique de f (ou de A), noté

$$\lambda \mapsto P_A(\lambda)$$

Les valeurs propres sont donc les racines du polynôme caractéristique.

EXERCICE Montrer à l'aide de la formule de changement de base que le polynôme caractéristique d'une application ne dépend pas de la base choisie pour le calcul.

4.2.2 Diagonalisation

Définition 34 Une matrice $M \in \mathcal{M}_n$ est dite diagonale, si les termes situés en dehors de la diagonale sont nuls ; autrement dit $i \neq j \Rightarrow m_{ij} = 0$

Proposition 4.2.2 Soit f une application linéaire de $\mathbb{R}^n$ dans lui-même de matrice A par rapport à une base $\mathcal{B}'$. La matrice A est diagonale si et seulement si $\mathcal{B}'$ est une base formée de vecteurs propres pour f (ou pour A). Les coefficients situés sur la diagonale sont alors les valeurs propres de f .

DÉMONSTRATION : immédiate puisque les colonnes de A correspondent aux images des vecteurs de $\mathcal{B}'$; on a alors $f(e'_i) = a_{ii}e'_i$. $\square$

Définition 35 Soit f une application linéaire de $\mathbb{R}^n$ dans lui-même de matrice A par rapport à une base $\mathcal{B}$. On dira que f (ou A) est diagonalisable s'il existe une base $\mathcal{B}'$ (éventuellement égale à $\mathcal{B}$) telle que la matrice de f relativement à $\mathcal{B}'$ soit une matrice diagonale D .

DIAGONALISER la matrice A , c'est donc trouver la base $\mathcal{B}'$ et la matrice D ; généralement on détermine $\mathcal{B}'$ en se donnant la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$, c'est-à-dire les coordonnées des vecteurs de $\mathcal{B}'$ dans $\mathcal{B}$. On a alors la relation

$$A = P D P^{-1} \quad \text{équivalente à} \quad D = P^{-1} A P$$

PRATIQUEMENT : On commence par déterminer les valeurs propres comme racines du polynôme caractéristique puis les vecteurs propres associés ; en conclusion on vérifie que ces vecteurs forment une base.

4.2.3 Un exemple

L'espace $\mathbb{R}^4$ est rapporté à la base canonique. Soit f linéaire de matrice A . Le polynôme caractéristique de f est alors

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad P_A(\lambda) = \begin{vmatrix} -\lambda & 1 & 1 & 1 \\ 1 & -\lambda & 1 & 1 \\ 1 & 1 & -\lambda & 1 \\ 1 & 0 & 1 & -\lambda \end{vmatrix}$$

On ajoute toutes les lignes à la première, on met en facteur $\lambda - 3$:

$$P_A = \begin{vmatrix} -\lambda + 3 & -\lambda + 3 & -\lambda + 3 & -\lambda + 3 \\ 1 & -\lambda & 1 & 1 \\ 1 & 1 & -\lambda & 1 \\ 1 & 1 & 1 & -\lambda \end{vmatrix} = (-\lambda + 3) \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -\lambda & 1 & 1 \\ 1 & 1 & -\lambda & 1 \\ 1 & 1 & 1 & -\lambda \end{vmatrix}$$

On retranche la première colonne aux trois autres :

$$P_A(\lambda) = (-\lambda + 3) \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & -\lambda - 1 & 0 & 0 \\ 1 & 0 & -\lambda - 1 & 0 \\ 1 & 0 & 0 & -\lambda - 1 \end{vmatrix}$$

En développant par rapport à la première ligne, puis encore par rapport à la première ligne, on obtient

$$P_A(\lambda) = (-\lambda + 3)(-\lambda - 1)^3$$

Il y a donc deux valeurs propres $\lambda_1 = 3$ et $\lambda_2 = -1$. On vérifie facilement que le sous-espace propre associé à $\lambda_1 = 3$ est la droite vectorielle engendrée par $X_1 = (1, 1, 1, 1)$. Pour $\lambda_2 = -1$ le sous-espace propre correspond à

$$\{(a, b, c, d) | a + b + c + d = 0\}$$

dont on peut prendre pour base $X_2 = (1, -1, 0, 0)$, $X_3 = (1, 0, -1, 0)$ et $X_4 = (1, 0, 0, -1)$.

Pour vérifier que (X_1, X_2, X_3, X_4) est une base de $\mathbb{R}^4$ on calcule le déterminant suivant (c.f. Corollaire 4.1.8) en ajoutant toutes les colonnes à la première :

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{vmatrix} = \begin{vmatrix} 4 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{vmatrix} = -4$$

Comme il est non nul, on peut affirmer que f est diagonalisable.

Deuxième partie

ANALYSE

Chapitre 5

Fonctions d'une variable

5.1 Continuité, Limite

Rappels

Définition 36 Soit une fonction f définie sur une partie $D \subset \mathbb{R}$, $x_0 \in \mathbb{R}$ telle que pour tout réel h assez petit, $D \cap]x_0 - h, x_0 + h[$ est non vide et non réduit à un point. On dira que f admet pour limite ℓ en x_0 si

$$\forall \varepsilon > 0 \quad \exists h > 0 \quad \text{tq} \quad 0 < |x - x_0| < h \implies |f(x) - \ell| < \varepsilon$$

REMARQUE : La condition sur x_0 et D sert à assurer qu'il existe des x assez voisins de x_0 pour lesquels $f(x)$ existe. Elle est par exemple satisfaite si f est définie sur une réunion d'intervalles et si x_0 appartient à l'un de ces intervalles, ou encore si x_0 est une borne d'un de ces intervalles. Le deuxième partie de la définition exprime que $f(x)$ peut être rendu arbitrairement voisin de ℓ pourvu que x soit assez près de x_0 .

Définition 37 On dira que f admet une limite à droite (resp. à gauche) en x_0 si la restriction de f à l'intervalle $]x_0, +\infty[$ (resp. à l'intervalle $]-\infty, x_0[$) admet une limite.

Notations : $\lim_{x \rightarrow x_0^+} = \ell$ (resp. $\lim_{x \rightarrow x_0^-} = \ell$).

Définition 38 (équivalente)

$$\lim_{x \rightarrow x_0^+} = \ell \quad \text{ssi} \quad \forall \varepsilon > 0 \quad \exists h > 0 \quad \text{tq} \quad x_0 < x < x_0 + h \implies |f(x) - \ell| < \varepsilon$$

$$\lim_{x \rightarrow x_0^-} = \ell \quad \text{ssi} \quad \forall \varepsilon > 0 \quad \exists h > 0 \quad \text{tq} \quad x_0 - h < x < x_0 \implies |f(x) - \ell| < \varepsilon.$$

Proposition 5.1.1 Si f admet une limite à droite en x_0 , une limite à gauche en x_0 et si ces limites sont égales, alors f admet une limite en x_0 .

DÉMONSTRATION : Admise $\square$

Définition 39 Soit f définie sur $D \subset \mathbb{R}$, $x_0 \in D$. On dira que f est continue en x_0 si et seulement si

$$\forall \varepsilon > 0 \quad \exists h > 0 \quad \text{tq} \quad x \in]x_0 - h, x_0 + h[\implies |f(x) - \ell| < \varepsilon.$$

REMARQUE : On définirait de manière évidente, continue à droite et continue à gauche.

Proposition 5.1.2 Soit f définie sur un intervalle I , $x_0 \in I$. On a l'équivalence

$$f \text{ continue en } x_0 \iff \lim_{x \rightarrow x_0} f(x) = f(x_0).$$

REMARQUE : Dans le cas où f n'est pas définie sur un intervalle, il y a un problème si x_0 est un point "isolé" du domaine de f ; par exemple $f(x) = \sqrt{x^2(x^2 - 1)}$ est définie sur $]-\infty, -1] \cup \{0\} \cup [1, +\infty[$: avec nos définitions elle ne possède pas de limite en 0, mais on elle est continue en 0.

Définition 40 On dira que f est continue sur l'intervalle $[a, b]$ si f est continue en tout point de l'intervalle $[a, b]$.

Théorème 10 Soit f une fonction continue sur un intervalle fermé $[a, b]$. L'image de $[a, b]$ par f est un intervalle fermé $[m, M]$. En particulier

- f possède un minimum m effectivement atteint : $\exists c_1$ tq $f(c_1) = m$
- f possède un maximum M effectivement atteint : $\exists c_2$ tq $f(c_2) = M$
- f prend toutes les valeurs comprises entre $f(a)$ et $f(b)$.

DÉMONSTRATION : Admise $\square$

5.2 Dérivabilité

Définition 41 Soit f définie sur un intervalle ouvert I , $x_0 \in I$. On dit que f est dérivable en x_0 si la quantité

$$t_{x_0}(h) = \frac{f(x_0 + h) - f(x_0)}{h}$$

admet une limite finie lorsque h tend vers 0. Cette limite, notée $f'(x_0)$, est appelée nombre dérivé de f en x_0 .

Si la quantité $t_{x_0}(h)$ possède une limite à droite en 0 (i.e. lorsque h tend vers 0^+) on dit que f est dérivable à droite en x_0 , la limite étant notée $f'_d(x_0)$. De même pour $f'_g(x_0)$ qui correspond à la limite à gauche (i.e. lorsque h tend vers 0^-).

REMARQUE : Si f est dérivable en x_0 , alors f est continue en x_0 . La réciproque est fautive ; il existe des fonctions continues mais non dérivables et ceci en tout point d'un intervalle.

Théorème 11 (Rolle) Soit f une fonction continue sur l'intervalle $[a, b]$ et dérivable sur l'intervalle $]a, b[$ telle que $f(a) = f(b)$. Il existe un réel $c \in]a, b[$ tel que $f'(c) = 0$.

DÉMONSTRATION : La fonction f est continue sur l'intervalle fermé $[a, b]$, elle possède donc un minimum m et un maximum M qui sont effectivement atteints. Comme $f(a) = f(b)$, au moins l'un des deux est atteint en un point c distinct de a et b . Considérons le taux d'accroissement de f en $x_0 = c$

$$t_c(h) = \frac{f(c + h) - f(c)}{h}$$

Supposons que c soit un minimum, alors le numérateur de $t_c(h)$ est toujours positif (pour h assez petit) : pour $h > 0$, $t_c(h)$ est positif et donc $f'_d(x_0) \geq 0$. De même pour $h < 0$, $f'_g(x_0) \leq 0$. Puisque f est dérivable en c , on a donc, par définition $f'_d(x_0) = f'_g(x_0)$, et ces deux quantités sont donc nulles. Même type de raisonnement si c est le maximum de f . $\square$

REMARQUE : La démonstration précédente montre que si une fonction dérivable f présente un extremum en un point x_0 intérieur à $[a, b]$, la dérivée en ce point est nulle.

Théorème 12 (Accroissements finis) Soit f une fonction continue sur l'intervalle $[a, b]$ et dérivable sur l'intervalle $]a, b[$. Il existe un réel $c \in]a, b[$ tel que

$$f(b) - f(a) = (b - a)f'(c).$$

Cela signifie qu'il existe un point c où la tangente à la courbe représentative de f au point d'abscisse c est parallèle à la corde AB.

DÉMONSTRATION : Considérons la fonction $\varphi : x \mapsto f(x) - \frac{f(b)-f(a)}{b-a}(x-a)$, qui représente la "distance" (mesurée parallèlement à l'axe des ordonnées) entre le point de la courbe et celui de la corde AB. Elle vérifie les hypothèses du théorème de Rolle, en particulier $\varphi(a) = \varphi(b) = 0$; il existe donc un réel c tel que $\varphi'(c) = 0$. Comme $\varphi'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$, on obtient $\frac{f(b)-f(a)}{b-a} = f'(c)$. $\square$

REMARQUE : Cette démonstration permet aussi de justifier le résultat classique selon lequel le bénéfice maximal est obtenu lorsque le profit marginal est égal au coût marginal.

Corollaire 5.2.1 Soit f dérivable sur un intervalle I , si

- $\forall x \in [a, b]$ $f'(x) \geq 0$ (resp. $f'(x) \leq 0$) alors f est croissante (resp. décroissante),
- $\forall x \in [a, b]$ $f'(x) = 0$ alors f est constante.

DÉMONSTRATION : immédiate. $\square$

5.3 Formules de Taylor

5.3.1 Les formules

Proposition 5.3.1 *Soit P un polynôme à coefficients réels de degré n . On a*

$$P(x) = P(0) + xP'(0) + \frac{x^2}{2!}P''(0) + \dots + \frac{x^k}{k!}P^{(k)}(0) + \dots + \frac{x^n}{n!}P^{(n)}(0)$$

où $P^{(k)}$ désigne la dérivée k ème du polynôme P .

DÉMONSTRATION : Rappelons d'abord que la dérivée n ème d'un polynôme de degré n est constante, les dérivées d'ordre $n+1, n+2, \dots$ sont donc identiquement nulles.

Considérons

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

. Pour $x = 0$ on a immédiatement $a_0 = P(0)$. En dérivant une fois on obtient :

$$P'(x) = a_1 + 2a_2x + 3a_3x^2 + \dots + na_nx^{n-1}$$

donc $a_1 = P'(0)$, et ainsi de suite En dérivant k fois on obtient

$$P^{(k)}(x) = k!a_k + \dots + n(n-1)(n-2)\dots(n-k+1)x^{n-k}$$

d'où $a_k = \frac{P^{(k)}(0)}{k!}$. $\square$

On peut étendre ce résultat à une fonction quelconque :

Théorème 13 (Taylor-Mac Laurin) *Soit f une fonction continue et dérivable n fois sur l'intervalle $[0, b]$, on suppose que la dérivée n ème $f^{(n)}$ est continue sur $[0, b]$ et dérivable sur $]0, b[$. Il existe un réel $c \in]a, b[$ tel que*

$$\begin{aligned} f(b) &= f(0) + bf'(0) + \frac{b^2}{2}f''(0) + \dots + \frac{b^k}{k!}f^{(k)}(0) \\ &+ \dots + \frac{b^n}{n!}f^{(n)}(0) + \frac{b^{n+1}}{(n+1)!}f^{(n+1)}(c) \end{aligned}$$

DÉMONSTRATION : Soit

$$P_n(x) = f(0) + xf'(0) + \frac{x^2}{2}f''(0) + \dots + \frac{x^k}{k!}f^{(k)}(0) + \dots + \frac{x^n}{n!}f^{(n)}(0)$$

et

$$\varphi(x) = f(x) - P_n(x) + Cx^{n+1}$$

où la constante C est déterminée par la condition $\varphi(b) = 0$. Par définition, φ et ses n premières dérivées sont nulles en 0. D'après le théorème de Rolle, φ s'annule pour 0 et b , donc φ' s'annule pour une valeur c_1 . On peut appliquer le théorème de Rolle à φ' qui s'annule en 0 et c_1 , il existe donc un réel c_2 tel que $\varphi''(c_2) = 0$. De proche en proche on arrive à montrer que la dérivée n ème s'annule pour 0 et un réel c_n ; en appliquant une dernière fois le théorème de Rolle, il existe un réel c tel que $\varphi^{(n+1)}(c) = 0$. Or

$$\varphi^{(n+1)}(x) = f^{(n+1)}(x) + C.(n+1)!$$

on peut donc affirmer que

$$C = \frac{f^{(n+1)}(c)}{(n+1)!}$$

d'où la formule.

Selon une méthode analogue, avec des hypothèses légèrement différentes, on peut démontrer :

Théorème 14 (Taylor-Young) *Soit f une fonction continue et dérivable n fois sur l'intervalle $[a, b]$, on suppose que la dérivée n ème $f^{(n)}$ est dérivable en a . Il existe une fonction ε telle que $\lim_{x \rightarrow a} \varepsilon(x) = 0$ vérifiant*

$$\begin{aligned} f(x) &= f(a) + (x-a)f'(a) + \dots + \frac{(x-a)^n}{n!}f^{(n)}(a) \\ &+ \frac{(x-a)^{n+1}}{(n+1)!} \left(f^{(n+1)}(a) + \varepsilon(x) \right) \end{aligned}$$

5.3.2 Développements limités

Définition 42 *On dira qu'une fonction f admet un développement limité à l'ordre n (n entier) au voisinage de 0 s'il existe un polynôme P_n de degré*

inférieur ou égal à n , appelé *partie régulière du développement*, et une fonction ε telles que

$$f(x) = P_n(x) + x^n \varepsilon(x) \\ \lim_{x \rightarrow 0} \varepsilon(x) = 0.$$

Il est clair que si une fonction f est dérivable indéfiniment, ses dérivées sont donc bornées (car continues) sur un voisinage de 0, et la formule de Taylor fournit alors le développement limité.

Exemples :

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^k}{k!} + \dots \\ \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + (-1)^{k+1} \frac{x^k}{k!} + \dots \\ (1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} + \frac{\alpha(\alpha-1)(\alpha-2)}{3!} + \dots$$

REMARQUE :

- On peut ajouter ou multiplier des développements limités, plus précisément les parties régulières de ces développements.
- Sous certaines conditions on peut les composer, au sens de la loi $\circ$ de composition des applications.
- On peut intégrer un développement limité, mais il ne faut jamais essayer de le dériver !
- L'existence d'un développement limité à l'ordre $n > 1$ au voisinage 0 pour une fonction f , n'implique pas l'existence des dérivées secondes, troisièmes, $\dots$, en 0.

5.4 Convexité

Définition 43 Une partie $\Omega \subset \mathbb{R}^2$ est dite *convexe* si pour tout couple (A, B) de points de Ω , le segment $[A, B]$ est inclus dans Ω .

Exemples :

- Un disque est convexe

FIG. 5.1 –

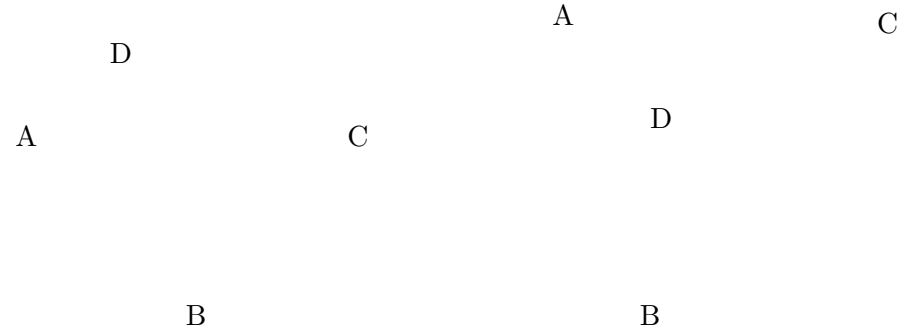
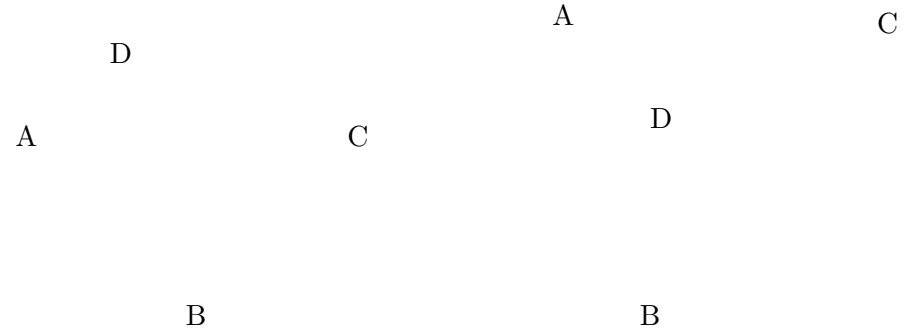


FIG. 5.2 –



- Un demi-plan est convexe
- Toute intersection de convexes est convexe
- Le polygone de la figure 5.1 est convexe.
- Le polygone de la figure 5.2 n'est pas convexe, car le segment $[AC]$ n'est pas inclus dans la figure.

Définition 44 Une fonction f définie sur un intervalle $[a, b]$ est dite *convexe* si l'ensemble

$$\mathcal{E} = \{(x, y) \mid a \leq x \leq b \text{ et } f(x) \leq y\}$$

(appelé *épigraphe* de f) est convexe. cf. figure 5.3.

Une fonction f est dite *concave*, si $-f$ est convexe.

Définition 45 (équivalente) f est convexe si et seulement si, pour tout couple (x_1, x_2) de $[a, b]$ et pour tout réel k avec $0 \leq k \leq 1$, on a :

$$f(kx_1 + (1-k)x_2) \leq kf(x_1) + (1-k)f(x_2)$$

DÉMONSTRATION : cf. figure 5.4. Le réel $kx_1 + (1-k)x_2$ représente un point du segment $[x_1, x_2]$; le réel $kf(x_1) + (1-k)f(x_2)$ représente l'ordonnée du point correspondant sur le segment M_1M_2 qui par définition doit être au dessus du point de la courbe. $\square$

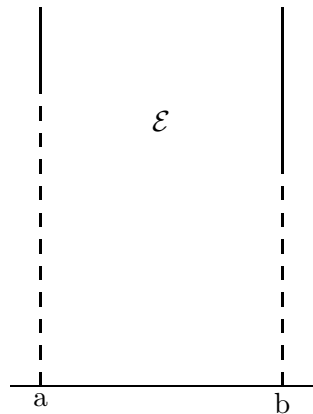


FIG. 5.3 –

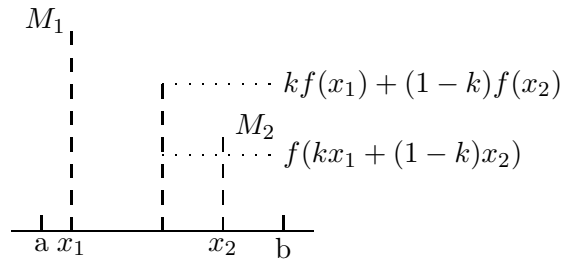


FIG. 5.4 –

Théorème 15 Soit f une fonction convexe sur $[a, b]$. Alors f est continue et admet en tout point de cet intervalle une dérivée à droite et une dérivée à gauche. En outre pour tout couple de points (x_1, x_2) avec $x_1 \leq x_2$ on a

$$f'_g(x_1) \leq f'_d(x_1) \leq f'_g(x_2) \leq f'_d(x_2) \quad (5.1)$$

DÉMONSTRATION : Elle repose sur la remarque suivante : si M_0 est un point fixe du graphique de f et M un point d'abscisse variable x , la fonction $\varphi : x \mapsto \text{pente}(M_0M)$ est une fonction croissante de x . Soit M' un point d'abscisse x' avec $x' < x$, il y a trois cas à envisager :

1. $x' < x_0 < x$ (cf figure 5.5) : Le point M_0 est en dessous du segment $M'M$ par définition de la convexité, donc : $\text{pente}(M'M_0) \leq \text{pente}(M_0M)$ et donc $\varphi(x') \leq \varphi(x)$.
2. $x' < x < x_0$ (cf figure 5.6) : Le point M est nécessairement en dessous de la corde $M'M_0$, on a donc $\text{pente}(M'M_0) \leq \text{pente}(MM_0)$ et donc $\varphi(x') \leq \varphi(x)$.
3. $x_0 < x' < x$ (cf. figure 5.7) : Le point M' est au dessous du segment M_0M par définition de la convexité, donc : $\text{pente}(M_0M') \leq \text{pente}(M_0M)$ et donc $\varphi(x') \leq \varphi(x)$.

Dans tous les cas, on a :

$$x < x' \implies \varphi(x) < \varphi(x')$$

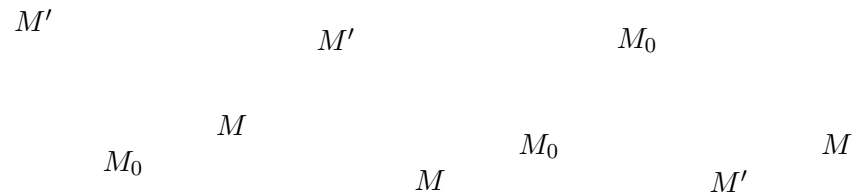


FIG. 5.5 –

FIG. 5.6 –

FIG. 5.7 –

ce qui est la définition de φ croissante.

Considérons maintenant le taux d'accroissement de f en x_0 ,

$$t_{x_0}(h) = \frac{f(x_0 + h) - f(x_0)}{h} = \varphi(x_0 + h)$$

Cette fonction est monotone, bornée sur $]a, b[$, elle admet donc en tout point une limite à droite et une limite à gauche avec $f'_g(x_0) \leq f'_d(x_0)$. On en déduit immédiatement la continuité de f , et la relation (5.1). $\square$

Corollaire 5.4.1 Si f est convexe, dérivable sur $[a, b]$, alors le graphique de f est au dessus de la tangente en x_0 .

DÉMONSTRATION : L'équation de la tangente en x_0 est $y(x) = f'(x_0)(x - x_0) + f(x_0)$. D'après le théorème des accroissements finis, on peut écrire $f(x) = f(x_0) + (x - x_0)f'(c)$ avec c compris entre x et x_0 . Considérons la différence

$$f(x) - y(x) = (x - x_0)(f'(c) - f'(x_0)). \quad (5.2)$$

Comme f' est croissante, $(f'(c) - f'(x_0))$ est du même signe que $c - x_0$ et donc du même signe que $(x - x_0)$; la quantité (5.2) est donc positive, ce qui montre bien que le point d'abscisse x de la courbe (et d'ordonnée $f(x)$) est

au dessus du point correspondant de la tangente. $\square$

Corollaire 5.4.2 *Si f est deux fois dérivable sur $[a, b]$ on a l'équivalence :*

$$f \text{ est convexe} \iff \forall x \in [a, b] \quad f''(x) \geq 0$$

DÉMONSTRATION : Si f est convexe, d'après la relation (5.1), la dérivée f' est une fonction croissante, sa dérivée f'' est donc positive sur $[a, b]$.

Réciproquement, soient x_1 et x_2 deux points de $[a, b]$; sans nuire à la généralité on peut supposer $x_1 < x_2$. Soit $x \in [x_1, x_2]$ qui s'écrit $x = kx_1 + (1 - k)x_2$ avec $0 \leq k \leq 1$. En utilisant le théorème des accroissements finis, on peut écrire :

$$f(x_1) = f(x) + (x_1 - x)f'(c_1) \quad (5.3)$$

$$f(x_2) = f(x) + (x_2 - x)f'(c_2) \quad (5.4)$$

avec $x_1 < c_1 < x < c_2 < x_2$. En multipliant (5.3) par k , (5.4) par $(1 - k)$ et en ajoutant, on obtient :

$$kf(x_1) + (1 - k)f(x_2) = f(x) + k(x_1 - x)f'(c_1) + (1 - k)(x_2 - x)f'(c_2).$$

En remplaçant x par $kx_1 + (1 - k)x_2$, on obtient

$$kf(x_1) + (1 - k)f(x_2) = f(kx_1 + (1 - k)x_2) + k(1 - k)(x_2 - x_1)(f'(c_2) - f'(c_1))$$

Comme f' est croissante, car f'' est positive, ce dernier terme est positif, on a donc

$$f(kx_1 + (1 - k)x_2) \leq kf(x_1) + (1 - k)f(x_2)$$

ce qui est la définition de la convexité. $\square$

Corollaire 5.4.3 *Soit f une fonction convexe sur $[a, b]$. Alors f présente un unique minimum sur $[a, b]$.*

DÉMONSTRATION : Pour simplifier la démonstration, supposons f dérivable. La dérivée f' est monotone sur $[a, b]$. Si f' est de signe constant, alors f est monotone et le minimum de f est atteint en $x = a$ ou $x = b$ selon que f est croissante ou décroissante. Si f' n'est pas de signe constant, elle change de signe sur $]a, b[$: il existe un $c \in]a, b[$ tel que $\forall x \in]a, c[\quad f'(x) \leq 0$ et $\forall x \in]c, b[\quad f'(x) \geq 0$: la fonction f est décroissante puis croissante : elle présente donc un minimum.

Pour l'unicité, supposons qu'il y en ait deux, m_1 et m_2 : la courbe doit être en dessous du segment $[m_1 m_2]$ ce qui contredit le fait que m_1 et m_2 sont des minima. $\square$

REMARQUE : : En particulier, si f est dérivable et si f' s'annule en un point, ce point est le minimum absolu de f . On peut aussi démontrer ce résultat à l'aide du Corollaire (5.4.1).

Exemples :

- Les fonctions $x \mapsto x^n$ pour $n > 1$ sont convexes sur $[0, +\infty[$.
- Les fonctions $x \mapsto a^x$ pour $a > 0$ sont convexes sur $\mathbb{R}$.
- La fonction logarithme népérien est concave sur $]0, +\infty[$.
- La fonction $x \mapsto \sqrt{x}$ est concave sur $[0, +\infty[$.

Chapitre 6

Fonctions de plusieurs variables

6.1 Topologie de $\mathbb{R}^n$

Pour pouvoir parler de limite ou de continuité d'une fonction, il est indispensable de définir une distance sur $\mathbb{R}^n$.

Définition 46 Une distance est une application d de $\mathbb{R}^n$ dans $\mathbb{R}_+$ qui vérifie les propriétés

$$d(x, y) = d(y, x) \quad (6.1)$$

$$d(x, y) = 0 \Leftrightarrow x = y \quad (6.2)$$

$$d(x, z) \leq d(x, y) + d(y, z) \quad (6.3)$$

Les distances usuelles entre deux éléments $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ sont les suivantes :

$$d_1(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$$

$$d_2(x, y) = \sum_{i=1}^n |x_i - y_i|$$

$$d_3(x, y) = \sup |x_i - y_i|$$

La première est la distance euclidienne bien connue, les autres sont plus commodes du point de vue calcul pratique. Le point important est que ces distances sont toutes équivalentes ce qui se traduit par les propriétés suivantes

$$\frac{1}{n}d_2 \leq d_3 \leq d_2$$

$$\frac{1}{\sqrt{n}}d_1 \leq d_3 \leq d_1$$

En clair, dans toute condition du type $\forall \epsilon > 0 \quad d(x, y) < \epsilon$ on peut utiliser indifféremment d_1, d_2, d_3

Définition 47 Soit un élément x de $\mathbb{R}^n$ et un réel ϵ , la boule ouverte $B(x, \epsilon)$ de centre x et de rayon ϵ est l'ensemble des y tels que $d(x, y) < \epsilon$

$$B(x, \epsilon) = \{y \in \mathbb{R}^n / d(x, y) < \epsilon\}$$

A titre d'exercice, on peut voir que si $n = 2$, les boules pour la distance d_1 sont des disques, celles pour d_2 sont des losanges et celles pour d_3 sont des carrés. L'équivalence des distances se traduit par le fait que chaque boule pour une distance contient des boules pour une des autres distances (de rayon éventuellement différent).

Définition 48 On dira qu'une partie Ω de $\mathbb{R}^n$ est ouverte si et seulement si pour chaque point x de Ω il existe un $\epsilon > 0$ tel que la boule $B(x, \epsilon)$ soit incluse dans Ω .

Définition 49 On dira qu'un point x_0 est adhérent à une partie Ω si n'importe quelle boule de centre x_0 rencontre Ω :

$$\forall \epsilon > 0 \quad B(x_0, \epsilon) \cap \Omega \neq \emptyset$$

Il est évident que si x appartient à Ω alors il est adhérent à Ω , puisque $B(x_0, \epsilon) \cap \Omega$ contient au moins le point x_0 , mais il peut exister des points adhérents à Ω qui ne lui appartiennent pas. Par exemple, lorsque $n = 2$, soit $\Omega = \{(x, y)/x^2 + y^2 < 2\}$ le point $(1, 1)$ est un point adhérent qui n'appartient pas à Ω .

Définition 50 On dira qu'une partie $F \subset \mathbb{R}^n$ est fermée si et seulement si son complémentaire est une partie ouverte.

Proposition 6.1.1 F est fermée si et seulement si elle contient tous ses points adhérents.

Démonstration : laissée en exercice.

Exemples : une boule ouverte est ouverte. La boule fermée $\bar{B}(x, \epsilon) = \{y/d(x, y) \leq \epsilon\}$ est fermée. Dans $\mathbb{R}^2$ le produit cartésien de deux intervalles fermés $[a; b] \times [c; d]$, appelé pavé fermé, est un fermé, le demi-plan ouvert $\{(x, y)/ax + by < c\}$ est un ouvert, le demi-plan fermé $\{(x, y)/ax + by \geq c\}$ est un fermé.

Proposition 6.1.2 L'intersection de deux fermés est un fermé, la réunion de deux fermés est un fermé.

Définition 51 La norme associée à la distance d est l'application de $\mathbb{R}^n$ dans $\mathbb{R}_+$ définie par $\|x\| = d(x, 0)$.

On a alors les propriétés caractéristiques d'une norme

$$\|x\| = 0 \Leftrightarrow x = 0 \quad (6.4)$$

$$\|\lambda x\| = |\lambda| \|x\| \quad \forall \lambda \in \mathbb{R} \quad (6.5)$$

$$\|x + y\| \leq \|x\| + \|y\| \quad (6.6)$$

qui correspondent à celles déjà connues lorsque $n = 2$. Réciproquement on peut définir la distance à partir de la norme :

Proposition 6.1.3 On a $d(x, y) = \|x - y\|$.

En effet, à partir des définitions de d_1 , d_2 et d_3 on a : $d(x, y) = d(x - y, 0)$, donc $d(x, y) = \|x - y\|$.

Définition 52 On dira d'une partie Ω de $\mathbb{R}^n$ qu'elle est bornée s'il existe un réel positif M tel que pour tout x de Ω on ait $\|x\| \leq M$.

6.2 Continuité Limite

6.2.1 Définitions

Soit f une fonction définie sur une partie $\Omega \subset \mathbb{R}^n$ à valeurs dans $\mathbb{R}^p$; à chaque n -uplet $x = (x_1, \dots, x_n)$ on associe un élément $y = f(x)$ de $\mathbb{R}^p$ que l'on peut définir par ses coordonnées $(y_1, \dots, y_p)$ dans une base $\mathcal{B}$.

Définition 53 Les p fonctions numériques $f_i : x \mapsto y_i$ ainsi définies sur Ω sont appelées fonctions coordonnées.

Dans toute la suite on supposera que $\mathcal{B}$ est la base canonique de $\mathbb{R}^p$, on a donc

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix}$$

Dans la suite, nous essaierons de traduire les propriétés d'une fonction f à l'aide des propriétés de ses fonctions coordonnées. On peut aussi associer à une fonction f , la fonction *numérique* définie par $x \mapsto \|f(x)\|$ appelée norme de f et notée $\|f\|$.

6.2.2 Limite

Intuitivement, on dira qu'une fonction f admet une limite ℓ lorsque x tend vers x_0 si les valeurs prises par f peuvent être rendues aussi *proches* que l'on veut de ℓ pourvu que x soit assez *proche* de x_0 . La notion de *proximité* se mesurant à l'aide d'une des distances ci-dessus ; d'autre part pour que la propriété soit significative il faut qu'il existe des points de Ω aussi *proches* que possible de x_0 : une condition nécessaire est donc que x_0 soit adhérent à Ω .

Définition 54 Soit f définie sur Ω à valeurs dans $\mathbb{R}^p$, x_0 un point adhérent à Ω , on dira que f admet pour limite ℓ lorsque x tend vers x_0 si

$$\forall \epsilon > 0 \quad \exists \eta > 0 \quad \text{tq } d(x, x_0) < \eta \Rightarrow d(f(x), \ell) < \epsilon$$

Compte tenu des remarques précédentes, on peut utiliser dans $\mathbb{R}^n$ comme dans $\mathbb{R}^p$ n'importe laquelle des distances ci-dessus. Comme conséquences immédiates de la définition on a :

Proposition 6.2.1 1. f tend vers 0 si et seulement si $\|f\|$ tend vers 0.
 2. f tend vers ℓ si et seulement si $f - \ell$ tend vers 0.

Posons $\ell = (\ell_1, \dots, \ell_p)$ et notons f_i les fonctions coordonnées de f , il est facile de démontrer :

Proposition 6.2.2 f tend vers ℓ lorsque x tend vers x_0 si et seulement si on a pour chaque fonction coordonnée $\lim_{x \rightarrow x_0} f_i = \ell_i$.

Exemples : Soient f et g définies sur $\mathbb{R}^2 \setminus (0,0)$ définies par

$$f(x, y) = \frac{x + y}{x^2 + y^2} \quad g(x, y) = \frac{xy}{\sqrt{x^2 + y^2}}$$

La fonction f n'a pas de limite en $(0,0)$, g a pour limite 0. En effet, dans toute boule centrée en $(0,0)$, il existe des points de la forme $(x, -x)$ où f est nulle, et des points de la forme $(x, 0)$ où f vaut $1/|x|$ et n'a donc pas de limite lorsque x tend vers 0. Pour g , on a

$$|g(x, y) - 0| \leq \frac{|x||y|}{\sqrt{x^2 + y^2}} \leq |x| \frac{|y|}{\sqrt{x^2 + y^2}} \leq |x|$$

car $|y| \leq \sqrt{x^2 + y^2}$. De même $|g(x, y) - 0| \leq |x|$. La définition de la limite est donc satisfaite avec $\eta = \epsilon$ pour la distance d_3 puisque

$$d_3(x, y) < \epsilon \Rightarrow |x| \leq \epsilon \text{ et } |y| \leq \epsilon \Rightarrow |g(x, y) - 0| \leq \epsilon$$

Les définitions concernant la continuité se déduisent de celles concernant la limite selon le même procédé que dans le cas d'une variable.

Définition 55 Soit f définie sur $\Omega \subset \mathbb{R}^n$ et $x_0 \in \Omega$. f est continue en x_0 si et seulement si on a $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

Proposition 6.2.3 f est continue en x_0 si et seulement si toutes les fonctions coordonnées sont continues en x_0 .

Démonstration : conséquence immédiate de la proposition 6.2.2 et de la définition précédente.

On démontre et nous admettrons les résultats suivants

Proposition 6.2.4 1. Si f et g sont continues en x_0 alors $f+g$ est continue en x_0 .

2. Si f et g sont continues en x_0 avec f à valeurs réelles, alors $f.g$ est continue en x_0 .

3. Si f est à valeurs réelles, continue en x_0 , avec $f(x_0) \neq 0$ alors $\frac{1}{f}$ est continue en x_0 .

4. Si f est continue en x_0 et g continue en $f(x_0)$ alors $g \circ f$ est continue en x_0 .

5. Les fonctions projections $p_i : (x_1, \dots, x_n) \mapsto x_i$ sont continues.

Une conséquence immédiate de ces propriétés est que les fonctions polynômes et rationnelles à plusieurs variables, et leurs composés avec des fonctions numériques continues sont continues partout où elles sont définies.

Exercice : Montrons qu'une application f linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est continue. Par définition, $f(x) - f(x_0) = f(x - x_0)$, il suffit donc de montrer la continuité en 0. Soit $(a_{i,j})$ la matrice de f ; posons $M = \sup |a_{i,j}|$. Les fonctions coordonnées de f correspondent alors aux vecteurs lignes de la matrice de f . En prenant la distance d_3 on a immédiatement

$$\begin{aligned} \|f(x)\| &= \sup_{1 \leq j \leq p} |f_j(x)| \\ &= \sup_j \left| \sum_{i=1}^n a_{j,i} x_i \right| \\ &\leq n.M. \sup_i |x_i| \\ &\leq n.M. \|x\| \end{aligned}$$

Il suffit alors dans la définition de prendre $\eta = \epsilon/n.M$ pour conclure.

Enfin signalons une propriété importante qui généralise le théorème sur l'image d'un intervalle fermé par une fonction continue :

Proposition 6.2.5 Si une fonction numérique f est continue en tous les points d'une partie K fermée et bornée de $\mathbb{R}^n$ alors f est bornée et atteint ses bornes.

Démonstration : admise (trop délicate).

6.3 Différentiabilité

6.3.1 Définition

Nous disposons maintenant de tous les outils pour étendre la définition de la différentiabilité d'une fonction numérique à une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^p$.

Définition 56 Soit f définie sur une partie Ω et x appartenant à Ω . On dira que f est différentiable en x s'il existe une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, notée Df_x , appelée différentielle de f en x , et une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^p$ notée ϵ telle que

$$\begin{cases} f(x+h) &= f(x) + Df_x(h) + \epsilon(h)\|h\| \\ \lim_{h \rightarrow 0} \epsilon(h) &= 0 \end{cases}$$

On montre que s'il existe un couple (Df_x, ϵ) répondant à ces conditions, il est nécessairement unique.

Remarques : L'accroissement h est un élément de $\mathbb{R}^n$ sous la forme $h = (h_1, \dots, h_n)$.

Df_x est une application qui peut être différente pour chaque valeur de x .

Lorsqu'on repère les espaces de départ et d'arrivée à l'aide de bases convenables, la différentielle en x possède alors une matrice.

Exemple : Soit l'application $f : (x, y) \mapsto x^3 - x^2y + xy^2$. Evaluons $Df_{(1,2)}$. En notant $h = h_1$ et $k = h_2$, on a

$$\begin{aligned} f(1+h, 2+k) &= (1+h)^3 - (1+h)^2(2+k) + (1+h)(2+k)^2 \\ &= 1 + 3h + 3h^2 + h^3 - (1 + 2h + h^2)(2+k) + \\ &\quad (1+h)(4 + 4k + k^2) \\ &= 3 + 3h + 3k + (2hk + h^2 + k^2 + hk^2 - kh^2 + h^3) \end{aligned}$$

Il est facile de voir que le terme entre parenthèses est négligeable, par identification on a donc $Df_{(1,2)}(h, k) = 3h + 3k$.

Définition 57 La matrice de l'application linéaire Df_x est appelée matrice Jacobienne¹ de f en x .

Nous retrouvons les propriétés auxquelles on peut s'attendre avec les fonctions coordonnées :

Proposition 6.3.1 f est différentiable en x si et seulement si les fonctions coordonnées f_i sont différentiables en x . De plus on a :

$$Df_x(h) = \begin{pmatrix} Df_{1x}(h) \\ \vdots \\ Df_{px}(h) \end{pmatrix}$$

Démonstration : utiliser dans l'espace d'arrivée la distance d_3 .

Ceci montre que si on sait calculer la différentielle d'une application de $\mathbb{R}^n$ dans $\mathbb{R}$ on peut alors généraliser à une fonction quelconque dont on connaît les fonctions coordonnées.

Ceci nous amène à considérer les dérivées partielles d'une fonction numérique. On suppose l'espace de départ $\mathbb{R}^n$ rapporté à une base $(e_1, \dots, e_n)$ qui sera pour nous la base canonique.

Définition 58 f admet une dérivée partielle selon x_i (ou dans la direction e_i) au point $x = (x_1, \dots, x_n)$ si la fonction numérique ϕ_i

$$\phi_i(k) = \frac{f(x_1, \dots, x_i + k, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{k}$$

possède une limite lorsque k tend vers 0. Cette limite est notée $\frac{\partial f}{\partial x_i}(x)$ ou $f'_i(x)$

Proposition 6.3.2 Si f est différentiable en x , alors elle admet en ce point des dérivées partielles

Démonstration : Dans la définition prenons un accroissement h de la forme

$$h = (0, \dots, k, \dots, 0) = k(0, \dots, 1, \dots, 0) = k.e_i$$

où k figure à la i ème place, e_i désignant le i ème vecteur de la base canonique. On a alors

$$f(x+h) = f(x) + Df_x(h) + \|h\|\epsilon(h)$$

¹JACOBI Carl Gustav : Mathématicien allemand 1804-1851

mais, par linéarité, on a $Df_x(h) = kDf_x(e_i)$ ce qui montre alors que

$$\phi_i(k) = Df_x(e_i) + \epsilon(h)$$

possède une limite lorsque k , et donc h tendent vers 0 : f admet une dérivée partielle $\frac{\partial f}{\partial x_i}$ qui est égale à $Df_x(e_i)$ ce qui permet de trouver ainsi la matrice de Df relativement à la base canonique de $\mathbb{R}^n$: c'est une matrice à une ligne et n colonnes dont les termes sont les $\frac{\partial f}{\partial x_i}(x)$.

Proposition 6.3.3 *Si f admet des dérivées partielles continues en x , alors f est différentiable en x*

Démonstration : Faisons-la dans le cas où $n = 2$. On se place en (x, y) et on suppose que $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ existent au voisinage de (x, y) et sont continues en (x, y) .

Ecrivons

$$f(x+h, y+k) - f(x, y) = f(x+h, y+k) - f(x+h, y) + f(x+h, y) - f(x, y)$$

Compte tenu de l'existence des dérivées partielles on a

$$f(x+h, y+k) - f(x+h, y) = k \cdot \frac{\partial f}{\partial y}(x+h, y) + k\epsilon(k) \quad (6.7)$$

$$f(x+h, y) - f(x, y) = h \cdot \frac{\partial f}{\partial x}(x, y) + h\epsilon_1(h) \quad (6.8)$$

$$(6.9)$$

Donc

$$f(x+h, y+k) = f(x, y) + h \cdot \frac{\partial f}{\partial x}(x, y) + k \cdot \frac{\partial f}{\partial y}(x, y) + h\epsilon(h) + k\epsilon(k)$$

Compte tenu de la continuité des dérivées partielles, on peut écrire

$$\frac{\partial f}{\partial y}(x+h, y) = \frac{\partial f}{\partial y}(x, y) + \epsilon_2$$

où ϵ_2 tend vers 0 lorsque h tend vers 0. Si on pose $Df : (h, k) \mapsto h \cdot \frac{\partial f}{\partial x}(x, y) + k \cdot \frac{\partial f}{\partial y}(x, y)$ on a alors

$$f(x+h, y+k) = f(x, y) + Df(h, k) + \epsilon_3(h, k)$$

où ϵ_3 , qui regroupe tous les ϵ précédents, vérifie bien

$$\lim_{(h,k) \rightarrow (0,0)} \frac{\epsilon_3(h, k)}{\|(h, k)\|} = 0$$

si on choisit par exemple la norme associée à d_3 .

Remarquons qu'il n'existe pas de condition équivalente à la différentiabilité exprimée en terme de dérivée partielle, on a seulement une condition suffisante. Nous avons donc démontré :

Proposition 6.3.4 *Avec les notations ci-dessus $h = (h_1, \dots, h_n)$ on a en désignant par f'_i les dérivées partielles de f en x par rapport à la i ème variable,*

$$Df_x(h) = f'_1(x) \cdot h_1 + \dots + f'_n(x) \cdot h_n$$

Remarquons que dans ces conditions la matrice jacobienne de f est la matrice à 1 ligne et n colonnes : $(f'_1(x), \dots, f'_n(x))$. Compte tenu de la Proposition 6.3.1, dans le cas d'une application de $\mathbb{R}^n$ dans $\mathbb{R}^p$, la matrice jacobienne est celle dont les lignes sont constituées des dérivées partielles des fonctions coordonnées, c'est-à-dire la matrice

$$\left(\frac{\partial f_i}{\partial x_j} \right)_{1 \leq i \leq p ; 1 \leq j \leq n}$$

Proposition 6.3.5 *Une application linéaire L de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est différentiable en tout point x et sa différentielle Df_x est égale à elle-même.*

Démonstration : revenir à la définition 56 et prendre $\epsilon \equiv 0$.

Proposition 6.3.6 *Si f et g sont différentiables en x , alors $f+g$ est différentiable en x avec*

$$D(f+g)_x = Df_x + Dg_x$$

Démonstration : On écrit

$$\begin{aligned} (f+g)(x+h) &= f(x+h) + g(x+h) \\ &= f(x) + Df_x(h) + \epsilon(h) + g(x) + Dg_x(h) + \epsilon_1(h) \\ &= (f+g)(x) + (Df_x + Dg_x)(h) + \epsilon_2(h) \end{aligned}$$

et la somme de deux applications linéaires étant linéaire, la conclusion est immédiate.

Le résultat le plus important, et le plus utile, de tout ce paragraphe est le théorème ayant trait à la différentielle d'un composé d'applications

Théorème 16 *Soit f différentiable en x et g différentiable en $f(x)$, alors $g \circ f$ est différentiable en x avec*

$$D(g \circ f)_x = (Dg_{f(x)})_o Df_x$$

Démonstration : En utilisant la différentiabilité de f , on a :

$$g \circ f(x + h) = g(f(x + h)) = g(f(x) + Df_x h + \epsilon(h))$$

Considérant $k = Df_x h + \epsilon(h)$ comme un accroissement à partir de $f(x)$ on obtient en utilisant la différentiabilité de g :

$$g \circ f(x + h) = g \circ f(x) + Dg_{f(x)}(k) + \epsilon_1(k)$$

On rappelle que le composé de deux applications linéaires est une application linéaire dont la matrice est le produit des matrices correspondantes. On a alors :

$$g \circ f(x + h) = g \circ f(x) + Dg_{f(x)}(Df_x h) + Dg_{f(x)}(\epsilon(h)) + \epsilon_1(k)$$

Il suffit alors pour conclure de montrer que les deux derniers termes tendent vers 0, pour cela on utilise la continuité de l'application linéaire Df .

Exercice : Soit f une application différentiable de $\mathbb{R}^2$ dans $\mathbb{R}$, u et v deux fonctions numériques dérivables ; quelle est la dérivée de la fonction numérique $\phi : t \mapsto f(u(t), v(t))$? Soit la fonction $\Psi : \mathbb{R} \rightarrow \mathbb{R}^2$ définie par

$$t \mapsto \Psi(t) = (u(t), v(t))$$

dont les fonctions coordonnées sont alors u et v . La matrice jacobienne de Ψ est alors

$$\begin{pmatrix} u'(t) \\ v'(t) \end{pmatrix}$$

De même la matrice jacobienne de f est la matrice

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

Comme $\phi = f \circ \Psi$, la matrice jacobienne de ϕ qui correspond à ϕ' est le produit des matrices jacobiniennes, ce qui donne :

$$\phi'(t) = \frac{\partial f}{\partial x}(u(t)).u'(t) + \frac{\partial f}{\partial y}(v(t)).v'(t)$$

6.4 Dérivées successives

6.4.1 Formule de Taylor

Dans tout ce qui suit on ne considère que des fonctions numériques. On rappelle :

Théorème 17 (Schwarz) *Si f possède des dérivées partielles secondes continues en x_0 alors on a :*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Démonstration : admise

Nous allons définir la notion de différentielle seconde d'une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, c'est à dire définir la différentielle de Df . Pour x fixé, Df_x est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$; on peut identifier cette application avec sa matrice jacobienne, et considérer que la correspondance $x \mapsto Df_x$ est à valeur dans $\mathbb{R}^n$ et que ses fonctions coordonnées sont les n fonctions $x \mapsto f'_i(x)$. La différentielle de cette nouvelle application sera appelée *différentielle seconde* de f , notée $D^2 f_x$, et l'on peut considérer qu'elle est représentée par sa matrice jacobienne (qui sera appelée *matrice Hessienne* de f) :

$$\text{Hess} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{1 \leq i, j \leq n}$$

puisque les dérivées partielles des fonctions coordonnées de Df sont les dérivées partielles secondes de f .

Si cela vous semble peu rigoureux, dites-vous que la difficulté pour dériver Df_x par rapport à x réside dans le fait que Df_x n'est pas un élément de $\mathbb{R}^n$.

En pratique on dispose de la formule suivante

Théorème 18 (Taylor)

$$f(x+h) = f(x) + Df_x(h) + \frac{1}{2!} {}^t h \cdot \text{Hess}.h + \epsilon(h) \|h\|^2$$

avec $\lim_{h \rightarrow 0} \epsilon(h) = 0$

Remarquez le carré autour de $\|h\|$!

Dans la formule précédente, h est un vecteur de $\mathbb{R}^n$ normalement écrit en colonne, H une matrice carrée $n \times n$, dont le produit avec h donne un autre vecteur colonne ; enfin le produit avec ${}^t h$ vecteur transposé de h (qui est donc formé d'une ligne et de n colonnes) finit bien par donner un réel !

Exemple : Vérifions la formule avec la fonction polynômiale f

$$f(x, y) = x^3 + 4x^2y + xy^2$$

On se place au point $(1, 2)$. On a alors

$$\frac{\partial f}{\partial x}(1, 2) = 23 \quad \frac{\partial f}{\partial y}(1, 2) = 8$$

$$\frac{\partial^2 f}{\partial x^2}(1, 2) = 22 \quad \frac{\partial^2 f}{\partial y^2}(1, 2) = 2 \quad \frac{\partial^2 f}{\partial x \partial y}(1, 2) = 12$$

Si l'on pose $h = (u, v)$, on a alors

$$Df_{(1,2)}(u, v) = 23u + 8v$$

$${}^t h \cdot \text{Hess}.h = (u, v) \begin{pmatrix} 22 & 12 \\ 12 & 2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \quad (6.10)$$

$$= (u, v) \begin{pmatrix} 22u + 12v \\ 12u + 2v \end{pmatrix} \quad (6.11)$$

$$= 22u^2 + 24uv + 2v^2 \quad (6.12)$$

On a alors d'après la formule de Taylor

$$f(1+u, 2+v) = 9 + 23u + 8v + 11u^2 + 12uv + v^2 + \epsilon(u, v)$$

Un calcul direct montre que

$$f(1+u, 2+v) = 9 + 23u + 8v + 11u^2 + 12uv + v^2 + u^3 + 4u^2v + 4uv^2$$

Par identification $\epsilon(u, v) = u^3 + 4u^2v + 4uv^2$; on a immédiatement

$$|\epsilon(u, v)| \leq (1 + 4 + 4) \text{Sup}(|u|, |v|)$$

et si on pose $\|h\| = \text{Sup}(|u|, |v|)$ on voit que

$$\frac{\epsilon(h)}{\|h\|^2} \leq 9 \|h\|$$

et la quantité tend bien vers 0 lorsque h tend vers 0.

Chapitre 7

Optimisation

7.1 Recherche des extrema d'une fonction f

Le même argument que dans le cas d'une variable, montre qu'une condition nécessaire pour que f présente un extrema en x_0 est $Df_{x_0} \equiv 0$ ce qui veut dire que le point est stationnaire. Si le point est stationnaire, la nature de celui-ci dépend du signe de la quantité ${}^t h \cdot \text{Hess}.h$

- S'il est strictement positif *pour tout h assez petit* : le point est un minimum.
- S'il est strictement négatif *pour tout h assez petit* : le point est un maximum
- S'il est de signe variable : le point est un point selle (ou point col)
- Sinon on ne peut pas conclure.

On démontre que la matrice Hessienne d'une application est toujours une matrice diagonalisable et les conditions ci-dessus s'interprètent encore de la façon suivante : si toutes les valeurs propres de la matrice Hessienne sont strictement positives le point considéré est un minimum, si toutes les valeurs propres sont strictement négatives le point considéré est un maximum.

Dans le cas de deux variables, introduisons les notations de Monge

$$r = \frac{\partial^2 f}{\partial x^2} \quad t = \frac{\partial^2 f}{\partial y^2} \quad s = \frac{\partial^2 f}{\partial x \partial y}$$

La matrice hessienne est alors

$$H = \begin{pmatrix} r & s \\ s & t \end{pmatrix}$$

et la quantité considérée s'écrit alors en posant $h = (u, v)$

$$R(u, v) = r.u^2 + 2.s.u.v + t.v^2$$

Supposons v non nul, on peut écrire

$$R(u, v) = v^2 \cdot \left\{ r \cdot \frac{u^2}{v^2} + 2s \cdot \frac{u}{v} + t \right\}$$

et le signe de l'expression est donné par le signe du trinôme entre crochets. Si le discriminant réduit $s^2 - rt$ est négatif, le trinôme est de signe constant : il y a extremum et la nature de celui-ci est donnée par le signe de r. Pour r positif c'est un minimum, pour r négatif c'est un maximum. Si le discriminant est positif, le trinôme change de signe et le point stationnaire est un point selle. On vérifiera qu'on obtient les mêmes résultats en supposant u non nul.

Dans le cas général, soit $H = (a_{i,j})_{1 \leq i,j \leq n}$ la matrice hessienne de f en x ; on note D_k le déterminant de la matrice extraite des k premières lignes et des k premières colonnes de H

$$D_k = \det(a_{i,j})_{1 \leq i,j \leq k} \quad 1 \leq k \leq n$$

Proposition 7.1.1 Si les D_k sont tous strictement positifs, x est un minimum.

Si les D_k sont de signe alternés avec $D_1 < 0$, x est un maximum.

On obtient le même résultat en considérant les k dernières lignes et colonnes.

La justification, délicate, de ce résultat vient du fait que D_k est le produit des k premières valeurs propres de H.

7.2 Optimisation Sans Contrainte

Soit f une fonction définie sur un ouvert $\Omega \in \mathbb{R}^n$. Le fait que Ω soit ouvert est important car il signifie que l'extremum sera obtenu en un point "intérieur", et non pas sur le bord. Dans le cas où on cherche aussi des extrema sur la frontière, il faut utiliser des contraintes d'inégalités.

Proposition 7.2.1 *On suppose que f est pourvue de dérivées partielles continues sur $\Omega \in \mathbb{R}^n$. Si f admet un extremum en x_0 alors $Df_{x_0} \equiv 0$. Autrement dit le point x_0 est un point stationnaire.*

DÉMONSTRATION : admise.

La propriété précédente est bien sûr équivalente à dire que toutes les dérivées partielles de f en x_0 sont nulles. Cette condition n'est pas suffisante : il existe des points stationnaires qui ne sont pas des minima ni des maxima, ce sont des points selles. Pour déterminer la nature d'un point x_0 stationnaire, deux méthodes sont possibles :

On étudie directement le signe de $f(x) - f(x_0)$: c'est la méthode à conseiller pour un mathématicien.

On étudie la matrice Hessienne au point x_0 : c'est la méthode généralement utilisée en économie. Elle est résumée ci-dessous.

Soit f pourvue de dérivées partielles secondes continues, x_0 un point **stationnaire** pour f , H la matrice hessienne de f en x_0 .

$$H = (h_{ij})_{1 \leq i, j \leq n} \quad \text{avec} \quad h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

On pose $D_k = \det(h_{ij})_{1 \leq i, j \leq k}$ pour $1 \leq k \leq n$.

Proposition 7.2.2 *Si tous les D_k sont strictement positifs, x_0 est un minimum.*

Si $D_1, D_3, D_5, \dots$ sont strictement négatifs et $D_2, D_4, D_6, \dots$ sont strictement positifs, alors x_0 est un maximum.

Si un des D_k est nul, on ne peut pas conclure

Dans tous les autres cas, x_0 est un point selle.

DÉMONSTRATION : admise

REMARQUE : On aura exactement le même résultat en considérant les déterminants D'_k emboîtés à partir du bas,

$$D'_k = \det(h_{ij})_{n-k+1 \leq i, j \leq n}.$$

La condition donnée par la Proposition 7.2.2 est suffisante, mais n'est pas équivalente : sur l'exemple suivant les D_k ne permettent pas de conclure, alors que les D'_k montrent que l'on a affaire à un point selle. **Exemple :**

Soit $f(x, y, z) = \frac{1}{2}(x^2 + y^2 + z^2) + xy - 2yz$. On a

$$\frac{\partial f}{\partial x} = x + y \quad \frac{\partial f}{\partial y} = y + x - 2z \quad \frac{\partial f}{\partial z} = z - 2y$$

ce qui donne comme point stationnaire $x_0 = (0, 0, 0)$. La matrice hessienne en x_0 est

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

On a $D_2 = 0$, par contre $D'_1 = 1$, $D'_2 = -3$ et $D'_3 = -4$.

Chapitre 8

Optimisation avec contraintes

8.1 Fonctions implicites

L'idée exprimée par la théorie des fonctions implicites est la suivante : si plusieurs quantités réelles $x_1, \dots, x_n$ sont liées par une relation de la forme $g(x_1, \dots, x_n) = 0$, il est en général possible d'exprimer n'importe laquelle des quantités x_i comme fonction des autres quantités. Par exemple, la relation $x^2 + y^2 = 1$ permet d'exprimer y en fonction de x par $y = \sqrt{1 - x^2}$. La relation $x^2 + y^2 = 1$ définit implicitement y en fonction de x , l'expression explicite étant $y = \sqrt{1 - x^2}$. Dans la majorité des cas, on ne connaît pas l'expression explicite (on ne sait pas l'exprimer à l'aide de fonctions usuelles). L'intérêt du théorème suivant, est qu'il permet de déterminer les dérivées de cette fonction implicite, même si on ne sait pas l'exprimer explicitement.

Théorème 19 Soit f une fonction numérique définie sur un ouvert $\Omega \in \mathbb{R}^n$, $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)$ un point de Ω tel que $f(\hat{x}_1, \dots, \hat{x}_n) = 0$. On suppose que f et ses dérivées partielles sont continues au voisinage de ce point.

Si $\frac{\partial f}{\partial \hat{x}_n}(\hat{x}_1, \dots, \hat{x}_n) \neq 0$ alors il existe une fonction de $n-1$ variables $x_1, \dots, x_{n-1}$, ϕ définie pour x assez proche de $\hat{x}$ telle que

$$f(x_1, \dots, x_n) = 0 \iff x_n = \phi(x_1, \dots, x_{n-1})$$

De plus ϕ est différentiable avec

$$\frac{\partial \phi}{\partial x_k}(\hat{x}_1, \dots, \hat{x}_{n-1}) = - \frac{\frac{\partial f}{\partial x_k}(\hat{x}_1, \dots, \hat{x}_n)}{\frac{\partial f}{\partial \hat{x}_n}(\hat{x}_1, \dots, \hat{x}_n)} \quad 1 \leq k \leq n-1$$

Bien entendu, la propriété se généralise à n'importe laquelle des x_i qui peut s'exprimer en fonction des autres, à condition que la dérivée partielle correspondante soit non nulle au point considéré.

DÉMONSTRATION : On admettra l'existence et la différentiabilité de ϕ . Il est facile de justifier le dernier résultat ; supposons $n = 3$, la fonction $f(x, y, \phi(x, y))$ est identiquement nulle, ainsi que sa différentielle ; on a donc

$$\left(\frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \frac{\partial f}{\partial z} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

ce qui donne bien

$$\begin{cases} \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z} \frac{\partial \phi}{\partial x} = 0 \\ \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z} \frac{\partial \phi}{\partial y} = 0 \end{cases}$$

On peut généraliser le théorème à un nombre quelconque de relations. Par exemple, soient f et g deux fonctions différentiables, et considérons les

(x, y, z) satisfaisant le système

$$\begin{cases} f(x, y, z) = 0 \\ g(x, y, z) = 0 \end{cases}$$

Il existe alors, sous la condition

$$\frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x} \neq 0$$

deux fonctions ϕ et ψ telles que le système ci-dessus soit équivalent à

$$x = \phi(z) \quad y = \psi(z)$$

Exercice Enoncer correctement les hypothèses correspondantes et vérifier que

$$\frac{\partial \phi}{\partial z} = \frac{\frac{\partial f}{\partial y} \frac{\partial g}{\partial z} - \frac{\partial f}{\partial z} \frac{\partial g}{\partial y}}{\frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x}} \quad \frac{\partial \psi}{\partial z} = \frac{\frac{\partial f}{\partial z} \frac{\partial g}{\partial x} - \frac{\partial f}{\partial x} \frac{\partial g}{\partial z}}{\frac{\partial f}{\partial x} \frac{\partial g}{\partial y} - \frac{\partial f}{\partial y} \frac{\partial g}{\partial x}}$$

8.2 Extrema avec contraintes d'égalité

Le problème est de chercher le maximum (ou le minimum) d'une fonction f de n variables $x_1, \dots, x_n$ soumises à une contrainte, c'est-à-dire satisfaisant une relation de la forme $g(x_1, \dots, x_n) = 0$. Sauf cas exceptionnel, lorsque f admet un maximum en x_0 , ce point ne vérifie pas la contrainte $g(x_0) = 0$. Il faut donc aborder le problème différemment.

Chaque fois que cela est possible, la méthode la plus efficace consiste à exprimer explicitement une des quantités en fonction des $n - 1$ autres, à reporter dans f : on est ainsi ramené à un problème avec $n - 1$ variables sans contrainte. Si cela est impossible, ou trop compliqué à mettre en œuvre, il faut utiliser le Lagrangien. Pour une fonction f de deux variables, la méthode de substitution ramène le problème à l'étude d'une fonction d'une variable, le Lagrangien est une fonction de trois variables...

8.2.1 Un exemple théorique

Considérons une fonction f de deux variables que l'on cherche à optimiser sous la contrainte $g(x, y) = 0$. En supposant que toutes les quantités écrites ont un sens, soit (a, b) un extremum de f sous la contrainte $g(x, y) = 0$; supposons que le théorème des fonctions implicites s'applique : il existe une fonction ϕ telle que : $g(x, y) = 0 \Leftrightarrow y = \phi(x)$.

La fonction $F : x \mapsto f(x, \phi(x))$ présente donc un extremum pour $x = a$ et sa dérivée en ce point est donc nulle :

$$F'(a) = \frac{\partial f}{\partial x}(a, b) + \frac{\partial f}{\partial y}(a, b)\phi'(a) = 0$$

Or

$$\phi'(a) = -\frac{\partial g}{\partial x}(a, b) / \frac{\partial g}{\partial y}(a, b)$$

ce qui donne finalement

$$\frac{\frac{\partial f}{\partial x}}{\frac{\partial g}{\partial x}} = \frac{\frac{\partial f}{\partial y}}{\frac{\partial g}{\partial y}} = \lambda \quad (8.1)$$

si on appelle λ la valeur commune des deux rapports. Donc si on pose $L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$, le point (a, b, λ) est stationnaire pour L puisque

$$\begin{aligned} \frac{\partial L}{\partial x}(a, b, \lambda) &= \frac{\partial f}{\partial x}(a, b) - \lambda \frac{\partial g}{\partial x}(a, b) = 0 \\ \frac{\partial L}{\partial y}(a, b, \lambda) &= \frac{\partial f}{\partial y}(a, b) - \lambda \frac{\partial g}{\partial y}(a, b) = 0 \\ \frac{\partial L}{\partial \lambda}(a, b, \lambda) &= g(a, b) = 0 \end{aligned}$$

Ce qui donne une condition nécessaire.

Cherchons si cette condition est suffisante. Soit (a, b) comme ci-dessus un point stationnaire de $L(x, y, \lambda)$, écrivons la formule de Taylor pour f , où on note H la matrice hessienne de f en (a, b) .

$$f(a + h, b + k) - f(a, b) = Df(a, b) \begin{pmatrix} h \\ k \end{pmatrix} + (h, k)H \begin{pmatrix} h \\ k \end{pmatrix} + \epsilon(h, k) \quad (8.2)$$

Il faut donc étudier le signe du second membre, non pas pour toutes les valeurs de h et k , mais seulement pour les valeurs qui respectent la contrainte $g(a+h, b+k) = 0$. Or

$$g(a+h, b+k) = g(a, b) + \frac{\partial g}{\partial x}(a, b)h + \frac{\partial g}{\partial y}(a, b)k + \epsilon_1(h, k)$$

En négligeant ϵ_1 on peut écrire,

$$k = -\frac{\partial g}{\partial x}(a, b)h / \frac{\partial g}{\partial y} = hg_x/g_y$$

En notant les dérivations en indice :

$$\frac{\partial f}{\partial x} = f_x \quad \frac{\partial^2 f}{\partial x^2} = f_{xx} \quad \frac{\partial^2 f}{\partial x \partial y} = f_{xy} \dots$$

le second membre de (8.2) devient

$$f_x h + f_y k + f_{xx} h^2 + 2f_{xy} h k + f_{yy} k^2$$

ce qui s'écrit encore

$$f_x h - f_y h g_x / g_y + f_{xx} h^2 - 2f_{xy} h^2 g_x / g_y + f_{yy} h^2 (g_x / g_y)^2$$

La somme des deux premiers termes s'écrit aussi

$$\frac{h}{g_y} (f_x g_y - f_y g_x)$$

et est nulle d'après (8.1). Après mise en facteurs de h^2/g_y^2 , le second membre de (8.2) est donc du signe de

$$f_{xx} g_y^2 - 2f_{xy} g_x g_y + f_{yy} g_x^2 \quad (8.3)$$

On remarque que la quantité (8.3) est exactement l'opposé du déterminant ci-dessous

$$\begin{vmatrix} f_{xx} & f_{xy} & g_x \\ f_{xy} & f_{yy} & g_y \\ g_x & g_y & 0 \end{vmatrix}$$

qui correspond à la matrice hessienne H "bordée" par les dérivées partielles de la contrainte g . Si ce déterminant est strictement négatif, il s'agit d'un minimum ; s'il est positif strictement, il s'agit d'un maximum.

8.2.2 Le Lagrangien

Soit à optimiser une fonction f de n variables $x_1, \dots, x_n$ soumises à p contraintes $g_i(x_1, \dots, x_n) = 0$, $1 \leq i \leq p < n$. On introduit p variables auxiliaires λ_i et on considère le Lagrangien du problème qui est la fonction de $n+p$ variables

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_p) = f(x_1, \dots, x_n) + \sum_{i=1}^p \lambda_i g_i(x_1, \dots, x_n)$$

On dispose du théorème suivant, qui fournit une condition nécessaire :

Théorème 20 *Si f présente un extremum sous les contraintes g_i en un point $(a_1, \dots, a_n)$, alors il existe p réels λ_i , $1 \leq i \leq p$ tels que le point*

$$(a_1, \dots, a_n, \lambda_1, \dots, \lambda_p)$$

soit un point stationnaire du Lagrangien.

DÉMONSTRATION : admise, mais analogue à celle de l'exemple précédent.

Chaque fois que cela est possible, on évite le recours aux conditions du second ordre.

Si le problème est tel que l'on sait par avance que la fonction présente un unique minimum sous contrainte et si le Lagrangien admet un seul point stationnaire, celui-ci est le minimum cherché : il est inutile de vérifier les conditions du second ordre.

Si la fonction f présente des propriétés de convexité, on peut montrer que les conditions nécessaires sont suffisantes dans certains cas.

Pour les cas désespérés, voici des conditions suffisantes :

Proposition 8.2.1 *Soit $a = (a_1, \dots, a_n, \lambda_1, \dots, \lambda_p)$ un point stationnaire du Lagrangien. Soit H la matrice hessienne de f au point $(a_1, \dots, a_n)$. On définit encore des déterminants emboîtés extraits de la matrice H , mais en outre on les borde avec les dérivées partielles des contraintes. Plus*

précisément soit D_k le déterminant

$$D_k = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_k} & \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_p}{\partial x_1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_k} & \cdots & \frac{\partial^2 f}{\partial x_k \partial x_k} & \frac{\partial g_1}{\partial x_k} & \cdots & \frac{\partial g_p}{\partial x_k} \\ \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_k} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial g_p}{\partial x_1} & \cdots & \frac{\partial g_p}{\partial x_k} & 0 & \cdots & 0 \end{vmatrix}$$

où toutes les dérivées sont évaluées au point stationnaire considéré.

Si tous les D_k sont négatifs, le point stationnaire est un minimum.

Si $D_1 < 0$, $D_2 > 0$, $D_3 < 0$, $D_4 > 0$, ..., alors le point stationnaire est un maximum.

DÉMONSTRATION : admise.

Remarquons que les multiplicateurs λ_i n'interviennent pas dans cette condition.

Par définition D_1 est toujours négatif.

Dans le cas $n = 2$, $p = 1$ on retrouve exactement le calcul traité dans l'exemple.

Pour $n = 3$, $p = 1$, les déterminants à considérer sont, avec les notations des dérivations en indices :

$$D_1 = \begin{vmatrix} f_{xx} & g_x \\ g_x & 0 \end{vmatrix} \quad D_2 = \begin{vmatrix} f_{xx} & f_{xy} & g_x \\ f_{xy} & f_{yy} & g_y \\ g_x & g_y & 0 \end{vmatrix} \quad D_3 = \begin{vmatrix} f_{xx} & f_{xy} & f_{xz} & g_x \\ f_{xy} & f_{yy} & f_{yz} & g_y \\ f_{xz} & f_{yz} & f_{zz} & g_z \\ g_x & g_y & g_z & 0 \end{vmatrix}$$

Pour $D_1 < 0$, $D_2 < 0$ et $D_3 < 0$: le point est un minimum.

Pour $D_1 < 0$, $D_2 > 0$ et $D_3 < 0$: le point est un maximum.

Exercice On suppose que $f(x, y, z) = x^2 + y^2 + z^2$. Montrer que si un point est stationnaire pour le Lagrangien, les conditions précédentes entraînent que c'est toujours un minimum (quelle que soit la contrainte). Vérifier que

pour $f(x, y, z) = -x^2 - y^2 - z^2$ on obtient toujours un maximum. Essayer de le justifier directement.

8.2.3 Un exemple numérique

Cherchons parmi tous les parallélépipèdes rectangles inscrits dans la sphère unité celui qui a le plus grand volume. Ceci est équivalent à chercher le maximum de $f(x, y, z) = xyz$ sous la contrainte $g(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$. Remarquons que pour $x = 0$ et $x = 1$ le volume considéré est nul, donc si on trouve un seul autre point stationnaire, il représentera bien le maximum cherché.

On a donc

$$\begin{aligned} L(x, y, z, \lambda) &= xyz + \lambda(x^2 + y^2 + z^2 - 1) \\ \frac{\partial L}{\partial x} &= yz + 2\lambda x \\ \frac{\partial L}{\partial y} &= xz + 2\lambda y \\ \frac{\partial L}{\partial z} &= xy + 2\lambda z \\ \frac{\partial L}{\partial \lambda} &= x^2 + y^2 + z^2 - 1 \end{aligned}$$

Les points stationnaires éventuels vérifient donc

$$0 = yz + 2\lambda x \quad (8.4)$$

$$0 = xz + 2\lambda y \quad (8.5)$$

$$0 = xy + 2\lambda z \quad (8.6)$$

$$0 = x^2 + y^2 + z^2 - 1 \quad (8.7)$$

En multipliant (8.4) par x , (8.5) par y , (8.6) par z et compte tenu de (8.7), on obtient

$$3xyz + 2\lambda = 0$$

En remplaçant $2\lambda = 3xyz$ dans (8.4), (8.5), (8.6) on obtient

$$\begin{aligned} yz(1 - 3x^2) &= 0 \\ xz(1 - 3y^2) &= 0 \\ xy(1 - 3z^2) &= 0 \\ x^2 + y^2 + z^2 &= 1 \end{aligned}$$

Il y a donc cinq points stationnaires pour le Lagrangien

$$O = (0, 0, 0) \quad A = (1, 0, 0) \quad B = (0, 1, 0) \quad C = (0, 0, 1)$$

$$M = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$$

Les points O, A, B, C correspondent manifestement à des minima ; seul M est susceptible d'être un maximum ; comme il en faut bien un, c'est le bon ! Vérifions néanmoins les conditions du second ordre ; au point considéré on a

$$H = \begin{pmatrix} 0 & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{3} & 0 & 1/\sqrt{3} \\ 1/\sqrt{3} & 1/\sqrt{3} & 0 \end{pmatrix} \quad g_x = g_y = g_z = 2/\sqrt{3}$$

Les déterminants à considérer sont donc, après mise en facteur de $1/\sqrt{3}$,

$$D_1 = \begin{vmatrix} 0 & 2 \\ 2 & 0 \end{vmatrix} \quad D_2 = \begin{vmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 0 \end{vmatrix} \quad D_3 = \begin{vmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 2 \\ 2 & 2 & 2 & 0 \end{vmatrix}$$

avec $D_1 = -4$, $D_2 = 8$, $D_3 = -12$, ce qui correspond effectivement à un maximum.

8.3 Contraintes d'inégalité

8.3.1 Introduction

Donnons une interprétation intuitive du problème suivant : on cherche le minimum d'une fonction f de x avec la contrainte $a(x) \leq 0$, où $x \in E \subset \mathbb{R}^n$. Dans de nombreux cas, on est en mesure de montrer que si un point $\hat{x}$ réalise

le minimum sous contrainte, alors $a(\hat{x}) = 0$; on peut alors se ramener au problème précédent avec contrainte d'égalité.

Voici une autre approche du problème : Supposons pour simplifier que f atteint son minimum absolu en un unique point $\hat{x}$. Si $\hat{x}$ vérifie la contrainte $a(\hat{x}) \leq 0$, on a résolu le problème ; supposons donc le contraire : $a(\hat{x}) > 0$. On introduit la famille de problèmes sans contrainte

$$P_\lambda : \inf_{x \in E} f(x) + \lambda a(x) \quad \forall \lambda \geq 0$$

Pour chaque valeur de $\lambda \geq 0$, on cherche le minimum par rapport à x du Lagrangien. Toujours pour simplifier, supposons que la fonction $x \mapsto f(x) + \lambda a(x)$ atteigne son minimum absolu en un unique point que nous noterons x_λ .

Le point important est que la fonction $\lambda \mapsto a(x_\lambda)$ est une fonction décroissante de λ ; en effet pour deux réels λ et μ avec $\lambda < \mu$ on a, par définition de x_λ et x_μ :

$$\begin{aligned} f(x_\mu) + \mu a(x_\mu) &\leq f(x_\lambda) + \mu a(x_\lambda) \\ f(x_\lambda) + \lambda a(x_\lambda) &\leq f(x_\mu) + \lambda a(x_\mu) \end{aligned}$$

En ajoutant ces deux relations, on obtient $a(x_\mu) \leq a(x_\lambda)$.

Par conséquent, $a(x_\lambda)$ est strictement positive pour $\lambda = 0$ (puisque le minimum correspond à $\hat{x}$) et décroît avec λ . Il y a donc intuitivement une valeur ρ de λ telle que $a(x_\rho) = 0$.

Montrons que x_{rho} est solution du problème d'optimisation considéré : d'abord x_ρ vérifie la contrainte, puis on a

$$f(x_\rho) = f(x_\rho) + \rho a(x_\rho) \leq f(x) + \rho a(x) \leq f(x)$$

pour tous les x vérifiant la contrainte $a(x) \leq 0$, ce qui est exactement ce que nous cherchions.

En résumé : Si $\hat{x}$ est solution du problème d'optimisation sous contrainte, deux cas se présentent :

Le point $\hat{x}$ est intérieur au domaine délimité par la contrainte, il réalise alors le minimum de f sans contrainte.

Le point $\hat{x}$ est sur la frontière du domaine délimité par la contrainte, il réalise

alors le minimum du Lagrangien.

Exemple 1 :

Chercher les extrema de $f(x, y) \mapsto (y - x)^3 + 6xy$ sur le domaine $-1 \leq x \leq y \leq 1$.

Le domaine considéré correspond au triangle rectangle isocèle délimité par les droites $x = -1$, $y = 1$ et $y = x$. Cherchons les points stationnaires de f . On a

$$\frac{\partial f}{\partial x} = -3x^2 + 6xy - 3y^2 + 6y = 0 \quad \frac{\partial f}{\partial y} = 3y^2 - 6xy + 3x^2 + 6x = 0$$

En ajoutant ces deux relations on obtient $x = -y$, en reportant dans la première on aboutit à : $6x(2x + 1) = 0$. Il y a donc deux points stationnaires $A(-\frac{1}{2}, \frac{1}{2})$ et l'origine O. Seul le point A est intérieur au domaine. La matrice hessienne correspondant au point A est

$$\begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}$$

Ce point est donc un minimum.

Il faut maintenant étudier les variations de f sur chacun des bords $x = -1$, $y = 1$ et $x = y$ les extrema de f . Il faut résoudre successivement 3 problèmes d'extrema avec une contrainte d'égalité soit en utilisant le Lagrangien soit par substitution directe.

Bord $x = -1$ On a $f(-1, y) = g(y) = (1 + y)^3 - 6y$. La dérivée $g'(y) = 3(y^2 + 2y - 1)$ s'annule sur $[-1, 1]$ pour la valeur $-1 + \sqrt{2}$. Le tableau de variation correspondant est

y	-1		$-1 + \sqrt{2}$		1
g	6	$\searrow$	$6 - 4\sqrt{2}$	$\nearrow$	2

Le maximum vaut 6, atteint pour $(-1, -1)$ et le minimum vaut $6 - 4\sqrt{2} \approx 0.34$.

Bord $y = 1$ On a $f(x, 1) = h(x) = (1 - x)^3 + 6x$. On a immédiatement

x	-1		$1 - \sqrt{2}$		1
h	2	$\searrow$	$6 - 4\sqrt{2}$	$\nearrow$	6

Bord $x=y$ On a $f(x, x) = k(x) = 6x^2$. Le minimum est 0 atteint à l'origine et le maximum 6 atteint en $(1, 1)$ et $(-1, -1)$.

En conclusion, la maximum absolu de f vaut 6, le minimum absolu vaut $f(-\frac{1}{2}, \frac{1}{2}) = -\frac{1}{2}$ atteint en A.

Exemple 2 :

Déterminer les extrema de $f(x, y) = x^2(y^2 + 1) + y(y + \frac{5}{2}x)$ sur le domaine

$$\mathcal{D} = \{(x, y) | x \leq 0 \text{ et } x^2 + y^2 \leq 1\}$$

Cherchons les points stationnaires de f :

$$\frac{\partial f}{\partial x} = 2xy^2 + 2x + \frac{5}{2}y = 0 \quad \frac{\partial f}{\partial y} = 2x^2y + 2y + \frac{5}{2}x = 0$$

En multipliant la première relation par x et la deuxième par y , on obtient par différence $x^2 = y^2$, soit $x = \pm y$. Pour $x = y$: en reportant dans la première on trouve $4x^3 + 9x = x(4x^2 + 9) = 0$. La seule possibilité est $x = 0$, ce qui donne l'origine comme point stationnaire : il n'est pas intérieur.

Pour $x = -y$: en reportant on obtient $4x^3 - x = x(4 - x^2) = 0$. Mis à part la solution $x = 0$, on trouve un autre point stationnaire $A(\frac{1}{2}, -\frac{1}{2})$. La matrice hessienne correspondant est alors

$$\frac{1}{2} \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$$

Il s'agit donc d'un minimum, et f vaut $-\frac{11}{16}$.

Regardons sur chacun des bords.

Bord $x = 0$ On a $f(0, y) = y^2$. Le maximum vaut 1, atteint pour $(0, -1)$ et $(0, 1)$; le minimum vaut 0.

Bord $x^2 + y^2 = 1$ La substitution ne paraissant pas simple, on va traiter le problème à l'aide du Lagrangien. Posons

$$L(x, y, \lambda) = f(x, y) + \lambda(x^2 + y^2 - 1)$$

La recherche des points stationnaires conduit au système

$$\begin{aligned} \frac{\partial L}{\partial x} &= 2xy^2 + 2x + \frac{5}{2}y + 2\lambda x = 0 \\ \frac{\partial L}{\partial y} &= 2yx^2 + 2y + \frac{5}{2}x + 2\lambda y = 0 \\ \frac{\partial L}{\partial \lambda} &= x^2 + y^2 - 1 = 0 \end{aligned}$$

En multipliant la première relation par x , la deuxième par y et en faisant la différence, on obtient

$$(x^2 - y^2)(2 + 2\lambda) = 0.$$

Il y a trois possibilités à envisager :

$\lambda = -1$: en remplaçant on trouve $xy = \frac{5}{4}$ ce qui est incompatible avec la contrainte $x^2 + y^2 = 1$.

$x = \pm y$: compte tenu de la contrainte $x^2 + y^2 = 1$ on trouve deux points stationnaires sous contrainte

$$B\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \quad \text{et} \quad C\left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right).$$

Etudions la nature de ces points. On a :

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= 2y^2 + 2 & \frac{\partial^2 f}{\partial y^2} &= 2x^2 + 2 \\ \frac{\partial^2 f}{\partial x \partial y} &= 4xy + \frac{5}{2} \end{aligned}$$

Au point B la matrice hessienne bordée est donc

$$\begin{pmatrix} 3 & \frac{9}{2} & \sqrt{2} \\ \frac{9}{2} & 3 & \sqrt{2} \\ \sqrt{2} & \sqrt{2} & 0 \end{pmatrix}$$

Le déterminant D_1 vaut -2, le déterminant D_2 vaut 6 : il s'agit d'un maximum. Au point B f vaut $\frac{5}{2}$.

Pour le point C la matrice hessienne bordée est

$$\begin{pmatrix} 3 & \frac{1}{2} & \sqrt{2} \\ \frac{1}{2} & 3 & -\sqrt{2} \\ \sqrt{2} & -\sqrt{2} & 0 \end{pmatrix}$$

Le déterminant D_1 vaut -2, le déterminant D_2 vaut -14 : il s'agit d'un minimum. Au point C f vaut 0.

En conclusion, la valeur minimum de f est $-\frac{11}{16}$ et la valeur maximum est $\frac{5}{2}$.

Annexe A

Appendice : La méthode de Khün et Tucker

A.1 Introduction

Nous avons vu que si $\hat{x}$ est solution du problème d'optimisation sous contrainte $(\hat{x}, \rho)$ est un point stationnaire de

$$L(x, \lambda) = f(x) + \lambda a(x)$$

qui vérifie soit $\lambda = 0$ soit $a(x) = 0$; le multiplicateur de "Lagrange" dans ce cas s'appelle un multiplicateur de Kuhn-Tucker. Par construction ce multiplicateur est positif, c'est en cela qu'il est différent d'un multiplicateur de Lagrange.

La méthode de Kuhn et Tucker est simplement une généralisation de ce qui précède : elle exprime que l'optimum cherché est soit intérieur au domaine, c'est à dire sans contrainte, soit obtenu au bord du domaine, où une ou plusieurs des contraintes sont des égalités. On dit que la contrainte g_i est saturée en un point x si on a : $g_i(x) = 0$.

L'intérêt des calculs suivants peut sembler minime. Pour des problèmes simples (que l'on peut traiter "à la main") cette méthode est bien lourde ; par contre dans les cas plus compliqués, se ramener à un système formé uniquement d'égalités permet un traitement informatique bien plus aisé.

Théorème 21 Soient $f, h_1, \dots, h_q, g_1, \dots, g_r$, des fonctions définies et continûment dérivables sur un ouvert $\Omega \in \mathbb{R}^n$. Soit D le domaine des solutions possibles défini par les x de Ω tels que $\forall i \ 1 \leq i \leq q \ h_i(x) = 0$ et $\forall j \ 1 \leq j \leq r \ g_j(x) \geq 0$.

Si f admet un maximum local en $\hat{x} = (x_1, \dots, x_n)$ et si les contraintes sont régulières au point $\hat{x} \in D$, c'est-à-dire que les q différentielles Dh_i et les k différentielles Dg_j correspondant aux contraintes saturées sont linéairement indépendantes, il existe des réels $\lambda_1, \dots, \lambda_q$ et $\mu_1, \dots, \mu_r$ tels que si on pose

$$L = f + \sum_{i=1}^q \lambda_i h_i + \sum_{j=1}^r \mu_j g_j$$

on ait :

- (i) $\forall j \ 1 \leq j \leq r \ \mu_j \geq 0$
- (ii) $\forall j \ 1 \leq j \leq r \ \mu_j g_j(\hat{x}) = 0$
- (iii) $(x_1, \dots, x_n, \lambda_1, \dots, \lambda_q, \mu_1, \dots, \mu_r)$ est un point stationnaire de L .

DÉMONSTRATION : admise

Il est possible qu'il existe des conditions suffisantes (i.e. du second ordre) correspondantes ...

Le signe des multiplicateurs de Kuhn et Tucker peut se retrouver ainsi : si le maximum est atteint sur le bord, et que l'on se dirige vers l'intérieur, le Lagrangien ne peut que diminuer ; mais si on se dirige vers l'intérieur les $\mu_j g_j$ augmentent, ce qui oblige donc f à diminuer ! Avec les mêmes contraintes, pour un minimum, les signes des multiplicateurs seront donc négatifs !

A.2 Exemples

Considérons l'exemple suivant : Extrema de $U(x, y) = (x + 1)(y + 1)$ sous les contraintes $6x + y = 3$, $x \geq 0$, $y \geq 0$. Formons le Lagrangien L

$$L(x, y, \lambda, \mu_1, \mu_2) = (x + 1)(y + 1) + \lambda(-6x - y + 3) + \mu_1 x + \mu_2 y$$

L'optimum (a, b) est solution du système :

$$\left\{ \begin{array}{lcl} \frac{\partial L}{\partial x}(a, b) & = & b + 1 - 6\lambda + \mu_1 = 0 \\ \frac{\partial L}{\partial y}(a, b) & = & a + 1 - \lambda + \mu_2 = 0 \\ & & 6a + b - 3 = 0 \\ & & \mu_1 a = 0 \\ & & \mu_2 b = 0 \end{array} \right.$$

La résolution du système se ramène à une discussion sur la nullité des μ_i qui fait intervenir 4 cas

1. $\mu_1 = \mu_2 = 0$. Le système devient

$$b + 1 - 6\lambda = 0 \quad a + 1 - \lambda = 0 \quad 6a + b - 3 = 0$$

avec la solution $a = -2/3$, $b = 1$ et $\lambda = 1/3$ qui ne convient pas.

2. $\mu_1 \neq 0$ et $\mu_2 = 0$ Le système conduit à $a = 0$, $b = 3$, $\lambda = 1$ et $\mu_1 = 2$.
3. $\mu_1 = 0$ et $\mu_2 \neq 0$. Le système conduit à $b = 0$, $a = 1/2$, $\lambda = 1/6$ et $\mu_2 = -4/3$ ce qui ne convient pas.
4. $\mu_1 > 0$ et $\mu_2 > 0$. On obtient $a = b = 0$ ce qui ne convient pas.

Si U est maximum, il ne peut l'être qu'au point $(0, 3)$. On montrerait d'une façon similaire que si U est minimum, il ne peut l'être qu'au point $(1/2, 0)$.

Exercice Substituer $y = 6x - 3$ dans f et vérifier.