Quelques éléments de calcul matriciel

Extrait de l'ouvrage : « Traitement des données statistiques »

L. Lebart   A. Morineau   J.-P. Fénelon   **paru chez Dunod en 1979**

III. RAPPELS DE CALCUL MATRICIEL

Pour fixer à la fois le vocabulaire et les notations, on énumère un ensemble de définitions et de propriétés concernant le calcul matriciel. La généralité a été sacrifiée au profit de la simplicité et de la concision ; aucune démonstration n'est donnée. Cette annexe ne saurait remplacer un cours ou un manuel de mathématiques, mais permettra de préciser l'étendue des connaissances nécessaires à la compréhension des chapitres de statistique (en particulier les chapitres III et IV).

1. *DÉFINITIONS ET OPÉRATIONS ÉLÉMENTAIRES SUR LES MATRICES*

Une *matrice* est un tableau rectangulaire dont les éléments sont rangés en lignes et colonnes. Notation pour une matrice **X** ayant n lignes et p colonnes :

$$\underset{(n,\,p)}{\mathbf{X}} = \{ x_{ij} \} \qquad \underset{\text{lignes}}{i = 1, 2, \ldots, n} \qquad \underset{\text{colonnes}}{j = 1, 2, \ldots, p}$$

où x_{ij} est l'élément situé à l'intersection de la *ligne* i (premier indice) et de la *colonne* j (deuxième indice).

On appelle *transposée* de **X**, et on note **X**′, la matrice obtenue en échangeant le rôle des lignes et des colonnes. Ainsi, si les dimensions de **X** sont (n, p), les dimensions de **X**′ seront (p, n).

Une matrice est dite *carrée* si le nombre de lignes n est égal au nombre de colonnes p.

Une matrice est dite *symétrique* si elle est égale à sa transposée (c'est-à-dire si les éléments ayant mêmes indices dans **A** et **A**′ sont égaux : $a_{ij} = a_{ji}$) ; on écrit $\mathbf{A} = \mathbf{A}'$. Une matrice symétrique est nécessairement carrée. Cas particuliers de matrices symétriques :

- la matrice *unité* **I** (des 1 dans la *diagonale principale* et des 0 ailleurs) ;
- la matrice *scalaire* **S** (une même valeur s dans la diagonale principale et des 0 ailleurs) ;
- la matrice *diagonale* **D** (des éléments quelconques dans la diagonale principale et des 0 ailleurs).

Une matrice est dite *nulle* si tous ses éléments sont nuls. Exemple : la matrice **0** ayant n lignes et p colonnes, définie par :

$$\underset{(n,\,p)}{\mathbf{0}} = \{ x_{ij} \} \quad \text{avec } x_{ij} = 0 \quad (i = 1, 2, \ldots, n \text{ et } j = 1, 2, \ldots, p)\,.$$

Une matrice (n, 1) est appelée *vecteur-colonne*, une matrice (1, p) est appelée *vecteur-ligne*. Le transposé d'un vecteur-ligne est un vecteur-colonne, et vice versa. Exemple :

$$\text{Vecteur-colonne } \underset{(n,\,1)}{\mathbf{u}} = \{ u_i \} \qquad (u_i = 1\,;\ i = 1, 2, \ldots, n)\,.$$

Si **X** et **Y** sont deux matrices de mêmes dimensions (n, p), on définit leur *somme* :

$$\underset{(n,\,p)}{\mathbf{Z}} = \underset{(n,\,p)}{\mathbf{X}} + \underset{(n,\,p)}{\mathbf{Y}}$$

par : $z_{ij} = x_{ij} + y_{ij} \qquad (i = 1, 2, \ldots, n \text{ et } j = 1, 2, \ldots, p)$.

Multiplication par un scalaire. Multiplier la matrice **X** par un scalaire s, c'est multiplier tous ses éléments par ce scalaire.

Multiplication matricielle. Soit **R** une matrice (n, p) et **S** une matrice (p, m) ; on définit la matrice-produit :

$$\underset{(n,\,m)}{\mathbf{T}} = \underset{(n,\,p)}{\mathbf{R}}\ \underset{(p,\,m)}{\mathbf{S}}$$

par

$$\mathbf{T} = \{ t_{ik} \} \qquad (i = 1, 2, \ldots, n \text{ et } k = 1, 2, \ldots, m)$$

avec

$$t_{ik} = \sum_{j=1}^{p} r_{ij}\, s_{jk}\,.$$

Pour que le produit **RS** soit possible, le nombre de colonnes de **R** doit être égal au nombre de lignes de **S**.

Remarques sur les opérations élémentaires

Lorsque les opérations sont possibles :

$$\mathbf{X} + \mathbf{Y} = \mathbf{Y} + \mathbf{X}$$

$$\mathbf{XY} \neq \mathbf{YX} \ \text{(en général)}$$

$$(\mathbf{X} + \mathbf{Y}) + \mathbf{Z} = \mathbf{X} + (\mathbf{Y} + \mathbf{Z})$$

$$(\mathbf{XY})\,\mathbf{Z} = \mathbf{X}(\mathbf{YZ})$$

$$\mathbf{X}(\mathbf{Y} + \mathbf{Z}) = \mathbf{XY} + \mathbf{XZ}$$

$$s(\mathbf{X} + \mathbf{Y}) = s\mathbf{X} + s\mathbf{Y}$$

transposition :

$$(\mathbf{X}')' = \mathbf{X}$$
$$(\mathbf{X} + \mathbf{Y})' = \mathbf{X}' + \mathbf{Y}'$$
$$(\mathbf{YZ})' = \mathbf{Z}'\,\mathbf{Y}'$$
$$(\mathbf{XYZ})' = \mathbf{Z}'\,\mathbf{Y}'\,\mathbf{X}' .$$

Matrices identité et diagonale :

- le produit de deux matrices diagonales est une matrice diagonale,
- si **A** est carrée, s un scalaire, une matrice identité **I** de rang convenable peut être introduite ou supprimée à volonté :

$$s\mathbf{A} = (s\mathbf{I})\,\mathbf{A} = \mathbf{A}(s\mathbf{I}) = s\mathbf{IA} = \mathbf{A}s\mathbf{I} = \mathbf{A}s .$$

Symétrie :

- Si **X** est une matrice (n, p), alors $\mathbf{X}'\,\mathbf{X}$ est une matrice (p, p) symétrique et $\mathbf{XX}'$ est une matrice (n, n) symétrique.

2. *LONGUEUR, PRODUIT SCALAIRE, FORME QUADRATIQUE*

Soit les vecteurs-colonnes : $\underset{(n,1)}{\mathbf{x}}$ et $\underset{(n,1)}{\mathbf{y}}$.

- *Longueur de* **x** :

$$\mathbf{x}'\,\mathbf{x} = \sum_{i=1}^{n} x_i^2 .$$

- *Produit scalaire* de **x** et **y** :

$$\mathbf{x}'\,\mathbf{y} = \mathbf{y}'\,\mathbf{x} = \sum_{i=1}^{n} x_i\,y_i$$
$$\cos(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\,\mathbf{y}/\sqrt{(\mathbf{x}'\,\mathbf{x})\,(\mathbf{y}'\,\mathbf{y})}$$

(d'après l'inégalité de Schwarz : $-1 \leqslant \cos(\mathbf{x}, \mathbf{y}) \leqslant 1$).

- *Vocabulaire* : $\mathbf{x}'\,\mathbf{y} = \mathbf{x}'\,\mathbf{I}\mathbf{y}$ = produit scalaire euclidien usuel.

$$(\mathbf{x}'\,\mathbf{x})^{1/2} = (\mathbf{x}'\,\mathbf{I}\mathbf{x})^{1/2} = \text{norme euclidienne usuelle de } \mathbf{x} .$$

- *Orthogonalité* de **x** et de **y** : si $\mathbf{x}'\,\mathbf{y} = 0$ (métrique euclidienne) ; géométriquement $\cos(\mathbf{x}, \mathbf{y}) = 0$ dans un espace à n dimensions.
- *Forme quadratique* : c'est la fonction réelle des x_i définie par :

$$\mathbf{x}'\,\mathbf{A}\mathbf{x} = \sum_i \sum_j x_i\,x_j\,a_{ij}$$

où **A** est une matrice (n, n) symétrique, et **x** un vecteur-colonne (n, 1).

Une forme quadratique est dite :

- *définie positive* si $\mathbf{x}'\,\mathbf{A}\mathbf{x} > 0$ quel que soit $\underset{(n,\,1)}{\mathbf{x}} \neq \mathbf{0}$;
- *semi-définie positive* si $\mathbf{x}'\,\mathbf{A}\mathbf{x} \geqslant 0$ quel que soit $\mathbf{x}$; alors il peut exister $\mathbf{x} \neq \mathbf{0}$ tel que $\mathbf{x}'\,\mathbf{A}\mathbf{x} = 0$.

Exemples :
- $\mathbf{x}'\,\mathbf{I}\mathbf{x} = \mathbf{x}'\,\mathbf{x}$ est une forme quadratique définie positive ;
- si $\mathbf{X}$ est une matrice quelconque, $\mathbf{X}'\,\mathbf{X}$ et $\mathbf{X}\mathbf{X}'$ sont des matrices symétriques semi-définies positives (c'est-à-dire associées à une forme quadratique semi-définie positive).

3. *INDÉPENDANCE LINÉAIRE, RANG D'UNE MATRICE*

Soit $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ des vecteurs-colonnes de dimension n, et soit $c_1, c_2, \ldots, c_p$ des coefficients réels ; on considère le vecteur *combinaison linéaire* défini par :

$$c_1\,\mathbf{x}_1 + c_2\,\mathbf{x}_2 + \cdots + c_p\,\mathbf{x}_p\,.$$

On dit que les vecteurs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ sont *linéairement indépendants* (ou indépendants) si :

$$c_1\,\mathbf{x}_1 + c_2\,\mathbf{x}_2 + \cdots + c_p\,\mathbf{x}_p = \mathbf{0} \text{ implique } c_1 = c_2 = \cdots c_p = 0\,.$$

Exemples :

$$c_1\begin{bmatrix}1\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}c_1\\c_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \text{ implique } c_1 = c_2 = 0$$

par contre :

$$c_1\begin{bmatrix}1\\2\end{bmatrix} + c_2\begin{bmatrix}2\\4\end{bmatrix} = \begin{bmatrix}c_1 + 2\,c_2\\2\,c_1 + 4\,c_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \text{ pour } \begin{matrix}c_1 = 1\\c_2 = -\,0{,}5\,.\end{matrix}$$

Rang d'une matrice

Soit $\mathbf{X}$ une matrice (n, p). Le rang en ligne de $\mathbf{X}$ est égal au plus grand nombre de vecteurs-lignes indépendants ; le rang en colonne, au plus grand nombre de vecteurs-colonnes indépendants. On peut montrer que le rang en ligne est *toujours* égal au rang en colonne, et c'est pourquoi on peut parler du *rang* de la matrice $\mathbf{X}$; on le note : rg $\mathbf{X}$.

Si le rang d'une matrice $\mathbf{X}$ est maximum, c'est-à-dire égal à la plus petite dimension de $\mathbf{X}$, on dit que la matrice est de *plein rang*. Une matrice carrée (n, n) est de plein rang si son rang est égal à n.

- Le rang d'une matrice nulle est égal à 0.
- $\text{rg}\,(\mathbf{AB}) \leqslant \min\{\text{rg}\,\mathbf{A}, \text{rg}\,\mathbf{B}\}$.

- rg $\mathbf{X}$ = rg $\mathbf{X}'$ = rg $(\mathbf{X}'\mathbf{X})$ = rg $(\mathbf{X}\mathbf{X}')$.
- rg $\mathbf{X}$ = dimension du plus grand *déterminant* non nul extrait de $\mathbf{X}$. (On désigne par det $\mathbf{A}$ le déterminant de $\mathbf{A}$.)

4. *MATRICE INVERSE*

Soit $\mathbf{A}$ une matrice *carrée* (p, p). On dit qu'elle est *non singulière* (ou régulière) si elle est de plein-rang, c'est-à-dire si rg $\mathbf{A}$ = p. Si $\mathbf{A}$ est non singulière (p, p), il existe une matrice $\mathbf{B}$ de dimension (p, p) telle que :

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

$\mathbf{B}$ est appelée l'*inverse* de $\mathbf{A}$, et notée $\mathbf{A}^{-1}$:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}\,.$$

- L'inverse de $\mathbf{A}$, si elle existe, est unique.
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
- $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.
- det $\mathbf{A} \neq 0$ si et seulement si $\mathbf{A}$ est non singulière.
- Si $\mathbf{A}$ et $\mathbf{B}$ sont deux matrices non singulières de mêmes dimensions, alors $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
- Une matrice diagonale est non singulière si et seulement si tous ses éléments diagonaux sont non nuls. L'inverse est alors la matrice diagonale des inverses des éléments diagonaux (voir le point 5).
- det $(\mathbf{AB})$ = (det $\mathbf{A}$) (det $\mathbf{B}$).
- Si rg $\mathbf{B}$ = rg $\mathbf{C}$ = rg $\mathbf{A}$ alors rg $(\mathbf{AB})$ = rg $(\mathbf{CA})$ = rg $\mathbf{A}$ pourvu que les produits $\mathbf{AB}$ et $\mathbf{CA}$ soient possibles.
- Si $\mathbf{A}$ est définie positive, $\mathbf{A}$ est non singulière, et $\mathbf{A}^{-1}$ est définie positive.
- Une matrice carrée $\mathbf{A}$ est dite *orthogonale* si $\mathbf{A}'\mathbf{A} = \mathbf{I}$. On a alors $\mathbf{A}' = \mathbf{A}^{-1}$ et det $\mathbf{A} = 1$.
- Pour toute matrice symétrique $\mathbf{A}$, il existe une matrice orthogonale $\mathbf{X}$ telle que $\mathbf{X}'\mathbf{AX} = \mathbf{\Lambda}$ où $\mathbf{\Lambda}$ est une matrice diagonale (voir le point 5).
- Si $\mathbf{A}$ est une matrice symétrique définie positive, il existe une matrice régulière $\mathbf{C}$ telle que $\mathbf{CAC}' = \mathbf{I}$ (ce qui implique $\mathbf{C}'\mathbf{C} = \mathbf{A}^{-1}$).

Solution d'un système d'équations linéaires

On considère un système de p équations dont les p inconnues x_i constituent le vecteur $\mathbf{x}$; soit $\mathbf{A}$ la matrice (p, p) des coefficients et $\mathbf{b}$ le vecteur (p, 1) des seconds membres ; le système s'écrit :

$$\underset{(p,\,p)}{\mathbf{A}}\ \underset{(p,\,1)}{\mathbf{x}} = \underset{(p,\,1)}{\mathbf{b}}\,.$$

Il existe une solution $\mathbf{x}$ unique si et seulement si $\mathbf{A}$ est de plein rang. On calcule alors cette solution par :

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} .$$

5. *VECTEURS ET VALEURS PROPRES*

Soit $\mathbf{A}$ une matrice carrée (p, p) ; soit λ un scalaire et $\mathbf{x}$ un vecteur (p, 1) *non nul* tels que :

$$\underset{(p,\,p)}{\mathbf{A}}\ \underset{(p,\,1)}{\mathbf{x}} = \lambda\ \underset{(p,\,1)}{\mathbf{x}} .$$

Alors $\mathbf{x}$ est déterminé à un facteur multiplicatif près car, pour tout scalaire s, $\mathbf{A}(s\mathbf{x}) = \lambda(s\mathbf{x})$; on peut donc imposer à $\mathbf{x}$ la longueur 1 sans perte de généralité :

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \qquad \text{avec } \mathbf{x}'\,\mathbf{x} = 1$$

ou

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0} \qquad \text{avec } \mathbf{x}'\,\mathbf{x} = 1 .$$

Autrement dit les colonnes de $\mathbf{A} - \lambda\mathbf{I}$ sont linéairement dépendantes et par conséquent :

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 .$$

Cette relation de définition de λ est appelée *équation caractéristique* de la matrice $\mathbf{A}$. Le scalaire λ est appelé une *valeur propre* et $\mathbf{x}$ un *vecteur propre* de $\mathbf{A}$ correspondant à λ.

Plus généralement, si $\mathbf{A}$ est une matrice (p, p) :

- L'équation caractéristique est un polynôme de degré p, donc $\mathbf{A}$ possède p valeurs propres (pas nécessairement distinctes).
- La *somme* des valeurs propres est égale à la *trace* de la matrice (somme des termes diagonaux).
- Le *produit* des valeurs propres est égal au déterminant de la matrice.
- Si $\mathbf{A}$ est une matrice carrée et $\mathbf{P}$ une matrice régulière, alors $\mathbf{A}$ et $\mathbf{PAP}^{-1}$ ont mêmes valeurs propres ; par suite, si $\mathbf{P}$ est orthogonale, $\mathbf{A}$ et $\mathbf{PAP}'$ ont mêmes valeurs propres.

  Si $\mathbf{P}$ est régulière :

  $$\operatorname{tr}\mathbf{A} = \operatorname{tr}(\mathbf{PAP}^{-1})$$

  $$\operatorname{rang}(\mathbf{A}) = \operatorname{rang}(\mathbf{PAP}^{-1}) .$$

- Si $\mathbf{A}$ est une matrice non symétrique, les valeurs propres peuvent être des nombres complexes ; mais si $\mathbf{A}$ est *symétrique* toutes ses valeurs propres sont réelles.
- Si $\mathbf{A}$ est *symétrique*, deux vecteurs propres quelconques correspondant à des valeurs propres distinctes sont *orthogonaux*.

- Si λ est valeur propre de $\mathbf{A}$ *symétrique*, et $\mathbf{x}$ vecteur propre correspondant, λ^2 est valeur propre de $\mathbf{AA} = \mathbf{A}^2$, et $\mathbf{x}$ est le vecteur propre correspondant.
- Si $\mathbf{A}$ est *symétrique* et non singulière, les valeurs propres de $\mathbf{A}^{-1}$ sont les inverses de celles de $\mathbf{A}$, et les vecteurs-propres sont les mêmes.
- Si $\mathbf{A}$ est *symétrique* et *définie positive*, ses valeurs propres sont toutes positives. Si $\mathbf{A}$ est *symétrique* et *semi-définie positive*, elles sont toutes non négatives (positives ou nulles).
- Si $\mathbf{A}$ est *symétrique*, le nombre de ses valeurs propres non nulles est égal au rang de $\mathbf{A}$.
- Soit $\mathbf{X}$ une matrice quelconque ; alors les valeurs propres non nulles de $\mathbf{X}'\mathbf{X}$ et $\mathbf{XX}'$ sont les mêmes (cf. § IV.1.2).

6. *TRACE D'UNE MATRICE CARRÉE*

Soit $\mathbf{A}$ une matrice (p, p), on appelle *trace* de A le scalaire

$$\operatorname{tr} \mathbf{A} = \sum_i a_{ii} .$$

La trace est un *opérateur linéaire*, c'est-à-dire :

- pour tout scalaire s : $\operatorname{tr}(s\mathbf{A}) = s \operatorname{tr} \mathbf{A}$,
- $\operatorname{tr}(\mathbf{A} + \mathbf{B}) = \operatorname{tr} \mathbf{A} + \operatorname{tr} \mathbf{B}$.

De plus :

- si les produits sont possibles $\operatorname{tr} \mathbf{AB} = \operatorname{tr} \mathbf{BA}$.

7. *MATRICE SYMÉTRIQUE ET IDEMPOTENTE*

Une matrice carrée $\mathbf{A}$ est dite *idempotente* si :

$$\mathbf{A} = \mathbf{A}^2 .$$

On ne considérera ici que des matrices idempotentes qui sont également symétriques : $\mathbf{A} = \mathbf{A}' = \mathbf{A}^2 = \mathbf{AA}' = \mathbf{A}'\mathbf{A}$. Par exemple la matrice unité $\mathbf{I}$ est symétrique et idempotente (et définie positive).

- Si $\mathbf{A}$ est symétrique et idempotente, $\mathbf{I} - \mathbf{A}$ aussi.
- Si $\mathbf{A}$ est symétrique et idempotente, elle est semi-définie positive ; ses éléments diagonaux sont non négatifs ; si un élément diagonal est nul, les éléments de toute la ligne et de toute la colonne sont nuls.
- Si $\mathbf{A}$ est symétrique et idempotente, ses valeurs propres sont égales à 1 ou à 0 ; puisque le nombre de valeurs propres non nulles d'une matrice symétrique est égal à son rang, le rang de $\mathbf{A}$ est égal à la trace de $\mathbf{A}$.

8. *INVERSE D'UNE MATRICE SYMÉTRIQUE PARTITIONNÉE*

Soit **H** une matrice *symétrique* partitionnée en 4 blocs **A**, **B**, **B**′, **D**, le bloc **A** étant lui-même une matrice carrée (et par conséquent symétrique) :

$$\mathbf{H} = \left|\begin{array}{c:c} \mathbf{A} & \mathbf{B} \\ \hdashline \mathbf{B}' & \mathbf{D} \end{array}\right| .$$

On suppose **H** non singulière, donc $\mathbf{H}^{-1}$ existe, et est symétrique. Partitionnons $\mathbf{H}^{-1}$ de la même façon :

$$\mathbf{H}^{-1} = \left|\begin{array}{c:c} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \hdashline \tilde{\mathbf{B}}' & \tilde{\mathbf{D}} \end{array}\right| .$$

Parmi les nombreuses façons d'écrire les blocs de $\mathbf{H}^{-1}$ en fonction de ceux de **H**, on retiendra la forme suivante :

$$\begin{aligned} \tilde{\mathbf{A}} &= (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{B}')^{-1} \\ \tilde{\mathbf{B}} &= -\tilde{\mathbf{A}}\mathbf{B}\mathbf{D}^{-1} \qquad (\text{donc } \tilde{\mathbf{B}}' = -\mathbf{D}^{-1}\mathbf{B}'\tilde{\mathbf{A}}) \\ \tilde{\mathbf{D}} &= \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{B}'\tilde{\mathbf{B}} \end{aligned}$$

(vérifier en dimensionnant les blocs que ces calculs sont possibles, et que $\mathbf{H}^{-1}$ est bien l'inverse de **H**).

Exemple. On peut retrouver avec ces formules l'inverse d'une matrice (2, 2) symétrique partitionnée en ses 4 éléments :

$$\mathbf{H} = \left|\begin{array}{c:c} a & b \\ \hdashline b & d \end{array}\right|, \qquad \mathbf{H}^{-1} = \left|\begin{array}{c:c} \tilde{a} & \tilde{b} \\ \hdashline \tilde{b} & \tilde{d} \end{array}\right|$$

$$\begin{aligned} \tilde{a} &= 1/(a - b^2/d) = d/(ad - b^2) \\ \tilde{b} &= \{-d/(ad - b^2)\}.(b/d) = -b/(ad - b^2)\,;\ \text{etc.} \end{aligned}$$

9. *DÉRIVATION VECTORIELLE ET MATRICIELLE*

On considère une fonction réelle f de p variables x_j, composantes d'un vecteur-colonne $\mathbf{x}$:

$$y = f(\mathbf{x}) .$$

Exemple. Si **a** est un vecteur-colonne de p coefficients a_j, on peut définir la fonction réelle

$$y = f(\mathbf{x}) = \mathbf{a}'\mathbf{x} = \sum_j a_j x_j .$$

Pour éviter toute difficulté technique on supposera toujours que les *dérivées partielles* par rapport aux x_i existent au moins jusqu'au second ordre, et sont continues.

On appelle *gradient* de la fonction f le vecteur-colonne à p composantes, dont la j-ième est la *dérivée partielle* de la fonction f par rapport à la variable x_j ; on note :

$$\frac{\partial f}{\partial \mathbf{x}} = \left(\frac{\partial f}{\partial x_j}\right) \qquad \text{vecteur } (p, 1) : j = 1, 2, \ldots, p\,.$$

Exemple : $y = f(\mathbf{x}) = \mathbf{a}'\,\mathbf{x} = \sum_j a_j\,x_j$; alors :

$$\frac{\partial f}{\partial \mathbf{x}} = \underset{(p,\,1)}{\mathbf{a}} \qquad \text{(vecteur-colonne)}\,.$$

Le gradient est donc la généralisation vectorielle de la dérivée première. De façon semblable l'expression vectorielle de la dérivée seconde sera la *matrice Hessienne* : matrice (p, p) dont l'élément d'indices i et j est la dérivée partielle seconde de f par rapport à x_i et x_j :

$$\frac{\partial^2 f}{\partial \mathbf{x}\,\partial \mathbf{x}'} = \left[\frac{\partial^2 f}{\partial x_i\,\partial x_j}\right] \qquad \text{matrice } (p, p) \ : \ i, j = 1, 2, \ldots, p\,.$$

Si les dérivées secondes sont continues on démontre que l'ordre de dérivation (d'abord x_i puis x_j, ou vice et versa) est indifférent (théorème de Young) ; donc la matrice Hessienne est *symétrique*.

Exemple :

$$\frac{\partial^2(\mathbf{a}'\,\mathbf{x})}{\partial \mathbf{x}\,\partial \mathbf{x}'} = \underset{(p,\,p)}{\mathbf{0}}\,.$$

Quelques formules utiles

- $\dfrac{\partial(\mathbf{a}'\,\mathbf{x})}{\partial \mathbf{x}} = \underset{(p,\,1)}{\mathbf{a}}$; $\dfrac{\partial(\mathbf{x}'\,\mathbf{a})}{\partial \mathbf{x}} = \underset{(p,\,1)}{\mathbf{a}}$.
- Si $\mathbf{A}$ est une matrice *symétrique* de coefficients :

$$\frac{\partial(\mathbf{x}'\,\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\,\mathbf{A}\mathbf{x} \qquad \text{vecteur } (p, 1)$$

$$\frac{\partial^2(\mathbf{x}'\,\mathbf{A}\mathbf{x})}{\partial \mathbf{x}\,\partial \mathbf{x}'} = 2\,\mathbf{A} \qquad \text{matrice } (p, p)\,.$$

10. *PROBLÈMES D'EXTREMA*

On rencontre souvent en calcul statistique le problème suivant : maximiser ou minimiser une fonction $y = f(\mathbf{x})$, où $\mathbf{x}$ est un vecteur-colonne de variables réelles.

Premier résultat. Si toutes les dérivées partielles $\partial f / \partial x_j$ sont continues, la fonction atteint un *extremum* (maximum, minimum, ou point-selle) au point où toutes les dérivées sont nulles, c'est-à-dire au point où le *gradient* s'annule :

$$\frac{\partial f}{\partial \mathbf{x}} = \underset{(p,\,1)}{\mathbf{0}}.$$

Ce résultat généralise le cas d'une fonction réelle d'une variable réelle.

Deuxième résultat. Si les dérivées partielles $\partial^2 f / \partial x_i \, \partial x_j$ sont continues, on a un *minimum* (local en général) au point $\mathbf{x}$ si le gradient s'annule en $\mathbf{x}$, et si de plus la matrice Hessienne est *définie positive* au point $\mathbf{x}$:

$$\frac{\partial^2 f}{\partial \mathbf{x} \, \partial \mathbf{x}'} \text{ définie positive}$$

(un maximum correspondrait à une matrice Hessienne définie négative).

On peut commenter ce résultat en considérant le développement de Taylor de f autour du point $\mathbf{x}_0$; pour $\mathbf{x}$ « proche » de $\mathbf{x}_0$, on a :

$$f(\mathbf{x}) - f(\mathbf{x}_0) = (\mathbf{x} - \mathbf{x}_0)' \frac{\partial f}{\partial \mathbf{x}} + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)' \frac{\partial^2 f}{\partial \mathbf{x} \, \partial \mathbf{x}'} (\mathbf{x} - \mathbf{x}_0) + \cdots$$

Si $\mathbf{x}_0$ localise un extremum, le gradient est nul et le premier terme disparaît. Si la matrice Hessienne est définie positive en $\mathbf{x}_0$, le second terme est une forme quadratique définie positive et par conséquent $f(\mathbf{x}) - f(\mathbf{x}_0)$ est positif, indiquant par là que $\mathbf{x}_0$ correspond à un *minimum* de $f(\mathbf{x})$. Il s'agit d'un minimum *local*, car si $\mathbf{x}$ s'éloigne de $\mathbf{x}_0$, les termes non écrits du développement de Taylor peuvent inverser le signe de $f(\mathbf{x}) - f(\mathbf{x}_0)$. Cependant :

Troisième résultat. Si f est une forme quadratique *définie positive*, c'est-à-dire si f est de la forme :

$$f(\mathbf{x}) = \mathbf{x}' \mathbf{A} \mathbf{x} \qquad (\text{avec } \mathbf{A} \text{ symétrique})$$

alors l'extremum est un *minimum absolu* (ce serait un maximum absolu pour une forme quadratique définie négative). En effet, dans ce cas les dérivées partielles d'ordre supérieur à 2 sont identiquement nulles, et le développement de Taylor s'arrête aux 2 premiers termes écrits plus haut.

Extremum sous contrainte. Il s'agit de maximiser (ou de minimiser) la fonction $f(\mathbf{x})$ sous la contrainte $g(\mathbf{x}) = 0$. On construit le lagrangien :

$$\mathscr{L} = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

où λ est un paramètre inconnu, appelé multiplicateur de Lagrange. Une condition nécessaire d'extremum de $f(\mathbf{x})$, qui satisfasse à la contrainte $g(\mathbf{x}) = 0$, est l'annulation simultanée des dérivées partielles premières de $\mathscr{L}$ par rapport aux x_i.