Un rappel sur les matrices

Patrice Wira

Université de Haute-Alsace
Faculté des **S**ciences et **T**echniques

2000 - 2001

Sommaire

Calculs matriciels (matrix algebra)

1. Représentation matricielle et notations

Une matrice est un tableau rectangulaire d'éléments, généralement des nombres ou des fonctions. Ces grandeurs sont généralement des réels ou des complexes. Dans la suite, nous ne considérerons que des grandeurs réelles. Une matrice $\mathbf{A}$ de dimension $m*n$ est notée $\mathbf{A} \in \mathbb{R}^{m*n}$. Cette matrice est une matrice de m lignes et de n colonnes :

$$\mathbf{A} = \left[a_{ij} \right] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

Une matrice $\mathbf{V}$ qui ne comporte qu'une seule colonne, $\mathbf{V} \in \mathbb{R}^{m*1}$, est appelé un vecteur colonne :

$$\mathbf{V} = \begin{bmatrix} v_1 \\ \vdots \\ v_2 \end{bmatrix}.$$

Une matrice $\mathbf{V}$ qui ne comporte qu'une seule ligne, $\mathbf{V} \in \mathbb{R}^{1*n}$, est appelé un vecteur ligne :

$$\mathbf{V} = \begin{bmatrix} v_1 & \cdots & v_2 \end{bmatrix}.$$

Par convention, tout vecteur est désigné comme une matrice colonne.

Si $m = n$, alors la matrice est carrée.

2. L'addition

L'addition matricielle

L'addition matricielle n'est définie qu'entre deux matrices de même dimensions. La matrice résultante est de la même dimension que les matrices additionnées et chacun de ses éléments est la somme des éléments des deux matrices correspondant à la même ligne et à la même colonne.

Soient $\mathbf{A} \in \mathbb{R}^{m*n}$ et $\mathbf{B} \in \mathbb{R}^{m*n}$. L'addition de ces deux matrices est donnée par :

$$\mathbf{C} = \left[c_{ij} \right] = \left[a_{ij} + b_{ij} \right].$$

On la note : $\mathbf{C} = \mathbf{A} + \mathbf{B}$, $\mathbf{C} \in \mathbb{R}^{m*n}$.

Il en va de même pour la soustraction, au signe près. La soustraction des matrices $\mathbf{A}$ et $\mathbf{B}$ est donnée par :

$$\mathbf{C} = \left[c_{ij} \right] = \left[a_{ij} - b_{ij} \right].$$

On la note : $\mathbf{C} = \mathbf{A} - \mathbf{B}$, $\mathbf{C} \in \mathbb{R}^{m*n}$.

Propriétés de l'addition

En considérant trois matrices, $\mathbf{A} \in \mathbb{R}^{m*n}$, $\mathbf{B} \in \mathbb{R}^{m*n}$ et $\mathbf{C} \in \mathbb{R}^{m*n}$, on a les trois propriétés suivantes :

$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$	Commutativité (commutative law).
$\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB}$	Distributivité (distributive law).
$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$	Associativité (associative law).

Exemples : l'addition et la soustraction matricielle

Si $\mathbf{A} \in \mathbb{R}^{2*3}$ et $\mathbf{B} \in \mathbb{R}^{2*3}$ tel que $\mathbf{A} = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 7 & 3 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 5 & 2 & 6 \\ 0 & 1 & 1 \end{bmatrix}$, alors :

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1+5 & 4+2 & 0+6 \\ 2+0 & 7+1 & 3+1 \end{bmatrix} = \begin{bmatrix} 6 & 6 & 6 \\ 2 & 8 & 4 \end{bmatrix} \text{ et}$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} -4 & 2 & -6 \\ 2 & 6 & 2 \end{bmatrix}.$$

Exemples : l'addition et la combinaison linéaire de deux vecteurs

L'addition de deux vecteurs (de dimensions 2, $\mathbf{v},\mathbf{w}\in\mathbb{R}^2$) permet de définir un parallélogramme. Le vecteur résultant $\mathbf{u}$ est une des diagonale du parallélogramme.

Prenons par exemple $\mathbf{v}=\begin{bmatrix}4\\2\end{bmatrix}$ et $\mathbf{w}=\begin{bmatrix}-1\\3\end{bmatrix}$, on obtient alors :

$$\mathbf{u}=\mathbf{v}+\mathbf{w}=\begin{bmatrix}4\\2\end{bmatrix}+\begin{bmatrix}-1\\3\end{bmatrix}=\begin{bmatrix}3\\5\end{bmatrix}$$

Une combinaison linéaire de deux vecteurs $\mathbf{v}\in\mathbb{R}^2$ et $\mathbf{w}\in\mathbb{R}^2$ est l'addition de deux vecteurs à qui on fait subir un changement de longueur par des coefficients scalaires.

Reprenons les vecteurs exemple $\mathbf{v}=\begin{bmatrix}4\\2\end{bmatrix}$ et $\mathbf{w}=\begin{bmatrix}-1\\3\end{bmatrix}$:

$$\mathbf{u}=-2\mathbf{v}+\mathbf{w}=-2\begin{bmatrix}4\\2\end{bmatrix}+\begin{bmatrix}-1\\3\end{bmatrix}$$
$$=\begin{bmatrix}-8\\-4\end{bmatrix}+\begin{bmatrix}-1\\3\end{bmatrix}=\begin{bmatrix}-9\\-1\end{bmatrix}=-\begin{bmatrix}9\\1\end{bmatrix}$$

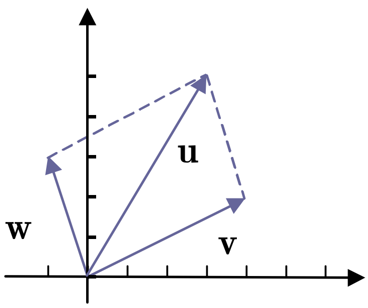

Figure 1: L'addition de deux vecteurs.

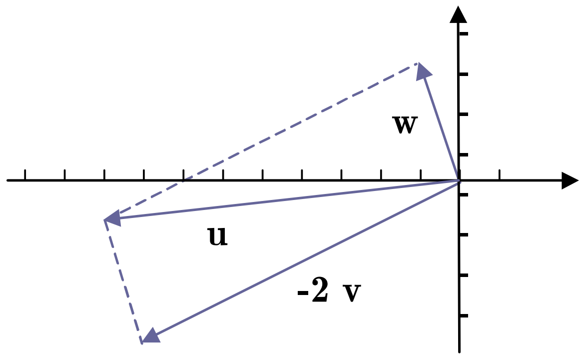

Figure 2 : La soustraction de deux vecteurs.

3. *Le produit*

La multiplication d'une matrice avec un scalaire

La multiplication entre une matrice et un nombre scalaire donne une matrice dont chaque élément de la matrice est multiplié par le scalaire.

Etant donné $\mathbf{A}\in\mathbb{R}^{m*n}$ une matrice, et b un scalaire, alors les éléments de la matrice $\mathbf{C}$ résultante sont donnés par :

$$c_{ij}=ba_{ik}\,.$$

La matrice $\mathbf{C}=b\mathbf{A}$ est de même dimension que $\mathbf{A}$, $\mathbf{C}\in\mathbb{R}^{m*n}$.

Propriétés de la multiplication entre une matrice et un scalaire

Si c est un scalaire, alors :

$c\mathbf{A} = \mathbf{A}c$	Commutativité, $\mathbf{A} \in \mathbb{R}^{m*n}$.
$c(\mathbf{A}+\mathbf{B}) = c\mathbf{A} + c\mathbf{B}$	Distributivité, $\mathbf{A} \in \mathbb{R}^{m*n}$ et $\mathbf{B} \in \mathbb{R}^{m*n}$.
$c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B}) = (\mathbf{AB})c$	Associativité, où $\mathbf{AB}$ est la multiplication matricielle entre les matrices $\mathbf{A} \in \mathbb{R}^{m*n}$ et $\mathbf{B} \in \mathbb{R}^{n*p}$.

Si c et b sont des scalaires, alors :

$(b+c)\mathbf{A} = b\mathbf{A} + c\mathbf{A}$	Distributivité, $\mathbf{A} \in \mathbb{R}^{m*n}$.
$(bc)\mathbf{A} = b(c\mathbf{A})$	Associativité, $\mathbf{A} \in \mathbb{R}^{m*n}$.

La multiplication matricielle

La multiplication entre deux matrices n'est définie que lorsque leurs dimensions son compatibles : le nombre de colonnes de la matrice à gauche de l'opérateur doit correspondre au nombre de lignes de la matrice à droite de l'opérateur.
Si $\mathbf{A} \in \mathbb{R}^{m*n}$, et si $\mathbf{B} \in \mathbb{R}^{n*p}$, la multiplication entre les matrices $\mathbf{A}$ et $\mathbf{B}$ donne une matrice $\mathbf{C}$ de dimensions $m*p$ telle que tous ses éléments :

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

On note cette opération : $\mathbf{C} = \mathbf{A} * \mathbf{B} = \mathbf{AB}$, . .

Propriétés élémentaires de la multiplication matricielle

$\mathbf{AB} \neq \mathbf{BA}$	La commutativité n'est pas toujours vraie (the commutative law is usually broken).
$\mathbf{C}(\mathbf{A}+\mathbf{B}) = \mathbf{CA} + \mathbf{CB}$	Distributivité à gauche (distributive law from the left) avec $\mathbf{A}$, $\mathbf{B}$ $\in \mathbb{R}^{m*n}$ et $\mathbf{C}$ $\in \mathbb{R}^{p*m}$.

$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$

Distributivité à droite
(distributive law from the right)
avec $\mathbf{A}$, $\mathbf{B}$ $\in \mathbb{R}^{m*n}$ et $\mathbf{C}$ $\in \mathbb{R}^{n*p}$.

$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

Associativité (associative law)
avec $\mathbf{A}$ $\in \mathbb{R}^{m*n}$, $\mathbf{B}$ $\in \mathbb{R}^{n*p}$ et $\mathbf{C}$ $\in \mathbb{R}^{p*q}$.

Quelques exemples

- Si $\mathbf{A} \in \mathbb{R}^{2*2}$ et $\mathbf{B} \in \mathbb{R}^{2*3}$ tel que $\mathbf{A} = \begin{bmatrix} 5 & -3 \\ 4 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 8 & -2 & 6 \\ 7 & 0 & 9 \end{bmatrix}$, alors

$$\mathbf{AB} = \begin{bmatrix} 5 & -3 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} 8 & -2 & 6 \\ 7 & 0 & 9 \end{bmatrix} = \begin{bmatrix} 5(8)+(-3)(7) & 5(-2)+(-3)0 & 5(6)+(-3)(9) \\ 4(8)+2(7) & 4(-2)+2(0) & 4(6)+2(9) \end{bmatrix} = \begin{bmatrix} 19 & -10 & 3 \\ 46 & -8 & 42 \end{bmatrix}$$

- Si on dispose d'un vecteur ligne $\mathbf{x} \in \mathbb{R}^4$ et d'un vecteur colonne $\mathbf{y} \in \mathbb{R}^4$ tel que $\mathbf{x} = \begin{bmatrix} 1 & 3 & 5 & -1 \end{bmatrix}$ et $\mathbf{y} = \begin{bmatrix} -2 & 4 & 2 & 6 \end{bmatrix}^T$, alors

$$\mathbf{xy} = \begin{bmatrix} 1 & 3 & 5 & -1 \end{bmatrix}\begin{bmatrix} -2 \\ 4 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 1(-2)+3(4)+5(2)+(-1)(6) \end{bmatrix} = 14.$$

Un produit matriciel particulier : le produit de Hadamard

La matrice résultante du produit de Hadamard entre deux matrices de même taille contient le résultat d'une multiplication élément par élément.

On note le produit entre les matrices $\mathbf{A} \in \mathbb{R}^{m*n}$ et $\mathbf{B} \in \mathbb{R}^{m*n}$:

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B}, \ \mathbf{C} \in \mathbb{R}^{m*n},$$
$$\mathbf{C} = \left[c_{ij} \right] = \left[a_{ij} b_{ij} \right].$$

4. ***<u>Les produits vectoriels</u>***

Si l'addition de vecteur n'est pas très différente de l'addition matricielle, il n'en va pas de même en ce qui concerne la multiplication. En effet, multiplier deux vecteurs n'a pas la même

signification que multiplier deux matrices, c'est pourquoi on distingue les produits vectoriels des produits matriciels.
La multiplication de deux vecteurs peut être définie de trois façon différentes, selon que le résultat est un scalaire, un vecteur ou une matrice. On les appelle respectivement produit interne (scalaire), vectoriel et produit externe.
Les produits interne et externe sont des cas particulier de la multiplication matricielle définie en toute généralité précédemment.

Le produit interne

Soient deux vecteurs de même dimension, $\mathbf{x} \in \mathbb{R}^n$ et $\mathbf{y} \in \mathbb{R}^n$. On appelle produit interne (en Anglais on parlera de « inner product » ou de « vector dot product ») la somme des multiplications éléments par éléments de chaque vecteur :

$$\left\langle \mathbf{x}^T, \mathbf{y} \right\rangle = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = \sum_{i=1}^{n} y_i x_i \ .$$

Cette opération retourne un scalaire. Le produit interne est aussi appelé produit scalaire, il n'est défini que si les deux vecteurs sont de même taille.
Le produit interne est un moyen de mesurer comment deux vecteurs sont parallèles. Deux vecteurs sont orthogonaux si leur produit scalaire est nul.

Remarque : angle en deux vecteurs.
Si deux vecteurs qui ne sont pas parallèles forment un angle entre eux.
Si $\mathbf{x}$ et $\mathbf{y}$ sont deux vecteurs Euclidiens, alors ils forment un angle θ définit par :

$$\cos\theta = \frac{\left\langle \mathbf{x}, \mathbf{y} \right\rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \ .$$

Les deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ sont dit orthogonaux quand leur produit interne est nul (θ =90°), on le note $\mathbf{x} \perp \mathbf{y}$.

Remarque : parfois le produit entre deux matrices en également appelé produit scalaire, mais l'ordre de $\mathbf{x}$ et $\mathbf{y}$ est important : $\mathbf{x}^T \mathbf{y} \neq \mathbf{y}^T \mathbf{x}$ qui donne un nombre, et qu'il faut différencier de $\mathbf{y}\mathbf{x}^T$ (produit externe) qui est une matrice.

Le produit vectoriel

Le produit vectoriel de deux vecteur se traduit en Anglais par « vector product » ou « cross product ». Le produit vectoriel de deux vecteurs de trois composantes est défini par :

$$\mathbf{x} \times \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{bmatrix}.$$

Formellement, ce produit peut être considéré comme un déterminant, avec les vecteurs $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ constituants la base dans la première colonne :

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{i} & x_1 & y_1 \\ \mathbf{j} & x_2 & y_2 \\ \mathbf{k} & x_3 & y_3 \end{vmatrix} = \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \mathbf{k},$$

$$\mathbf{x} \times \mathbf{y} = (x_2 y_3 - x_3 y_2)\mathbf{i} + (x_3 y_1 - x_1 y_3)\mathbf{j} + (x_1 y_2 - x_2 y_1)\mathbf{k}.$$

Remarque : ce n'est pas un vrai déterminant dans le sens où il ne retourne pas une grandeur scalaire.

Le produit mixte

Soit trois vecteurs $\mathbf{x}$, $\mathbf{y}$ et $\mathbf{z} \in \mathbb{R}^n$. Le nombre réel $(\mathbf{x} \times \mathbf{y}).\mathbf{z}$, c'est à dire le produit scalaire du vecteur $\mathbf{x} \times \mathbf{y}$ par le vecteur $\mathbf{z}$, s'appelle le produit mixte des trois vecteurs $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ (dans cet ordre). Le résultat est un scalaire. On parle en Anglais de « scalar triple product ».

Le produit externe

Considérons deux vecteurs de dimensions différentes, $\mathbf{x} \in \mathbb{R}^m$ et $\mathbf{y} \in \mathbb{R}^n$. On appelle produit externe (traduit en Anglais par « outer product » ou « dyadic product ») :

$$\rangle \mathbf{x}, \mathbf{y} \langle = \mathbf{x}\mathbf{y}^T = \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & x_i y_j & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{bmatrix}, \text{ avec } i = 1,...,m,\ j = 1,...,n.$$

$$\mathbf{z} = \left[z_{ij} \right] = \mathbf{x}\mathbf{y}^T = \left[x_i y_j \right], \text{ avec } i = 1,...,m,\ j = 1,...,n.$$

Cette opération retourne une matrice de m lignes et n colonnes. Ce produit peut être défini comme un produit matricielle.

Le produit externe joue un rôle important lorsqu'il s'agit de déterminer comment sont corrélés les éléments d'un vecteur avec ceux d'un autre.

Remarque : si $\mathbf{x} = [1 \quad 2 \quad 3 \quad \ldots \quad 10]^T$ contient toutes les valeurs possibles entre 1 et 10, alors $\mathbf{z} = \mathbf{x}\mathbf{x}^T$ contient les tables de multiplication jusqu'à 10 :

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 & 27 & 30 \\ 4 & 8 & 12 & 16 & 20 & 24 & 28 & 32 & 36 & 40 \\ 5 & 10 & 15 & 20 & 25 & 30 & 35 & 40 & 45 & 50 \\ 6 & 12 & 18 & 24 & 30 & 36 & 42 & 48 & 54 & 60 \\ 7 & 14 & 21 & 28 & 35 & 42 & 49 & 56 & 63 & 70 \\ 8 & 16 & 24 & 32 & 40 & 48 & 56 & 64 & 72 & 80 \\ 9 & 18 & 27 & 36 & 45 & 54 & 63 & 72 & 81 & 90 \\ 10 & 20 & 30 & 40 & 50 & 60 & 70 & 80 & 90 & 100 \end{bmatrix}.$$

Propriétés des produits scalaire et vectoriel

Soient $\mathbf{x}$, $\mathbf{y}$ et $\mathbf{z}$ trois vecteurs $\in \mathbb{R}^n$ et soit a un scalaire.

Le produit scalaire est distributif : $\mathbf{x}.(\mathbf{y}+\mathbf{z}) = \mathbf{x}.\mathbf{y} + \mathbf{x}.\mathbf{z}$.

est commutatif : $\mathbf{x}.\mathbf{y} = \mathbf{y}.\mathbf{x}$.

est associatif : $(a\mathbf{x}).\mathbf{y} = a(\mathbf{x}.\mathbf{y})$.

Le produit vectoriel est distributif : $\mathbf{x} \times (\mathbf{y}+\mathbf{z}) = \mathbf{x} \times \mathbf{y} + \mathbf{x} \times \mathbf{z}$.

N'est PAS commutatif : $\mathbf{x} \times \mathbf{y} = -\mathbf{y} \times \mathbf{x}$.

est associatif : $(a\mathbf{x}) \times \mathbf{y} = a(\mathbf{x} \times \mathbf{y})$.

Interprétation géométrique

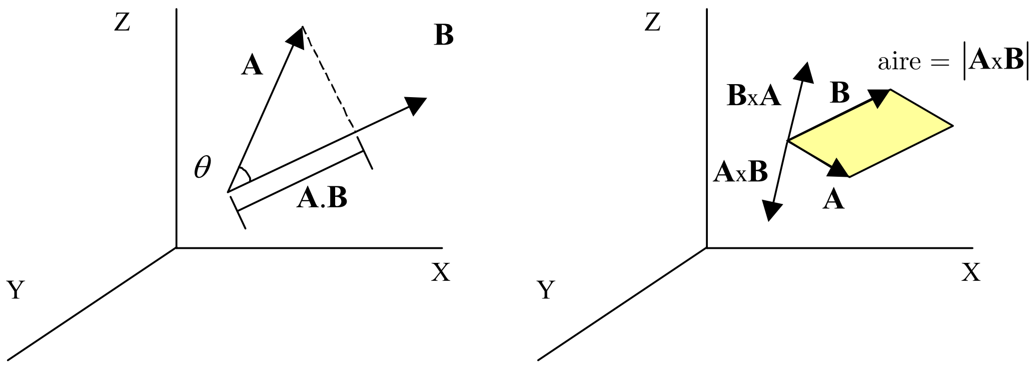

Figure 3: Les produits scalaire (à gauche) et vectoriel (à droite) de deux vecteurs.

Le produit scalaire :

Le scalaire qui est retourné est proportionnel à la projection du premier vecteur sur le second. Cet opération est fréquemment utilisée pour mesurer l'orthogonalité de deux vecteurs mais aussi pour décrire la composante d'un vecteur dans une direction particulière.

Le produit vectoriel :

Le produit vectoriel, noté $\mathbf{A} \times \mathbf{B}$ est un vecteur orthogonal à $\mathbf{A}$ et à $\mathbf{B}$. Il est nul si $\mathbf{A}$ et $\mathbf{B}$ sont parallèles. Il est utilisé pour déterminer la normale de deux droites. C'est ainsi qu'on défini le vecteur normal d'une surface en un point quelconque.

Le produit mixte :

La valeur absolue du produit mixte est le produit de la longueur de la projection du vecteur $\mathbf{z}$ sur la perpendiculaire au plan de même direction que les vecteurs $\mathbf{x}$ et $\mathbf{y}$ sur la surface du parallélogramme construit sur les vecteurs $\mathbf{x}$ et $\mathbf{y}$. Autrement dit, c'est le volume du parallélépipède construit sur les trois vecteurs $\mathbf{x}$, $\mathbf{y}$ et $\mathbf{z}$. On dit aussi que le produit mixte représente le volume algébrique de ce parallélépipède.

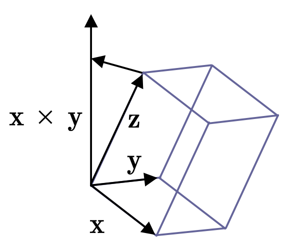

Figure 4: Le produit mixte de trois vecteurs.

Exemples :

Considérons un espace de dimensions trois défini par les vecteurs $\mathbf{i}$, $\mathbf{j}$ et $\mathbf{k}$.

Les vecteurs $\mathbf{x}$ et $\mathbf{y}$ s'écrivent par rapport à cette base :

$$\mathbf{x} = x_i\mathbf{i} + x_j\mathbf{j} + x_k\mathbf{k} \text{ et } \mathbf{y} = y_i\mathbf{i} + y_j\mathbf{j} + y_k\mathbf{k}.$$

Le produit scalaire de $\mathbf{x}$ et $\mathbf{y}$ vaut alors :

$$\mathbf{x}.\mathbf{y} = |\mathbf{x}||\mathbf{y}|\cos(\mathbf{x},\mathbf{y}) = x_i y_i + x_j y_j + x_k y_k.$$

Le produit vectoriel vaut quand à lui

$$\mathbf{x}\text{x}\mathbf{y} = (x_j y_k - x_k y_j)\mathbf{i} + (x_k y_i - x_i y_k)\mathbf{j} + (x_i y_j - x_j y_i)\mathbf{k},$$

$$\text{et } |\mathbf{x}\text{x}\mathbf{y}| = |\mathbf{x}||\mathbf{y}|\sin(\mathbf{x},\mathbf{y}).$$

On rappelle que $|\mathbf{x}| = (x_i^2 + x_j^2 + x_k^2)^{1/2}$.

5. *L'opérateur de transposition*

La transposée d'une matrice $\mathbf{A}$ est la matrice $\mathbf{A}^T$ (notée parfois aussi $\mathbf{A}'$) définie par :

$$\mathbf{A} \in \mathbb{R}^{m*n},\ \mathbf{A}^T \in \mathbb{R}^{n*m},$$

$$\mathbf{A} = \left[a_{ij}\right]_{m*n},\ \mathbf{A}^T = \left[a_{ji}\right]_{n*m}.$$

Propriétés élémentaires de la transposition

$(\mathbf{A}^T)^T = \mathbf{A}$ où $\mathbf{A} \in \mathbb{R}^{m*n}$.

$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$ avec $\mathbf{A}$, $\mathbf{B} \in \mathbb{R}^{m*n}$.

$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ où $\mathbf{A} \in \mathbb{R}^{m*n}$, $\mathbf{B} \in \mathbb{R}^{n*p}$ et $\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right]_{m*p}$.

6. *Matrices particulières*

La matrice nulle

La matrice dont tous les termes sont nuls est appelée matrice nulle. On la note $\mathbf{A} = \mathbf{0}$: $a_{ij} = 0$, $\forall i$, $\forall j$.

Matrice diagonale

Une matrice diagonale $\mathbf{A} \in \mathbb{R}^{n*n}$ est une matrice carrée dont tous les éléments non diagonaux sont nuls, c'est à dire que $a_{ij} = 0$, $\forall i \neq j$. Elle est de la forme :

$$\mathbf{A} = \left[a_{ij}\right] = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & a_{ij} & 0 \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}.$$

La matrice identité

Une matrice identité est une matrice diagonale $\mathbf{A} \in \mathbb{R}^{n*n}$ (donc carrée) qui voit tous ses éléments nuls sauf ceux de sa diagonale principale qui sont unitaires. On la note I_n ou $\mathbf{I}$:

$$\mathbf{A} = \mathbf{I} \Leftrightarrow \begin{cases} a_{ij} = 0 & \forall i \neq j \\ a_{ii} = 1 & i = 1,...,n \end{cases}.$$

Matrice triangulaire

Une matrice triangulaire supérieure (respectivement inférieure) est une matrice carrée $\mathbf{A} \in \mathbb{R}^{n*n}$ dont les éléments qui se trouvent au-dessous (respectivement au-dessus) de la diagonale principale sont nuls, c'est à dire telle que $a_{ij} = 0$, $\forall i > j$ ($\forall i < j$). Elle est donc de la forme :

$$\mathbf{A} = \left[a_{ij}\right] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & & \vdots \\ \vdots & & & \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} \text{ (respectivement } \mathbf{A} = \left[a_{ij}\right] = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & & \vdots \\ \vdots & & & 0 \\ a_{n1} & \cdots & & a_{nn} \end{bmatrix}\text{).}$$

Matrices symétrique et antisymétrique

Soit $\mathbf{A} \in \mathbb{R}^{n*n}$ une matrice carrée.

- Si $\mathbf{A}^T = \mathbf{A}$, alors $\mathbf{A}$ est dite matrice symétrique.
- Si $\mathbf{A}^T = -\mathbf{A}$, alors $\mathbf{A}$ est dite matrice antisymétrique.

Matrice idempotente

Considérons une matrice carrée $\mathbf{A} \in \mathbb{R}^{n*n}$. Si $\mathbf{A}^T\mathbf{A} = \mathbf{A}$, alors $\mathbf{A}$ est dite idempotente.

Propriétés de la matrice nulle

Soit $\mathbf{A}$ une matrice, $\mathbf{A} \in \mathbb{R}^{m*n}$:

$$\mathbf{A0} = \mathbf{0A} = \mathbf{0}$$ C'est l'opérateur nul de la multiplication

$$\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$$ C'est l'opérateur neutre de l'addition

Propriété de la matrice identité

Soit $\mathbf{A}$ une matrice, $\mathbf{A} \in \mathbb{R}^{m*n}$:

$$\mathbf{AI} = \mathbf{IA} = \mathbf{I}$$ C'est l'opérateur neutre de la multiplication

7. Matrices partitionnées

La composante de base d'une matrice $\mathbf{A}$ de dimensions quelconques est a_{ij}. Il est parfois intéressant de considérer une matrice en tant que tableau de matrices élémentaires.

Soit :

$$\mathbf{A} = \left[a_{ij} \right] = \begin{array}{cc} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} & \begin{matrix} m \\ n \end{matrix} \\ \begin{matrix} p & \quad q \end{matrix} & \end{array},$$

avec $\mathbf{A}_{11} \in \mathbb{R}^{m*p}$, $\mathbf{A}_{12} \in \mathbb{R}^{m*q}$, $\mathbf{A}_{21} \in \mathbb{R}^{n*p}$, $\mathbf{A}_{22} \in \mathbb{R}^{n*q}$.

La matrice $\mathbf{A}$ est une matrice de $m+n$ lignes et de $p+q$ colonnes, $\mathbf{A} \in \mathbb{R}^{(m+n)*(p+q)}$.

Considérons une seconde matrice $\mathbf{B}$ partitionnée ainsi : $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}$, où $\mathbf{B}_1 \in \mathbb{R}^{p*1}$, $\mathbf{B}_2 \in \mathbb{R}^{q*1}$.

On peut alors écrire :

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_1 + \mathbf{A}_{12}\mathbf{B}_2 \\ \mathbf{A}_{21}\mathbf{B}_1 + \mathbf{A}_{22}\mathbf{B}_2 \end{bmatrix}.$$

Cette relation est souvent utilisée avec les notations suivantes :

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{E} \\ \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{AE} + \mathbf{BF} \\ \mathbf{CE} + \mathbf{DF} \end{bmatrix}.$$

Toutes les sous-matrices doivent comporter des dimensions compatibles avec les règles du produit matriciel.

La transposée d'une matrice partionnée est donnée par :

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^T = \begin{bmatrix} \mathbf{A}^T & \mathbf{C}^T \\ \mathbf{B}^T & \mathbf{D}^T \end{bmatrix}.$$

8. *Le déterminant d'une matrice*

On appelle déterminant d'une matrice $\mathbf{A}$ carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, le nombre noté $\det(\mathbf{A})$ ou $|\mathbf{A}|$ et égal à :

$$\det(\mathbf{A}) = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(\mathbf{A}_i)$$

où $\mathbf{A}_i$ est la matrice obtenue en rayant la 1ère colonne et la i-ième ligne.

Si $n = 2$, $\det(\mathbf{A}) = \left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| = a_{11}a_{22} - a_{12}a_{21}$.

Si $n = 3$, $\det(\mathbf{A}) = a_{11}a_{22}a_{33} + a_{13}a_{32}a_{21} + a_{12}a_{23}a_{31} - a_{11}a_{23}a_{32} - a_{22}a_{13}a_{31} - a_{33}a_{12}a_{21}$.

Lorsque le déterminant d'une matrice est nul, on dit que la matrice est singulière.

Remarques :

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{kj} c_{kj}$$ pour tous les lignes k de $\mathbf{A}$.

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a_{il} c_{il}$$ pour tous les colonnes l de $\mathbf{A}$.

Quelques propriétés du déterminant d'une matrice :

$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ avec $\mathbf{A} \in \mathbb{R}^{n*n}$.

$\det(\mathbf{A}^{-1}) = \dfrac{1}{\det(\mathbf{A})}$ avec $\mathbf{A} \in \mathbb{R}^{n*n}$.

$$\det(\mathbf{A}) = \det\left(\begin{bmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{F} & \mathbf{G} \end{bmatrix}\right) = \det(\mathbf{G})\det(\mathbf{D} - \mathbf{E}\mathbf{G}^{-1}\mathbf{F}) = \det(\mathbf{D})\det(\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})$$

où $\mathbf{A} = \begin{bmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{F} & \mathbf{G} \end{bmatrix} \in \mathbb{R}^{n*n}$, $\mathbf{D} \in \mathbb{R}^{m*m}$ et les dimensions des matrices $\mathbf{E}$, $\mathbf{F}$ et $\mathbf{G}$ sont en accord avec celles de $\mathbf{A}$ et $\mathbf{D}$; $\mathbf{D}^{-1}$ doit exister

Quelques théorèmes

1. Si tous les éléments d'une ligne (colonne) d'une matrice $\mathbf{A} \in \mathbb{R}^{n*n}$ sont nuls alors $\det(\mathbf{A}) = 0$.
2. Si tous les éléments d'une ligne (colonne) du déterminant d'une matrice $\mathbf{A} \in \mathbb{R}^{n*n}$ sont multiplié par un scalaire k, alors le déterminant est multiplié par k.
3. Si $\mathbf{B}$ est obtenue à partir de $\mathbf{A} \in \mathbb{R}^{n*n}$ en échangeant deux de ces lignes (colonnes), alors $\det(\mathbf{B}) = -\det(\mathbf{A})$.
4. Si $\mathbf{B}$ est obtenue à partir de $\mathbf{A} \in \mathbb{R}^{n*n}$ en faisant passer la i-ème ligne (colonne) par dessus p lignes (colonnes), alors $\det(\mathbf{B}) = (-1)^p \det(\mathbf{A})$.

5. Si deux lignes (colonnes) de $\mathbf{A} \in \mathbb{R}^{n*n}$ sont identiques, alors $\det(\mathbf{A}) = 0$.
6. Si, aux éléments d'une ligne (colonne) on ajoute *k* fois les éléments correspondants d'une autre ligne (colonne), la valeur du déterminant reste inchangée.

Pour tous ces théorèmes, on trouvera la démonstration dans (Ayres, 1991).

Le calcul du déterminant d'une matrice $\mathbf{A} \in \mathbb{R}^{3*3}$ peut être calculé par la règle dite de Sarrus (Christol et al. 1996, p. 108).

9. Les mineurs, les cofacteurs et la matrice adjointe

Les mineurs

Les mineurs m_{ij} des éléments a_{ij} d'une matrice $\mathbf{A}$ carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, sont les déterminants de la partie restante de $\mathbf{A}$ lorsqu'on ne tient pas compte de la ligne *i* et de la colonne *j*.

Les mineurs directeurs

Les mineurs directeurs d'une matrice $\mathbf{A}$ carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, appelés aussi mineurs principaux (en Anglais « leading minors ») sont définis comme suit :

$$
\begin{aligned}
m_1 &= a_{11} \\
m_2 &= \det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) \\
m_3 &= \det\left(\begin{bmatrix} a_{11} & \cdots \\ \cdots & a_{33} \end{bmatrix}\right) \\
&\ldots \\
m_n &= \det(\mathbf{A})
\end{aligned}
$$

Les cofacteurs

Les cofacteurs c_{ij} des éléments a_{ij} d'une matrice $\mathbf{A}$ carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, sont donnés par :

$$\mathrm{c}_{ij} = (-1)^{i+j} m_{ij} .$$

Le cofacteur est souvent aussi noté Δ_{ij}.

La matrices adjointe

La matrice des cofacteurs c_{ij} des éléments a_{ij} d'une matrice $\mathbf{A}$ carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, lorsqu'elle est transposée, est appelé la matrice adjointe de $\mathbf{A}$. On note généralement cette matrice $\tilde{\mathbf{A}}$ ou encore $\mathrm{adj}(\mathbf{A})$.

$$\tilde{\mathbf{A}} = \mathrm{adj}(\mathbf{A}) = \left[c_{ij} \right]^T = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & c_{ij} & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}^T,$$

$$\mathrm{adj}(\mathbf{A}) = \mathbf{C}^T \text{, avec } \mathbf{C} = \left[c_{ij} \right]$$

Propriétés

$\tilde{\mathbf{A}}\,\mathbf{A} = \mathbf{A}\,\tilde{\mathbf{A}} = |\mathbf{A}|\mathbf{I}$ où $|\mathbf{A}|$ est le déterminant de la matrice $\mathbf{A}$.

$\mathrm{adj}(\mathbf{AB}) = \mathrm{adj}(\mathbf{A}).\mathrm{adj}(\mathbf{B})$.

Exemples

En ce qui concerne la matrice $\mathbf{A} \in \mathbb{R}^{3*3}$, $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, nous avons le mineur $m_{23} = \left| \begin{bmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{bmatrix} \right|$ et le cofacteur $c_{ij} = (-1)^{2+3} m_{23} = -\left| \begin{bmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{bmatrix} \right|$.

Calcul du déterminant d'une matrice avec les cofacteurs

La valeur d'un déterminant d'ordre n d'une matrice carrée $\mathbf{A} \in \mathbb{R}^{n*n}$ est la somme des n produits obtenus en multipliant chaque élément d'une ligne (colonne) donnée de la matrice par son cofacteur.

On peut donc, par exemple, calculer le déterminant d'ordre 3 de la matrice $\mathbf{A} \in \mathbb{R}^{3*3}$, $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ par un développement selon la seconde colonne :

$$\det(\mathbf{A}) = a_{12}c_{12} + a_{22}c_{22} + a_{32}c_{32}$$

$$= -a_{12} \left| \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} \right| + a_{22} \left| \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} \right| - a_{32} \left| \begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix} \right|.$$

En choisissant judicieusement la ligne (ou la colonne) par laquelle en effectue le développement, on peut très largement simplifier les calculs. On cherchera notamment à effectuer les développement selon les lignes ou les colonnes qui comportent le plus de valeurs nulles.

Ainsi, si $\mathbf{A} = \begin{bmatrix} 3 & 5 & 8 \\ 1 & 0 & 2 \\ 4 & 0 & 3 \end{bmatrix}$, alors $\det(\mathbf{A}) = -5\left|\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}\right| = -5\{1(3) - 2(4)\} = 25$.

Calcul de la solution d'un système d'équations linéaires - Règle de Cramer

On peut déterminer la solution d'équations linéaires par les déterminants. On appelle cette méthode la règle de Cramer.

Soit le système de trois équations linéaires à trois inconnues x_1, x_2 et x_3 :

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Soit $|\mathbf{A}|$ le déterminant de la matrice $\mathbf{A} \in \mathbb{R}^{3*3}$.

La valeur numérique des coefficients de $|\mathbf{A}|$ est multipliée par x_1 si chaque élément de la première colonne est multipliée par x_1 (théorème 2) :

$$x_1 |\mathbf{A}| = \left|\begin{bmatrix} x_1 a_{11} & a_{12} & a_{13} \\ x_1 a_{21} & a_{22} & a_{23} \\ x_1 a_{31} & a_{32} & a_{33} \end{bmatrix}\right|$$

En ajoutant à chaque élément de la première colonne de ce dernier déterminant, x_2 fois l'élément correspondant d la seconde colonne et x_3 fois l'élément de la troisième colonne (théorème 6), on obtient :

$$x_1 |\mathbf{A}| = \left|\begin{bmatrix} x_1 a_{11} + x_2 a_{12} + x_3 a_{13} & a_{12} & a_{13} \\ x_1 a_{21} + x_2 a_{22} + x_3 a_{23} & a_{22} & a_{23} \\ x_1 a_{31} + x_2 a_{32} + x_3 a_{33} & a_{32} & a_{33} \end{bmatrix}\right| = \left|\begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix}\right|,$$

c'est à dire

$$x_1 = \frac{\left|\begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix}\right|}{|\mathbf{A}|}, \text{ à condition que } |\mathbf{A}| \neq 0.$$

Il en va de même pour x_2 et x_3 :

$$x_2 = \frac{\left|\begin{bmatrix} a_{11} & b_1 a_{12} & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{bmatrix}\right|}{|\mathbf{A}|},$$

$$x_3 = \frac{\left|\begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{bmatrix}\right|}{|\mathbf{A}|}.$$

Cette règle peut être appliquée à n'importe quel système de n équations linéaires à n inconnues, pourvu que le déterminant des coefficients a_{ij} soit différent de zéro.

10. Matrice inverse et pseudo-inverse

Deux matrices $\mathbf{A}$ et $\mathbf{B}$ sont inverses si leur produit est égal à la matrice identité : $\mathbf{AB} = \mathbf{I}$, alors $\mathbf{B} = \mathbf{A}^{-1}$.

Les matrices inverse, et plus généralement les pseudo-inverses, trouvent leurs applications à la résolution des systèmes d'équations linéaires quelles que soient leurs dimensions :

$$\mathbf{y} = \mathbf{Ax},$$

$\mathbf{A} \in \mathbb{R}^{m*n}$, $\mathbf{y} \in \mathbb{R}^m$ est le vecteur cherché, $\mathbf{x} \in \mathbb{R}^n$ est le vecteur des connaissances.

L'inverse généralisée d'un tel système est noté $\mathbf{A}^+$.

L'inverse généralisée $\mathbf{A}^+$ satisfait les conditions suivantes :

$$\mathbf{AA}^+\mathbf{A} = \mathbf{A}$$

$$\mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+$$

$$(\mathbf{AA}^+)^T = \mathbf{AA}^+ \qquad \text{Condition de symétrie}$$

$$(\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A}$$

La solution d'un système linéaire à partir de la pseudo-inverse $\mathbf{A}^+$ s'écrit alors : $\mathbf{x} = \mathbf{A}^+\mathbf{y}$.

La résolution d'un tel système met en évidence trois cas. Selon les dimensions m et n, on définira les matrices inverse, pseudo-inverse à gauche et pseudo-inverse à droite.

La matrice inverse est la solution d'un problème qui possède autant d'inconnues (variables à déterminées) que de contraintes. Cela ne signifie pas pour autant qu'il existe une solution.

La pseudo-inverse à gauche est une solution d'un problème sur-déterminé, qui contient des informations redondantes, elle minimise l'erreur quadratique moyenne, « mean square error » en Anglais.

La pseudo-inverse à droite est une solution d'un problème sous-déterminé, qui ne contient pas suffisamment d'information pour donner une solution unique, mais donne une solution particulière, celle qui minimise la norme quadratique.

En fait, ces deux pseudo-inverses qu'on appelle généralement matrice inverse généralisée de Moore-Penrose, sont données par :

$$\mathbf{A}^{+} = \lim_{\delta\to 0}(\mathbf{A}^T\mathbf{A}+\delta^2\mathbf{I})^{-1}\mathbf{A}^T = \lim_{\delta\to 0}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T+\delta^2\mathbf{I})^{-1}.$$

Si les colonnes de $\mathbf{A}$ sont linéairement indépendantes, on peut prendre $\delta = 0$.

1er cas, $m = n$:

C'est la cas d'une matrice carrée. Si $\mathbf{A}$ est une matrice non singulière (c'est à dire $\det(\mathbf{A}) \neq 0$), alors $\mathbf{A}^{+} = \mathbf{A}^{-1}$ est la matrice inverse de $\mathbf{A}$. La matrice $\mathbf{A}^{-1}$ vérifie la propriété $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = I_n$.

Propriétés de la matrice inverse :

Si $\mathbf{A}$ et $\mathbf{B}$ sont deux matrices non singulières, $\mathbf{A} \in \mathbb{R}^{n*n}$ et $\mathbf{B} \in \mathbb{R}^{n*n}$, alors :

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

$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$

$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ (par extension $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$ si la matrice $\mathbf{C}$ est inversible.

$\det(\mathbf{A}^{-1}) = \dfrac{1}{\det(\mathbf{A})}$

$\mathbf{A}(I_n+\mathbf{A})^{-1} = (I_n+\mathbf{A})^{-1}\mathbf{A}$ si $(I_n+\mathbf{A})^{-1}$ existe.

$\mathbf{A}(I_n+\mathbf{A})^{-1} + (I_n+\mathbf{A})^{-1} = I_n$ si $(I_n+\mathbf{A})^{-1}$ existe.

$\mathbf{A}(I_n+\mathbf{BA})^{-1} = (I_m+\mathbf{AB})^{-1}\mathbf{A}$ si $\mathbf{A} \in \mathbb{R}^{m*n}$, $\mathbf{B} \in \mathbb{R}^{n*m}$ et si $(I_n+\mathbf{BA})^{-1}$ et $(I_m+\mathbf{AB})^{-1}$ existent.

$(I_m+\mathbf{AB})^{-1} + \mathbf{A}(I_n+\mathbf{BA})^{-1}\mathbf{B} = I_m$ si $\mathbf{A} \in \mathbb{R}^{m*n}$, $\mathbf{B} \in \mathbb{R}^{n*m}$ et si $(I_n+\mathbf{BA})^{-1}$ et $(I_m+\mathbf{AB})^{-1}$ existent.

Calcul de la matrice inverse :

La matrice inverse $\mathbf{A}^{-1}$ d'une matrice $\mathbf{A}$ carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, est donnée par la relation :

$$\mathbf{A}^{-1} = \frac{\mathrm{adj}(\mathbf{A})}{\det(\mathbf{A})}.$$

Pour qu'une matrice soit inversible, il faut que sont déterminant soit différent de 0. On dit alors que la matrice est régulière ou non singulière.

Exemple lorsque $n = 2$:

Si $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, on détermine $\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$ puis on remplace chaque élément de $\mathbf{A}^T$ par son cofacteur pour déterminer $\mathrm{adj}(\mathbf{A}) = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$.

On calcule $\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$, puis on détermine la matrice inverse $\mathbf{A}^{-1}$ en divisant chaque élément de $\mathrm{adj}(\mathbf{A})$ par $\det(\mathbf{A})$.

On peut également calculer l'inverse d'une matrice $\mathbf{A}$ par la méthode du pivot de Gauss (Christol et al. 1996, p. 19) dans le cas particulier où la matrice est carrée et est inversible. Le système, caractérisé par la matrice $\mathbf{A}$, est alors dit système de Cramer ou système régulier dans ce cas.

Calcul de l'inverse d'une matrice partitionnée

Soit $\mathbf{A} \in \mathbb{R}^{n*n}$ une matrice partitionnée en quatre sous-matrices telle que : $\mathbf{A} = \begin{bmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{F} & \mathbf{G} \end{bmatrix}$, où $\mathbf{D} \in \mathbb{R}^{m*m}$ et les dimensions des matrices $\mathbf{E}$, $\mathbf{F}$ et $\mathbf{G}$ sont en accord avec celles de $\mathbf{A}$ et $\mathbf{D}$.

L'inverse $\mathbf{A}^{-1}$ de la matrice $\mathbf{A}$ est donnée par :

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{E}(\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})^{-1}\mathbf{F}\mathbf{D}^{-1} & -\mathbf{D}^{-1}\mathbf{E}(\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})^{-1} \\ -(\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})^{-1}\mathbf{F}\mathbf{D}^{-1} & (\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})^{-1} \end{bmatrix}$$

si $\mathbf{D}^{-1}$ existe.

2ème cas, $m < n$:

Il existe une infinité de solution, parmi lesquelles ont peut retenir celle qui minimise la norme $\|\mathbf{x}\|$:

$$\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$$

Cette matrice est appelée pseudo-inverse à droite, elle vérifie la propriété $\mathbf{A}\mathbf{A}^+ = \mathrm{I}_\mathrm{m}$.

3ème cas, $m > n$:

Il existe une solution approchée au sens des moindres carrés :

$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

Cette matrice est appelée pseudo-inverse à gauche, elle vérifie la propriété $\mathbf{A}\mathbf{A}^+ = \mathrm{I}_n$.

Propriétés de la pseudo-inverse :

Si $\mathbf{A}$ est une matrice et α un scalaire, $\alpha \in \mathbb{R}^*$:

$$(\mathbf{A}^+)^+ = \mathbf{A}$$
$$(\alpha\mathbf{A})^+ = \alpha^{-1}\mathbf{A}^+$$
$$(\mathbf{A}^+)^T = (\mathbf{A}^T)^+$$
$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^+\mathbf{A}^T = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+$$
$$\mathbf{A}\mathbf{A}^T(\mathbf{A}^+)^+\mathbf{A}^T = \mathbf{A}$$
$$\mathbf{A}^T\mathbf{A}\mathbf{A}^+ = \mathbf{A}^T$$
$$\mathrm{rang}(\mathbf{A}^+) = \mathrm{rang}(\mathbf{A}) = \mathrm{rang}(\mathbf{A}^T)$$

Détermination récursive de la pseudo-inverse (théorème de Gréville) :

Soit $\mathbf{A} \in \mathbb{R}^{m*n}$. La détermination récursive de la matrice pseudo-inverse de $\mathbf{A}$ repose sur l'idée de partition de $\mathbf{A}$ en colonnes.

A un instant k donné, la pseudo-inverse est calculée à partir de $\mathbf{A}_{k-1}^+$ et de la colonne courante d'indice k de la matrice $\mathbf{A}$.

Soit $\mathbf{A}_k = \begin{bmatrix}\mathbf{A}_{k-1} & a_k\end{bmatrix}$ où a_k désigne la k-ième colonne de $\mathbf{A}_k$, alors (Kohonen, 1984, p. 51) :

$$\mathbf{A}_k^+ = \begin{pmatrix}\mathbf{A}_{k-1}(\mathbf{I} - \mathbf{a}_k\mathbf{p}_k^T) \\ \mathbf{p}_k^T\end{pmatrix}$$

où $\mathbf{p}_k = \begin{cases}(\mathbf{I} - \mathbf{A}_{k-1}\mathbf{A}_{k-1}^+)\mathbf{a}_k / \left\|(\mathbf{I} - \mathbf{A}_{k-1}\mathbf{A}_{k-1}^+)\mathbf{a}_k\right\|^2 & \text{si} \quad (\mathbf{I} - \mathbf{A}_{k-1}\mathbf{A}_{k-1}^+)\mathbf{a}_k \neq 0 \\ (\mathbf{A}_{k-1}^+)^T\mathbf{A}_{k-1}^+\mathbf{a}_k / (1 + \left\|\mathbf{A}_{k-1}^+\mathbf{a}_k\right\|^2) & \text{sinon}\end{cases}$

Les conditions initiales sont les suivantes :

$$\mathbf{A}_1 = \mathbf{a}_1$$
$$\mathbf{A}_1^+ = \begin{cases}\mathbf{a}_1^T(\mathbf{a}_1^T\mathbf{a}_1)^{-1} & \text{si} \quad \mathbf{a}_1 \neq 0 \\ 0^T & \text{sinon}\end{cases}$$

Le Lemme d'inversion matricielle (The matrix inversion Lemma)

Considérons l'expression suivante :

$$\mathbf{A}^{-1} \stackrel{\Delta}{=} \mathbf{B}^{-1} + \mathbf{C}^T\mathbf{D}^{-1}\mathbf{C} \qquad [1].$$

Quelle est alors l'expression de $\mathbf{A} = (\mathbf{A}^{-1})^{-1}$?

Multiplions [1] à gauche par $\mathbf{A}$:

$$\mathbf{I} = \mathbf{A}\mathbf{B}^{-1} + \mathbf{A}\mathbf{C}^T\mathbf{D}^{-1}\mathbf{C} \qquad [2].$$

Multiplions [2] à droite par $\mathbf{B}$:

$$\mathbf{B} = \mathbf{A} + \mathbf{A}\mathbf{C}^T\mathbf{D}^{-1}\mathbf{C}\mathbf{B} \qquad [3].$$

Multiplions [3] à droite par $\mathbf{C}^T$:

$$\mathbf{BC}^T = \mathbf{AC}^T + \mathbf{AC}^T\mathbf{D}^{-1}\mathbf{CBC}^T = \mathbf{AC}^T\mathbf{D}^{-1}(\mathbf{D}+\mathbf{CBC}^T) \qquad [4].$$

Multiplions [4] à droite par $[\mathbf{D}+\mathbf{CBC}^T]^{-1}$:

$$\mathbf{BC}^T[\mathbf{D}+\mathbf{CBC}^T]^{-1} = \mathbf{AC}^T\mathbf{D}^{-1} \qquad [5].$$

Multiplions [5] à droite par $\mathbf{CB}$:

$$\mathbf{AC}^T\mathbf{D}^{-1}\mathbf{CB} = \mathbf{BC}^T[\mathbf{D}+\mathbf{CBC}^T]^{-1}\mathbf{CB} \qquad [6].$$

Retranchons [6] de $\mathbf{B}$:

$$\mathbf{B}-\mathbf{AC}^T\mathbf{D}^{-1}\mathbf{CB} = \mathbf{B}-\mathbf{BC}^T[\mathbf{D}+\mathbf{CBC}^T]^{-1}\mathbf{CB} \qquad [7].$$

D'après [3], on tire que $\mathbf{AC}^T\mathbf{D}^{-1}\mathbf{CB} = \mathbf{B}-\mathbf{A}$ et on peut déduire que :

$$\mathbf{A} = \mathbf{B}-\mathbf{BC}^T[\mathbf{D}+\mathbf{CBC}^T]^{-1}\mathbf{CB} \qquad [8].$$

Finalement, on retiendra une des deux formulations suivantes :

$$\mathbf{A} \stackrel{\Delta}{=} \mathbf{B}+\mathbf{CD}^T \quad\Rightarrow\quad \mathbf{A}^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{CD}^T\mathbf{B}^{-1}[1+\mathbf{D}^T\mathbf{B}^{-1}\mathbf{C}]^{-1}$$

$$\mathbf{A}^{-1} \stackrel{\Delta}{=} \mathbf{B}^{-1}+\mathbf{C}^T\mathbf{D}^{-1}\mathbf{C} \quad\Rightarrow\quad \mathbf{A} = \mathbf{B}-\mathbf{BC}^T[\mathbf{D}+\mathbf{CBC}^T]^{-1}\mathbf{CB}$$

Le noyau d'une matrice

Pour un système d'équations homogènes $\mathbf{Ax}=0$, $\mathbf{A}\in\mathbb{R}^{m*n}$, $\mathbf{x}\in\mathbb{R}^n$, l'ensemble des solutions $\mathbf{x}$ constitue un espace vectoriel appelé noyau de $\mathbf{A}$, noté $\mathrm{Ker}(\mathbf{A})$. La dimension de cet espace sera notée $\mathrm{N}_\mathbf{A}$.

Remarque : Une équation linéaire $a_1x_1+a_2x_2+a_3x_3+...a_nx_n=h$ est dite non homogène si $h\neq 0$.

Théorème : pour $\mathbf{A}\in\mathbb{R}^{m*n}$ on a $\mathrm{rang}(\mathbf{A})+\mathrm{N}_\mathbf{A}=n$.

Résolution de l'équation $\mathbf{AXB}=\mathbf{C}$

Considérons l'équation $\mathbf{AXB}=\mathbf{C}$ (Kohonen, 1984, p.53), avec $\mathbf{A}$, $\mathbf{B}$ et $\mathbf{C}$ des matrices quelconques dont les dimensions sont en accords avec les règles de la multiplication matricielle et où l'inconnu à déterminer est $\mathbf{X}$.
La condition nécessaire et suffisante pour avoir une solution à cette équation est :

$$\mathbf{AA^+CB^+B} = \mathbf{C}\,.$$

La solution générale est de la forme :

$$\mathbf{X} = \mathbf{A^+CB^+} + \mathbf{Y} - \mathbf{A^+AYBB^+}\,,$$

avec $\mathbf{Y}$ de mêmes dimensions que $\mathbf{X}$.

Une solution particulière est $\mathbf{X} = \mathbf{A^+CB^+}$ (cas pour $\mathbf{Y} = 0$), c'est la solution dont la norme Euclidienne est minimale.

11. Indépendance linéaire de vecteurs, la singularité

Soit un ensemble $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ de n vecteurs de dimension m, et $\alpha_1, \alpha_2, ..., \alpha_n$ un ensemble de scalaires.

Le vecteur défini par $\mathbf{c} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$ forme une combinaison linéaire de vecteurs, notée $\{\mathbf{a}_i\}$.

L'ensemble des vecteurs $\{\mathbf{a}_i\}$ est linéairement indépendants si :

$$\mathbf{c} = 0 \Leftrightarrow \alpha_1 = \alpha_2 = ... = \alpha_i = ... = \alpha_n = 0$$

Soit $\mathbf{A}$ une matrice, $\mathbf{A} \in \mathbb{R}^{m*n}$.

- Les colonnes de $\mathbf{A}$ sont linéairement indépendantes si et seulement si $\mathbf{A}^T\mathbf{A}$ est une matrice non singulière de rang plein,
- Les lignes de $\mathbf{A}$ sont linéairement indépendantes si et seulement si $\mathbf{A}^T\mathbf{A}$ est non singulière.

Une matrice carrée $\mathbf{A} \in \mathbb{R}^{n*n}$, est dite non singulière (ou inversible) si $\exists\ \mathbf{B} \in \mathbb{R}^{n*n}$ telle que :

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I} \Leftrightarrow \det(\mathbf{A}) \neq 0\,.$$

12. Le rang d'une matrice

Le rang d'une matrice correspond au nombre maximum de colonnes ou de lignes linéairement indépendantes. C'est aussi l'ordre du plus grand déterminant non nul. Si k est cet ordre, on dit que la matrice est de rang k.

Une matrice $\mathbf{A}$, $\mathbf{A} \in \mathbb{R}^{m*n}$, est dite de rang plein si : $\text{rang}(\mathbf{A}) = \min(m, n)$.

Propriété :

Quelle que soit une matrice $\mathbf{A}$, $\mathbf{A} \in \mathbb{R}^{m*n}$,

$$\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}^T) = \text{rang}(\mathbf{AA}^T) = \text{rang}(\mathbf{A}^T\mathbf{A})\,.$$

13. La trace d'une matrice

La trace d'une matrice $\mathbf{A}$, $\mathbf{A} \in \mathbb{R}^{n*n}$ est égale à la somme des éléments de la diagonale de cette matrice :

$$\text{trace}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} \quad .$$

Propriété :

$$\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA}) \qquad \mathbf{A} \in \mathbb{R}^{n*n}, \ \mathbf{B} \in \mathbb{R}^{n*n} .$$

$$\frac{\partial(\text{trace}(\mathbf{AB}^T))}{\partial \mathbf{A}} = \mathbf{B} \qquad \mathbf{A} \in \mathbb{R}^{n*m}, \ \mathbf{B} \in \mathbb{R}^{n*m} .$$

$$\mathbf{D} = \frac{\partial(\text{trace}(\mathbf{ABA}^T))}{\partial \mathbf{A}} = 2\mathbf{AB} \qquad \mathbf{D} \in \mathbb{R}^{n*m} \text{ et } \mathbf{A} \in \mathbb{R}^{n*m}, \ \mathbf{B} \in \mathbb{R}^{m*m} .$$

14. La norme

La norme vectorielle

On appelle p-norme, la norme notée également L_p d'un vecteur $\mathbf{x} \in \mathbb{R}^n$:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \quad .$$

Les normes les plus usuelles sont :

- la norme L_1 également appelée norme absolue, $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$,
- la norme L_2 ou norme Euclidienne, $\|\mathbf{x}\| = \|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2} = \left(\mathbf{x}^T\mathbf{x}\right)^{1/2}$.

Voici quelques propriétés de la norme Euclidienne, $\mathbf{x}$ et $\mathbf{y} \in \mathbb{R}^n$:

- $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$,
- $\|\mathbf{x}^T\mathbf{y}\| \leq \|\mathbf{x}\|.\|\mathbf{y}\|$,
- $|\mathbf{v}.\mathbf{w}| \leq \|\mathbf{v}\|\|\mathbf{w}\|$ si $\mathbf{v} \in \mathbb{R}^n$ et $\mathbf{w} \in \mathbb{R}^n$, on appelle cette propriété l'inégalité de Schwartz.

La norme Euclidienne est aussi appelé norme quadratique (c'est la distance Euclidienne).
Une norme peut être pondérée. Ceci est surtout utilisé en Physique, lorsque les éléments d'un vecteur sont de nature différente (représentent des grandeurs dans des unités différentes).

Soit $\mathbf{x}$ un vecteur à normer, $\mathbf{x} \in \mathbb{R}^n$, on pose, $\mathbf{y} \in \mathbb{R}^n$: $\mathbf{y} = \mathbf{Dx}$.
On calcule alors la norme (Euclidienne par exemple) du vecteur $\mathbf{y}$ plutôt que celle de $\mathbf{x}$:

$$\|\mathbf{y}\| = \left(\mathbf{y}^T\mathbf{y}\right)^{1/2} = \|\mathbf{D}\mathbf{x}\| = \left(\mathbf{x}^T\mathbf{D}^T\mathbf{D}\mathbf{x}\right)^{1/2} = \left(\mathbf{x}^T\mathbf{Q}\mathbf{x}\right)^{1/2}.$$

La matrice $\mathbf{Q} = \mathbf{D}^T\mathbf{D}$ est une matrice de pondération de la forme quadratique $\mathbf{x}^T\mathbf{Q}\mathbf{x}$ présente dans l'expression de la norme $\|\mathbf{y}\|$.

Pour la matrice $\mathbf{D} \in \mathbb{R}^{n*n}$, on prend souvent une matrice diagonale avec chaque terme comme étant l'inverse de la valeur maximale que peut prendre l'élément concerné :

$$\mathbf{D} = \begin{bmatrix} 1/x_{1\max} & \cdots & 0 \\ \vdots & 1/x_{i\max} & \vdots \\ 0 & \cdots & 1/x_{n\max} \end{bmatrix}.$$

La norme matricielle

Il existe plusieurs normes matricielles. La norme d'une matrice peut être définie dans différent sens, elle doit satisfaire les contraintes posées par n'importe quel espace (Kohonen, 1984, p.47).

Les normes matricielles, sont définies de la même façon que les normes vectorielles.

On appelle p-norme d'une matrice $\mathbf{A}$, $\mathbf{A} \in \mathbb{R}^{m*n}$, la norme :

$$\|\mathbf{A}\|_p = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \left| a_{ij} \right|^p \right)^{1/p} .$$

La norme matricielle la plus connue et la plus utilisées est l'équivalente de la norme Euclidienne chez les vecteurs, on l'appelle la norme de Frobenius :

$$\|\mathbf{A}\| = \|\mathbf{A}\|_E = \sqrt{\text{trace}(\mathbf{A}^T\mathbf{A})} ,$$

$$\|\mathbf{A}\| = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \left| a_{ij} \right|^2 \right)^{1/2} .$$

La norme de Frobenius est la norme Euclidienne au sens matricielle, qui n'est pas compatible avec la norme Euclidienne vectorielle (it is not consistent with the Euclidean vector norm).

Si $\mathbf{A}$ est une matrice carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, et $\mathbf{x}$ un vecteur de n éléments, :

$$\|\mathbf{A}\mathbf{x}\| = \|\mathbf{x}\| .$$

La norme suivante est elle compatible avec la norme vectorielle :

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| .$$

15. L'orthogonalité

L'ensemble des vecteurs $\{\mathbf{a}_i\}$, $\mathbf{a}_i \in \mathbb{R}^n$, $\forall i = 1,...,k$, est orthogonal si :

$$\mathbf{a}_i^T \mathbf{a}_j = 0, \ \forall i \neq j.$$

Ce même ensemble est orthogonal si :

$$\mathbf{a}_i^T \mathbf{a}_j = 0, \ \forall i \neq j,$$
$$\mathbf{a}_i^T \mathbf{a}_i = 1, \ \forall i.$$

Ceci se note également :

$$\mathbf{a}_i^T \mathbf{a}_j = \delta_{ij},$$

où δ_{ij} est l'opérateur de Kronecker.

Une matrice carrée $\mathbf{A} \in \mathbb{R}^{n*n}$ est dite orthogonale si ses colonnes forment un ensemble orthogonale :

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = I_n \Rightarrow \mathbf{A}^T = \mathbf{A}^{-1}.$$

Propriétés

Si $\mathbf{A}$ est une matrice carrée orthogonale, $\mathbf{A} \in \mathbb{R}^{n*n}$, et $\mathbf{x}$ un vecteur de n éléments, $\mathbf{x} \in \mathbb{R}^n$:

$\|\mathbf{A}\mathbf{x}\| = \|\mathbf{x}\|$

$(\mathbf{A}\mathbf{x}^T)(\mathbf{A}\mathbf{y}) = \mathbf{x}^T\mathbf{y}$

Si $\mathbf{A} \in \mathbb{R}^{m*n}$, la condition d'orthogonalité se traduit par $\mathbf{A}^T\mathbf{A} = I_n$.

Quelques théorèmes

1. Si $\mathbf{A}$ est une matrice orthogonale, son inverse et sa transposée le sont également.
2. Le produit de plusieurs matrices orthogonales est une matrice orthogonale.
3. Le déterminant d'une matrice orthogonale est égal à ± 1.

16. Valeurs et vecteurs propres

Soit $\mathbf{A}$ une matrice carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$. Les valeurs propres λ_i de cette matrices (appelées aussi valeurs caractéristiques) sont les racines de l'équation polynomiale :

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

L'ensemble des valeurs propres d'une matrice constitue son spectre.

Les vecteurs propres (ou vecteurs caractéristiques) $\mathbf{v}_i$ de cette matrices se déduisent de la définition suivante :

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

Comme le système est indéterminé, les vecteurs propres sont obtenus à une constante multiplicative près, c'est à dire que l'on détermine des directions propres.

Remarques :

$\det(\mathbf{A} - \lambda_i \mathbf{I}) = 0$ est appelé polynôme caractéristique de $\mathbf{A}$.

$\lambda_i \mathbf{I} - \mathbf{A}$ est appelé la matrice caractéristique de $\mathbf{A}$.

Deux matrices sont dites semblables si elles ont le même polynôme caractéristique.

Propriétés

Si $\mathbf{A}$ est une matrice carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, ayant pour valeurs propres λ_i, $i = 1,...,n$:

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$$

$$\mathrm{trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$$

17. Diagonalisation d'une matrice

Si les valeurs propres d'une matrice $\mathbf{A} \in \mathbb{R}^{n*n}$ sont réelles et simples, on obtient n directions propres distinctes que l'on peut utiliser comme axes de coordonnées. Dans ce nouveau système d'axes, la matrice $\mathbf{A}$ devient une matrice $\mathbf{D}$ diagonale telle que :

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots \lambda_i & \\ 0 & & \ddots \lambda_n \end{bmatrix}.$$

Soit $\mathbf{P}$ la matrice formée par les composantes des vecteurs propres $\mathbf{v}_i$:

$$\mathbf{P} = \left[\mathbf{v}_1 ... \mathbf{v}_i ... \mathbf{v}_n \right].$$

Cette matrice ($\mathbf{P} \in \mathbb{R}^{n*n}$) est appelée la matrice modèle de $\mathbf{A}$. On montre alors que (Christol et al., 1996, p 141) :

$$\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}.$$

La diagonalisation d'une matrice permet de résoudre de façon élégante des problèmes qui peuvent sembler complexes au premier abord (Rotella et Borne, 1995).

Elle est souvent, dans le cas des problèmes complexes (les problèmes basés sur l'inversion de matrices singulières), l'unique façon de procéder. On utilise alors d'autres méthodes pour calculer les valeurs propres (Strang, 1993).
Les méthodes numériques prennent également de plus en plus d'importance dans le calcul d'inversion de telles matrices (Stoer et Bulirsch, 1993).

Intérêt de la diagonalisation

Considérons le système différentiel linéaire suivant, représenté dans une base quelconque :

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}.$$

Il est constitué de n équations différentielles du premier ordre dépendantes. Effectuons le changement de base défini par :

$$\mathbf{x} = \mathbf{P}\mathbf{y} \text{ ou } \mathbf{y} = \mathbf{P}^{-1}\mathbf{x}.$$

Le système différentiel devient alors :

$$\dot{\mathbf{y}} = \mathbf{P}^{-1}\dot{\mathbf{x}} = \mathbf{P}^{-1}\mathbf{A}\mathbf{x} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{y},$$

donc en fait : $\dot{\mathbf{y}} = \mathbf{D}\mathbf{y}$,

où $\mathbf{D}$ est une matrice diagonale. La résolution du système différentiel obtenu par changement de base se réduit alors à n équations différentielles du premier ordre indépendantes.
La diagonalisation permet également la simplification et l'étude des formes quadratiques (Cairoli, 1991). Une autre application directe de la diagonalisation d'une matrice est la calcul de ces puissances entières.

Puissance *n*-ème d'une matrice

La formule $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ avec $\mathbf{A}$ une matrice diagonalisable s'écrit aussi $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. On en déduit que :

$$\mathbf{A}^n = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}\mathbf{P}\mathbf{D}\mathbf{P}^{-1}...\mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \mathbf{P}\mathbf{D}^n\mathbf{P}^{-1}.$$

La matrice $\mathbf{D}$ (respectivement $\mathbf{A}$) représente l'endomorphisme (application linéaire) u dans une base **f** formée de vecteurs propres (respectivement dans la base canonique **e**). La formule précédente exprime simplement le fait que les matrices $\mathbf{D}^n$ et $\mathbf{A}^n$ représentent le même endomorphisme u^n respectivement dans la base **f** et dans la base **e**.
Le produit de deux matrices est très facile à calculer : la matrice produit est diagonale et le *i*-ème terme de sa diagonale est le produit des *i*-èmes termes des diagonales des deux matrices.

On en déduit que la puissance n-ème de la matrice diagonale $\mathbf{D}$ s'obtient simplement en élevant à la puissance n chacun des termes de sa diagonale.

18. Formes quadratiques

Une forme quadratique est un polynôme homogène du second degré en n variables $x_1, x_2, \ldots, x_n$:

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \; .$$

On peut représenter la forme quadratique grâce aux notations matricielles. Pour cela, considérons $\mathbf{A}$ une matrice carrée, $\mathbf{A} \in \mathbb{R}^{n*n}$, et $\mathbf{x}$ un vecteur de n éléments, $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x}^T = \begin{bmatrix} x_1 \; \ldots \; x_n \end{bmatrix}, \; \mathbf{A} = \begin{bmatrix} a_{ij} \end{bmatrix},$$

alors

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} \; .$$

Remarque : On dit que de l'expression $f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{y}$ qu'elle est de la forme bilinéaire, avec $\mathbf{y}$ un vecteur de dimension n, $\mathbf{y} \in \mathbb{R}^n$.

Supposons que $\mathbf{A}$ soit une matrice réelle.
Toute matrice carrée peut être décomposée en une matrice $\mathbf{A}_s$ symétrique, $\mathbf{A}_s = \mathbf{A}_s^T$, et en une matrice $\mathbf{A}_a$ anti-symétrique, $\mathbf{A}_a = -\mathbf{A}_a^T$.
Ainsi, $\mathbf{A} = \mathbf{A}_s + \mathbf{A}_a$ avec :

$$\mathbf{A}_s = \frac{\mathbf{A} + \mathbf{A}^T}{2} \; \text{ et } \; \mathbf{A}_a = \frac{\mathbf{A} - \mathbf{A}^T}{2} \; .$$

Si la matrice $\mathbf{A}$ n'est pas symétrique au départ, on peut écrire :

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T (\mathbf{A}_s + \mathbf{A}_a) \mathbf{x} = \mathbf{x}^T \mathbf{A}_s \mathbf{x} + \mathbf{x}^T \mathbf{A}_a \mathbf{x} \; .$$

Dans cette expression, les deux termes sont des scalaires, ils sont donc égaux à leur transposé, en particulier :

$$\mathbf{x}^T \mathbf{A}_a \mathbf{x} = (\mathbf{x}^T \mathbf{A}_a \mathbf{x})^T = -\mathbf{x}^T \mathbf{A}_a^T \mathbf{x} = -\mathbf{x}^T \mathbf{A}_a \mathbf{x}$$

qui doit être nul, d'où

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{A}_s \mathbf{x} \; .$$

On peut donc affirmer sans perdre de généralité que la matrice $\mathbf{A}$ de la forme quadratique est symétrique.

19. Définitivité

On dit qu'une matrice $\mathbf{A} \in \mathbb{R}^{n*n}$, est définie positive, $\mathbf{A} > 0$, si la forme quadratique qui lui est associée est définie positive :

$$\mathbf{A} > 0 \Leftrightarrow \mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \;\; , \;\; \forall \; \mathbf{x} \neq 0 \; .$$

$\mathbf{A}$ est définie semi-positive, $\mathbf{A} \geq 0$, si $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ pour tout vecteur $\mathbf{x} \neq 0$.

$\mathbf{A}$ est définie négative, $\mathbf{A} < 0$, si $\mathbf{x}^T \mathbf{A} \mathbf{x} < 0$ pour tout $\mathbf{x} \neq 0$.

$\mathbf{A}$ est définie semi- négative, $\mathbf{A} \leq 0$, si $\mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0$ pour tout vecteur $\mathbf{x} \neq 0$.

On parle, selon la cas, de positivité ou de négativité.

En fait, une forme quadratique est dite non définie si son signe dépend du vecteur $\mathbf{x}$.

Pour tester la positivité d'une matrice, on peut utiliser le critère de Sylvester qui dit qu'une matrice est définie positive si tous ses mineurs principaux sont positifs :

$$\mathbf{A} > 0 \text{ ssi } m_i > 0, \ \forall i = 1,...,n.$$

Ce critère n'est plus valable pour la semi-positivité :

$\mathbf{A} \geq 0$ ssi $\mathrm{m}_i \geq 0$, $\forall i = 1,...,n$ n'est pas vrai, mais :

$$\mathbf{A} \geq 0 \text{ ssi } \mathrm{m}_i \geq 0 \text{ et } m_{ij} \geq 0, \ \forall i = 1,...,n, \ \forall j = 1,...,n.$$

Les mêmes raisonnements sont valables pour les cas négatifs.

Cela peut également se vérifier en utilisant les valeurs propres λ_i, $\forall i = 1,...,n$, de la matrice $\mathbf{A}$:

$$\mathbf{A} > 0 \text{ ssi } \lambda_i > 0,$$
$$\mathbf{A} \geq 0 \text{ ssi } \lambda_i \geq 0,$$
$$\mathbf{A} < 0 \text{ ssi } \lambda_i < 0,$$
$$\mathbf{A} \leq 0 \text{ ssi } \lambda_i \leq 0.$$

20. Changement de base

Définition d'une base

Lorsqu'on travaille avec les combinaisons linéaires, il est commode d'introduire la notion de famille de vecteurs d'un espace E qui généralise la notion de partie. Dans une famille, les vecteurs sont rangés dans un certain ordre et le même vecteur peut apparaître plusieurs fois (dans le langage des dénombrements mathématiques, une famille est un arrangement avec répétition, alors qu'une partie est une combinaison). Les familles les plus intéressantes sont celles ayant un nombre n fini d'éléments et pour lesquelles les vecteurs sont distincts. Ces considérations faites, la différence entre famille et partie ne concerne alors plus que l'ordre des vecteurs. On note une famille de n vecteurs indicés par l'ensemble $I = \{1,...,n\}$ par $(a_i)_{i \in I}$ ou plus simplement, (a_i).

Une famille de vecteurs libre est un ensemble de vecteurs linéairement indépendants. A l'inverse, une famille liée contient au moins un vecteur qui est une combinaison linéaire des autres.

On dit que la famille A est génératrice d'un sous-espace vectoriel F de E si A est contenu dans F et si tout vecteur de F est une combinaison linéaire de vecteurs de A.

Une famille de vecteurs d'un espace vectoriel E qui est à la fois libre et génératrice de E est appelé base de E.

Exemple : la base canonique.
La famille $(\mathbf{e}_1, \mathbf{e}_2, ...\mathbf{e}_n)$ formée des n vecteurs $\mathbf{e}_1 = (1, 0, ...0)$, $\mathbf{e}_2 = (0, 1, 0, ...0)$, … $\mathbf{e}_i = (0, 0, ..., 1, 0, ...0)$, … $\mathbf{e}_n = (0, 0, ..., 0, 1)$ de $\mathbb{R}^n$ est une base de $\mathbb{R}^n$ appelée base canonique.

Application : plusieurs repères dans l'espace 3D

L'objectif que l'on se fixe est de représenter n'importe point de l'espace 3D dans plusieurs bases.

On se place dans un repère 3D R_1, défini par une base orthonormée. A partir de ce premier repère, on définit par rapport à R_1 un second repère R_2, également défini par une base orthonormée.
Le passage d'un repère à un autre fait intervenir deux opérations essentielles : la translation et la rotation.
Le repère R_2 est donc défini par la position de son origine O_2 dans le repère R_1 (c'est la translation, définie par un vecteur $\mathbf{t} \in \mathbb{R}^3$) et par l'orientation des vecteurs formant sa base (c'est la rotation, définie par une matrice $\mathbf{A} \in \mathbb{R}^{3*3}$) par rapport à ceux qui constituent la base de R_1.

Considérons dans un premier temps qu'une simple translation de $\mathbf{t} = \begin{bmatrix} 4 & 2 & 2 \end{bmatrix}^T$ comme le montre la figure suivante.
S'il n'y a pas de rotation pour passer d'un repère à l'autre, la matrice de rotation est la matrice identité : $\mathbf{A} = I_3$.

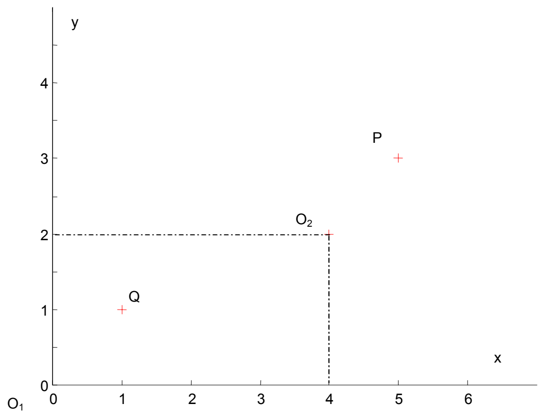

Figure 5 : Les points P et Q dans le premier repère par rapport aux axes x et y.

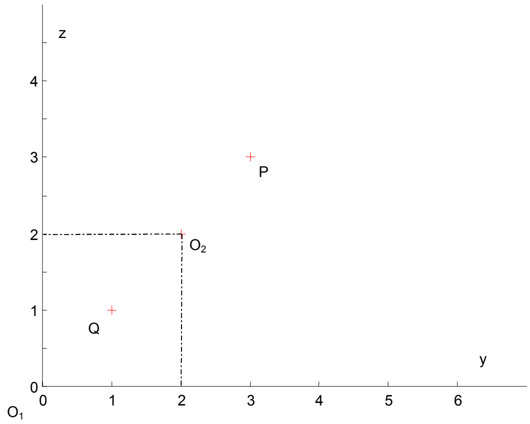

Figure 6 : Les points P et Q dans le premier repère par rapport aux axes y et z.

Connaissant les coordonnées d'un point $\mathrm{P}_{\mathrm{R}_2} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^T$ dans le repère R_2, on veut connaître ses coordonnées $\mathrm{P}_{\mathrm{R}_1}$ dans le repère R_1 :

$$\mathrm{P}_{\mathrm{R}_1} = \mathbf{A}\mathrm{P}_{\mathrm{R}_2} + \mathbf{t}.$$

On trouve alors : $\mathrm{P}_{\mathrm{R}_1} = \begin{bmatrix}5 & 3 & 3\end{bmatrix}^T$.

On peut poser le problème inverse : quelles sont les coordonnées dans R_2 d'un point défini dans R_1 ?

Dans R_1 définissons le point Q tel que $\mathrm{Q}_{\mathrm{R}_1} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^T$. Les coordonnées de Q dans R_2 sont données par l'application inverse à celle définie précédemment :

$$\mathrm{Q}_{\mathrm{R}_1} = \mathbf{A}Q_{\mathrm{R}_2} + \mathbf{t},$$
$$Q_{\mathrm{R}_2} = \mathbf{A}^{-1}(\mathrm{Q}_{\mathrm{R}_1} - \mathbf{t}).$$

L'application numérique donne $\mathrm{Q}_{\mathrm{R}_2} = \begin{bmatrix}-3 & -1 & -1\end{bmatrix}^T$.

Considérons maintenant, que le passage de R_1 à R_2 ne se fasse non seulement par la translation de $\mathbf{t} = \begin{bmatrix}4 & 2 & 2\end{bmatrix}^T$ mais aussi par une rotation définie par la matrice $\mathbf{A} = \begin{bmatrix}0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0\end{bmatrix}$ (les angles d'Euler équivalents sont : $\begin{bmatrix}\psi & \vartheta & \varphi\end{bmatrix}^T = \begin{bmatrix}0 & \pi/2 & \pi/2\end{bmatrix}^T$).

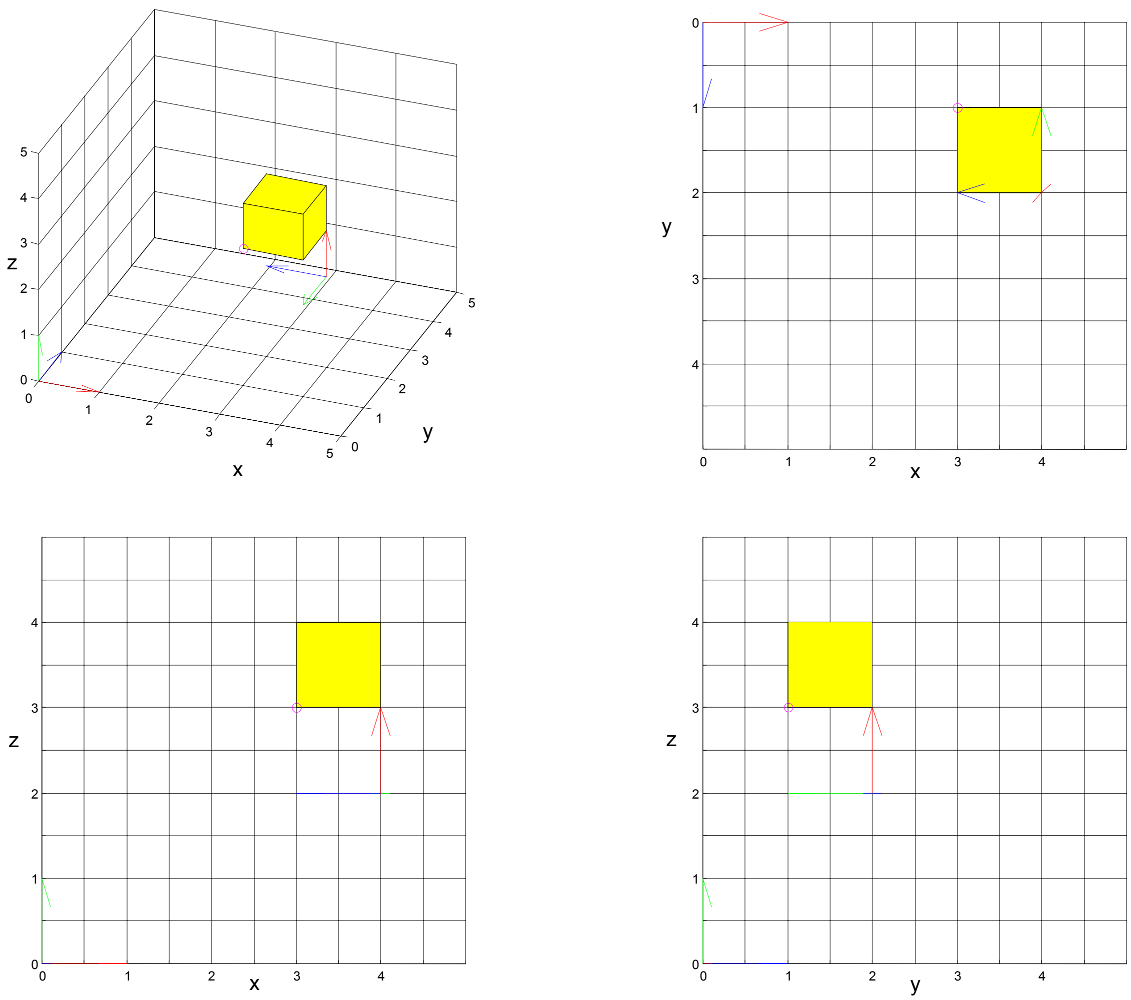

Figure 7 : Un cube dans un espace 3D, le point Q est représenté par un cercle magenta.

Le point S, défini dans R_2 par $S_{R_2} = [1 \quad 1 \quad 1]^T$ et représenté sur les figures précédentes par un cercle (un des sommets du cube) a pour coordonnées dans R_1 S_{R_1} selon la formule :

$$S_{R_1} = \mathbf{A}S_{R_2} + \mathbf{t} \,,$$

soit après calcul, $S_{R_1} = [3 \quad 1 \quad 3]^T$.

Le point T, défini dans le repère R_1 par $T_{R_1} = [3 \quad -8 \quad 3]^T$ voit ses coordonnées T_{R_2} dans R_2 données par l'application inverse :

$$T_{R_2} = \mathbf{A}^{-1}(T_{R_1} - \mathbf{t}) \,.$$

Le calcul donne $T_{R_2} = [1 \quad 1 \quad 10]^T$.

On voit finalement, que le passage d'un repère à l'autre utilise une multiplication à gauche, que ce soit pour passer de R_1 à R_2, ou inversement de R_2 à R_1. La multiplication à cause peut donc de façon générale être assimilé à un changement de base.

Application : les axes principaux d'une ellipse

L'objectif que l'on se fixe est de représenter n'importe quelle ellipse en coordonnées polaires.

Dans un premier temps, on rappelle un cas particulier, celui où les axes de l'ellipse sont parallèles aux axes de coordonnées. On généralisera ensuite pour le cas des axes quelconques.

Axes de l'ellipse parallèles aux axes de coordonnées :

L'équation d'une ellipse est donnée par :

$$a(x-x_0)^2 + b(y-y_0)^2 = d \ ,$$

avec $\{x_0, y_0\}$ le centre de l'ellipse et a , b et d positifs.

On peut aussi écrire :

$$\frac{(x-x_0)^2}{\rho_a^2} + \frac{(y-y_0)^2}{\rho_b^2} = 1 \ ,$$

avec $\rho_a = \sqrt{d/a}$ et $\rho_b = \sqrt{d/b}$.

Les coordonnées polaires qui représentent cette ellipse sont alors :

$$\begin{cases} x = \rho_a \cos(\vartheta) + x_0 \\ y = \rho_b \cos(\vartheta) + y_0 \end{cases}$$

avec $0 \leq \vartheta \leq 2\pi$.

Axes de l'ellipse quelconques :

L'équation générale d'une ellipse quelconque décrite dans un référentiel est :

$$a(x-x_0)^2 + b(y-y_0)^2 + 2c(x-x_0)(y-y_0) = d \ .$$

On peut aussi l'écrire sous forme quadratique :

$$\begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix} \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} = d \ .$$

Soit, en notation matricielle :

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = d \ .$$

La courbe correspondante est une ellipse si la matrice $\mathbf{A}$ est définie positive, c'est à dire si a et b sont positifs et si $\det(\mathbf{A}) > 0$, soit $ab > c^2$.

Caractériser une ellipse revient à déterminer ses axes principaux qui sont les directions propres de la matrice $\mathbf{A}$. Les vecteurs propres de $\mathbf{A}$ sont définis par :

$$\det(\mathbf{A} - \lambda_i \mathbf{I}) = 0 \ ,$$

soit,

$$\lambda_i = \frac{\mathrm{a+b} \pm \sqrt{(a-b)^2 + 4c^2}}{2}.$$

Soient d_1 et d_2 les directions propres et $\mathbf{v}_1$ et $\mathbf{v}_2$ les vecteurs propres $\mathbf{A}$. Les vecteurs propres vérifient la relation :

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

Ces vecteurs sont définis à une constante près : on ne connaît que leurs directions d_1 et d_2, soit :

$$\begin{cases} d_1 : y = \alpha_1 x \quad \text{avec} \quad \alpha_1 = (\lambda_1 - a)/c \\ d_2 : y = \alpha_2 x \quad \text{avec} \quad \alpha_2 = (\lambda_2 - a)/c \end{cases}$$

Soient v et w les coordonnées d'un point dans les axes définis par les directions d_1 et d_2. Avec ces coordonnées, l'équation de l'ellipse s'écrit :

$$\lambda_1 v^2 + \lambda_2 w^2 = d\,,$$

ou

$$\frac{v^2}{\rho_1^2} + \frac{w^2}{\rho_2^2} = 1\,,$$

avec $\rho_1 = \sqrt{d/\lambda_1}$ et $\rho_2 = \sqrt{d/\lambda_2}$.
On peut donc facilement tracer l'ellipse avec les coordonnées $\{v, w\}$, puis passer dans les axes $\{x, y\}$ par une matrice de changement de base.

Soit $\mathbf{B}$ la matrice de changement de base telle que :

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{B} \begin{bmatrix} v \\ w \end{bmatrix}$$

On démontre alors que :

$$\mathbf{B} = \begin{bmatrix} \dfrac{1}{\sqrt{1+\alpha_1^2}} & \dfrac{1}{\sqrt{1+\alpha_2^2}} \\ \dfrac{\alpha_1}{\sqrt{1+\alpha_1^2}} & \dfrac{\alpha_2}{\sqrt{1+\alpha_1^2}} \end{bmatrix}.$$

21. Matrice racine carrée

Toute matrice $\mathbf{A}$, $\mathbf{A} \in \mathbb{R}^{n*n}$, définie semi-positive, $\mathbf{A} \geq 0$, peut être factorisée en deux matrices racines carrées tel que :

$$\mathbf{A} = \sqrt{\mathbf{A}}\sqrt{\mathbf{A}}^T \text{ ou } \mathbf{A} = \sqrt{\mathbf{A}}^T\sqrt{\mathbf{A}} .$$

Les matrices racines carrées « gauche » et « droite » des deux expressions précédentes ne sont pas forcément les mêmes.

Il peut y avoir plusieurs matrices racines carrées, et de toutes sortes. En fait, si $\mathbf{M}$ est une matrice orthogonale, $\mathbf{M}^T = \mathbf{M}^{-1}$, $\mathbf{M} \in \mathbb{R}^{n*n}$, alors on montre que :

$$\mathbf{A} = \sqrt{\mathbf{A}}^T\mathbf{M}^T\mathbf{M}\sqrt{\mathbf{A}} .$$

$\sqrt{\mathbf{A}}^T\mathbf{M}^T$ est une matrice racine carrée de $\mathbf{A}$.

Si $\mathbf{A} > 0$, alors toutes les matrices racines carrées de $\mathbf{A}$ sont non singulières.

Les fonctions vectorielles (matrix calculus)

1. La différentiation et l'intégration par rapport à un scalaire

Soit $\mathbf{x}$ vecteur, $\mathbf{x} \in \mathbb{R}^n$ et soit $\mathbf{A}$ une matrice, $\mathbf{A} \in \mathbb{R}^{m*n}$.

La dérivée du vecteur $\mathbf{x}$ par rapport à un scalaire t vaut :

$$\frac{\partial \mathbf{x}}{\partial t} = \begin{bmatrix} \partial x_1 / \partial t \\ \vdots \\ \partial x_n / \partial t \end{bmatrix} \text{ et } \frac{\partial \mathbf{x}^T}{\partial t} = \left[\partial x_1 / \partial t \quad \cdots \quad \partial x_n / \partial t \right].$$

La dérivée de la matrice $\mathbf{A}$ par rapport à un scalaire t vaut :

$$\frac{\partial \mathbf{A}}{\partial t} = \left[\frac{\partial a_{ij}}{\partial t} \right].$$

Quelques propriétés sur la dérivation :

Soient $\mathbf{A} \in \mathbb{R}^{m*n}$ et $\mathbf{B} \in \mathbb{R}^{m*n}$ des matrices et soit $\mathbf{x} \in \mathbb{R}^n$ un vecteur :

$$\frac{\partial}{\partial t}(\mathbf{A}+\mathbf{B}) = \frac{\partial \mathbf{A}}{\partial t} + \frac{\partial \mathbf{B}}{\partial t}$$

$$\frac{\partial}{\partial t}(\mathbf{AB}) = \frac{\partial \mathbf{A}}{\partial t}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial t}$$

$$\frac{\partial \mathbf{A}^n}{\partial t} = \frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{n-1} + \mathbf{A}\frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{n-2} + \cdots + \mathbf{A}^{n-1}\frac{\partial \mathbf{A}}{\partial t}$$

$$\frac{\partial \mathbf{A}^{-1}}{\partial t} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{-1}$$

$$\frac{\partial |\mathbf{A}|}{\partial t} = \mathrm{trace}(\frac{\partial \mathbf{A}}{\partial t}\tilde{\mathbf{A}})$$

$$\frac{\partial (\mathbf{A}\mathbf{A}^{-1})}{\partial t} = \frac{\partial \mathbf{I}}{\partial t} = 0$$

Remarque : dérivée d'un déterminant par un scalaire

La dérivée $\frac{\partial}{\partial t}\det(\mathbf{A})$ du déterminant d'une matrice $\mathbf{A} \in \mathbb{R}^{n*n}$ est la somme des n déterminants obtenus en remplaçant de toutes les manières possibles les éléments d'une ligne (colonne) de $\det(\mathbf{A})$ par leurs dérivées par rapport à *t*.

Si $\mathbf{x}$ est un vecteur, $\mathbf{x} \in \mathbb{R}^n$ et si $\mathbf{A}$ est une matrice, $\mathbf{A} \in \mathbb{R}^{m*n}$, leur intégration par rapport à un scalaire t donne :

$$\int \mathbf{x}\partial t = \begin{bmatrix} \int x_1 \partial t \\ \vdots \\ \int x_n \partial t \end{bmatrix} \text{ et } \int \mathbf{x}^T \partial t = \begin{bmatrix} \int x_1 \partial t & \cdots & \int x_n \partial t \end{bmatrix},$$

$$\int \mathbf{A}\partial t = \left[\int a_{ij} \partial t \right].$$

2. Définition d'une fonction vectorielle

Un ensemble de fonctions multivariables à valeurs réelles, $f_1(\mathbf{x})$, $f_2(\mathbf{x})$, ..., $f_m(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, peut être représenté par une unique fonction vectorielle, $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, $f_i : \mathbb{R}^n \to \mathbb{R}$:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) & f_2(\mathbf{x}) & \cdots & f_m(\mathbf{x}) \end{bmatrix}^T .$$

Si $\mathbf{f}(\mathbf{x})$ est une fonction vectorielle, $\mathbf{y} = \mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) & f_2(\mathbf{x}) & \cdots & f_m(\mathbf{x}) \end{bmatrix}^T$, où $\mathbf{x} \in \mathbb{R}^n$ et $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$.

La dérivée de la fonction $\mathbf{f}$ relative à une constante t vaut :

$\frac{\partial(\mathbf{yQy})}{\partial t} = \dot{\mathbf{y}}^T \mathbf{Q}\mathbf{y} + \mathbf{y}^T \mathbf{Q}\dot{\mathbf{y}} = 2\mathbf{y}^T \mathbf{Q}\dot{\mathbf{y}}$ avec $\dot{\mathbf{y}} = \frac{\partial \mathbf{y}}{\partial t} = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial t}$ et avec $\mathbf{Q}$ une matrice de pondération.

3. Le gradient d'une fonction scalaire

Soit une fonction multivariable à valeurs scalaires, définie par :

$$f(\mathbf{x}) = f(x_1, x_2, \cdots, x_m), \ f : \mathbb{R}^n \to \mathbb{R}.$$

$\mathbf{x}$ est un vecteur de dimension n, $\mathbf{x} \in \mathbb{R}^n$.

Le gradient de la fonction f par rapport à $\mathbf{x}$ a pour expression :

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_i} \end{bmatrix}_{n*1} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_i} & \frac{\partial f(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}^T .$$

Le vecteur gradient représente le vecteur des dérivées premières de la fonction f.

Le gradient de la fonction scalaire $f(\mathbf{x}) = \mathbf{x}^T \mathbf{Q}\mathbf{x}$, avec $\mathbf{Q}$ une matrice de pondération, $\mathbf{Q} \in \mathbb{R}^{n*n}$, $\mathbf{x} \in \mathbb{R}^n$, vaut :

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{x}^T \mathbf{Q}\,.$$

4. La Jacobienne d'une fonction vectorielle

Soit une fonction vectorielle $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) & f_2(\mathbf{x}) & \cdots & f_m(\mathbf{x}) \end{bmatrix}^T$, où $\mathbf{x} \in \mathbb{R}^n$ et $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$.

La matrice des dérivées premières de la fonction $\mathbf{f}$ relatives aux composantes x_j du vecteur $\mathbf{x}$ s'écrit :

$$\mathbf{J}(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \left[\frac{\partial f_i(\mathbf{x})}{\partial x_j} \right].$$

On appelle cette matrice la matrice Jacobienne de $\mathbf{f}$. Cette matrice s'écrit également :

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \nabla_{\mathbf{x}} f_1(\mathbf{x})^T \\ \vdots \\ \nabla_{\mathbf{x}} f_m(\mathbf{x})^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ & \ddots & \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

5. La Hessienne d'une fonction scalaire

Etant donné une fonction scalaire $f(\mathbf{x})$, $f : \mathbb{R}^n \to \mathbb{R}$. La matrice Hessienne de $f(\mathbf{x})$ relative au vecteur $\mathbf{x} \in \mathbb{R}^n$ est définie comme la matrice symétrique suivante :

$$\mathbf{H}(\mathbf{x}) = \left[\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right]_{n*n}.$$

En notant $\mathbf{f}(\mathbf{x})$ la fonction vectorielle qui vaut :

$$\mathbf{f}(\mathbf{x}) = \nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_i} \right]_{n*1} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_i} & \frac{\partial f(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}^T ,$$

La matrice Hessienne s'écrit :

$$\mathbf{H}(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} \left[\nabla_{\mathbf{x}} f(\mathbf{x}) \right]^T = \frac{\partial}{\partial \mathbf{x}} \left[\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right]^T = \nabla^2_{\mathbf{xx}} f(\mathbf{x})$$

$$= \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ & \ddots & \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix}_{n*n} .$$

La matrice Hessienne est la matrice des dérivées secondes de la fonction scalaire f relativement au vecteur $\mathbf{x} \in \mathbb{R}^n$. Elle représente la matrice Jacobienne du gradient de la fonction scalaire *f*.

6. La dérivation chainée

Soit la fonction scalaire composée : $f(\mathbf{x}) = h(g((\mathbf{x}))$, $\mathbf{x} \in \mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}$.

La règle de dérivation chaînée permet d'écrire :

$$\frac{\partial f}{\partial x_i} = \frac{\partial h}{\partial g} \frac{\partial g}{\partial x_i} .$$

Dans le cas général, si $\mathbf{h} : \mathbb{R}^r \to \mathbb{R}^m$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^r$ et $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, alors :

$$\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}) = \nabla_{\mathbf{g}} \mathbf{h}(\mathbf{g}(\mathbf{x})) \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}) .$$

7. Expansions en série de Taylor et de Maclaurin

Expansion en série de Taylor d'une fonction scalaire multivariable

Soit une fonction scalaire multivariable $f(\mathbf{x})$, $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$.

L'expansion en série de Taylor de $f(\mathbf{x})$ au voisinage d'un point $\mathbf{x}_0$ en termes de $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_0 = \Delta x_i = x_i - x_{0i}$ (i = 1, ..., n) s'exprime ainsi :

$$f(\mathbf{x}) = f(\mathbf{x}_0 + \Delta \mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \Delta x_i \Delta x_j + \varepsilon(\mathbf{x}_0)$$

où $\varepsilon(\mathbf{x}_0)$ sont les termes d'ordres supérieur impliquants les dérivées partielles d'ordre supérieur, négligeables lorsque $\|\Delta \mathbf{x}\|$ est suffisamment petit.

En notation vectorielle, on obtient :

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \Delta \mathbf{x}^T \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}_0) + \frac{1}{2} \Delta \mathbf{x}^T \nabla^2_{\mathbf{x}} f(\mathbf{x}_0) \Delta \mathbf{x} + \varepsilon(\mathbf{x}_0)$$

Série de Taylor à deux variables

Soit une fonction scalaire $f(x,y)$ à deux variables admettant des dérivées partielles d'ordre n+1 dans un certain domaine, pour des points (x_0,y_0) et (x,y).

L'expansion en série de Taylor de $f(x,y)$ au voisinage d'un point (x_0,y_0) est :

$$f(x,y) = f(x_0,y_0) + (x-x_0)\frac{\partial f(x_0,y_0)}{\partial x} + (y-y_0)\frac{\partial f(x_0,y_0)}{\partial y}$$
$$+\frac{(x-x_0)^2}{2}\frac{\partial^2 f(x_0,y_0)}{\partial x^2} + (x-x_0)(y-y_0)\frac{\partial^2 f(x_0,y_0)}{\partial x \partial y} + \frac{(y-y_0)^2}{2}\frac{\partial^2 f(x_0,y_0)}{\partial y^2} + \varepsilon(x_0,y_0)$$

Avec $\Delta x = x - x_0$ et $\Delta y = y - y_0$, on a :

$$f(x_0+\Delta x, y_0+\Delta y) = f(x_0,y_0) + \Delta x\frac{\partial f(x_0,y_0)}{\partial x} + \Delta y\frac{\partial f(x_0,y_0)}{\partial y}$$
$$+\frac{\Delta x^2}{2}\frac{\partial^2 f(x_0,y_0)}{\partial x^2} + \Delta x\Delta y\frac{\partial^2 f(x_0,y_0)}{\partial x \partial y} + \frac{\Delta y^2}{2}\frac{\partial^2 f(x_0,y_0)}{\partial y^2} + \varepsilon(x_0,y_0)$$

Par rapport au cas multivariable, on a posé : $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T$, $\mathbf{x}_0 = \begin{bmatrix} x_0 & y_0 \end{bmatrix}^T$.

Série de Taylor à une variable (à l'ordre *n*)

L'expansion en série de Taylor à l'ordre n d'une une fonction scalaire $f(x)$ dépendant d'une unique variable x est :

$$f(x) = f(x_0) + (x-x_0)\frac{\partial f(x_0)}{\partial x} + \frac{(x-x_0)^2}{2}\frac{\partial^2 f(x_0)}{\partial x^2} + \ldots + \frac{(x-x_0)^n}{n!}\frac{\partial^n f(x_0)}{\partial x^n} + \varepsilon(x_0),$$

c'est à dire, avec $\Delta x = x - x_0$ et $\Delta y = y - y_0$:

$$f(x_0+\Delta x) = f(x_0) + \Delta x\frac{\partial f(x_0)}{\partial x} + \frac{\Delta x^2}{2}\frac{\partial^2 f(x_0)}{\partial x^2} + \ldots + \frac{\Delta x^n}{n!}\frac{\partial^n f(x_0)}{\partial x^n} + \varepsilon(x_0).$$

Série de Maclaurin à une variable (à l'ordre *n*)

L'expansion en série de Maclaurin est un cas particulier de celle de la série de Taylor. Cette particularité est marquée par le fait que l'expansion se fait autour de $x_0 = 0$ ce qui revient à dire que $\Delta x = x$:

$$f(x) = f(0) + x\frac{\partial f(0)}{\partial x} + \frac{x^2}{2}\frac{\partial^2 f(0)}{\partial x^2} + \ldots + \frac{x^n}{n!}\frac{\partial^n f(0)}{\partial x^n}.$$

Série de Maclaurin à deux variables

Soit une fonction scalaire $f(x, y)$ à deux variables. Son expansion en série de Maclaurin s'écrit :

$$f(x,y) = f(0,0) + x\frac{\partial f(0,0)}{\partial x} + y\frac{\partial f(0,0)}{\partial y} + \frac{x^2}{2}\frac{\partial^2 f(0,0)}{\partial x^2} + xy\frac{\partial^2 f(0,0)}{\partial x \partial y} + \frac{y^2}{2}\frac{\partial^2 f(0,0)}{\partial y^2} + \ldots .$$

Expansion en série de Maclaurin d'une fonction scalaire multivariable

Si $f(\mathbf{x})$ est une fonction scalaire multivariable, $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$, son expansion en série de Maclaurin au voisinage de $\mathbf{x}_0$ et en termes de $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0 = \Delta x_i = x_i - x_{0i}$ $(i = 1, \ldots, n)$ est donné par :

$$f(\mathbf{x}) = f(0) + \sum_{i=1}^{n} \frac{\partial f(0)}{\partial x_i} x_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f(0)}{\partial x_i \partial x_j} x_i x_j + \ldots .$$

La notation vectorielle est plus adaptée :

$$f(\mathbf{x}) = f(0) + \mathbf{x}^T \nabla_{\mathbf{x}} \mathbf{f}(0) + \frac{1}{2}\mathbf{x}^T \nabla_{\mathbf{x}}^2 f(0)\mathbf{x} + \ldots .$$

Bibliographie

Ayres, F., Jr., 1991. Matrices : cours et problèmes, McGraw Hill, New York.

Cairoli, R., 1996. Algèbre linéaire, Presses Polytechniques et Universitaires Romandes, Lausanne.

Christol, G., Pilibossian, P., and Yammine, S., 1996. Algèbre, cours et exercices corrigés, Ellipses-Edition Marketing, Paris.

Kohonen, T., 1984. Self-Organization and Associative Memory, Springer-Verlag, Berlin.

Rotella, F., and Borne, P., 1995. Théorie et pratique du calcul matriciel, Editions Technip, Paris.

Stoer, J., and Bulirsch, R., 1993. Introduction to Numerical Analysis, Springer-Verlag, New York.

Strang G., 1993. Introduction to Linear Algebra, .Welleley Cambridge Press, Wellesley.

Annexe

Alphabet Grec

Alpha	A	α	Nu	N	ν
Bêta	B	β	Xi	Ξ	ξ
Gamma	Γ	γ	Omicron	O	o
Delta	Δ	δ	Pi	Π	π
Epsilon	E	ε	Rhô	P	ρ
Zêta	Z	ζ	Sigma	Σ	σ
Êta	H	η	Tau	T	τ
Thêta	Θ	θ	Upsilon	Y	υ
Iota	I	ι	Phi	Φ	ϕ, φ
Kappa	K	κ	Khi	X	χ
Lambda	Λ	λ	Psi	Ψ	ψ
Mu	M	μ	Oméga	Ω	ω