

# Math S2 PeiP

## Chapitre 6

### Quelques applications de l'algèbre linéaire

Michel Rumin

Dans ce dernier chapitre nous allons illustrer les techniques d'algèbre linéaire que nous avons abordées par des exemples issus de différents domaines scientifiques. Cela commencera par une étape de modélisation, ou de mise en équation, propre à chaque domaine. Des exemples explicites seront traités en TP de math avec le logiciel Scilab.

## 1 Analyse entrée-sortie en économie

Nous commençons nos illustrations par un problème d'économie qui met naturellement en jeu de grands systèmes linéaires à résoudre.

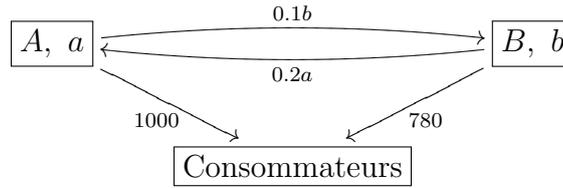
### 1.1 De quoi s'agit-il ?

L'économie d'un pays dépend de la production d'un certains nombres de secteurs, par exemple l'énergie, l'agriculture, l'industrie, etc, mais aussi de la consommation.

L'économiste américain Wassily Leontief (1905-1999), prix Nobel d'économie en 1973, s'est intéressé au problème suivant, qu'elle doit être la production de chaque secteur pour que la demande totale soit satisfaite ?

Par exemple, pour fixer les idées, supposons qu'il y ait deux secteurs A et B et que la demande des consommateurs soit respectivement de 1000 et 780 millions d'€ de ces produits par an. Il n'est pas vrai en général qu'il suffise que A et B produisent respectivement pour 1000 et 780 millions d'€ pour satisfaire la demande ! Les secteurs peuvent aussi dépendre l'un de l'autre. Par exemple si A produit de l'électricité, et B des câbles, chaque secteur doit consommer une partie de la production de l'autre secteur pour produire lui même. Si par exemple A a besoin de 20 centimes d'euro de B par € de production et B de 10 centimes de A, le problème est le suivant : on cherche les productions a et b (en millions d'€) de A et B telles que

$$\begin{cases} a = 0,1b + 1000 \\ b = 0,2a + 780 \end{cases} .$$



Ce système a pour solution  $a = 1100$  et  $b = 1000$  ce qui est plus que la demande destinée à la consommation. Le surplus de production est consommé en interne.

## 1.2 Modélisation

Pour traiter des problèmes plus complexes, on utilise des vecteurs et des matrices. Supposons qu'il y ait  $n$  secteurs de productions  $I_1, I_2, \dots, I_n$ , de production  $x_1, x_2, \dots, x_n$ .

– Le vecteur de production est alors  $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ ,

– et le vecteur de demande des consommateurs est  $C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$ .

– Chaque secteur  $I_j$  a besoin de  $a_{ij}$  € de biens de  $I_i$  par euro produit. Les coefficients  $A = (a_{ij})$  forment le tableau d'entrée-sortie de l'économie. Chaque colonne  $C_j$  de  $A$  représente la demande (interne) de  $I_j$  par € produit.

La demande totale pour produire  $X$  et satisfaire une demande de consommateurs  $C$  est donc

$$x_1 C_1 + x_2 C_2 + \dots + x_n C_n + C = AX + C.$$

La production satisfait exactement à la demande si

$$X = AX + C \Leftrightarrow (I_n - A)X = C \quad (1)$$

où  $I_n$  est la matrice identité.

## 1.3 Un résultat

Pour que l'on puisse trouver une production répondant à la demande, il faut que le système (1) possède une solution. C'est le cas si la matrice  $I_n - A$  est inversible. On va voir que c'est le cas si l'économie est *productive*.

**Définition 1.1.** On dit que l'économie est productive si pour chaque secteur  $I_j$  de l'économie la production de 1€ de bien nécessite moins d'1€ de dépense! Autrement dit, la somme des coefficients de chaque colonne  $C_j$  de la matrice  $A$  est inférieure à 1 :

$$\forall j, \sum_{i=1}^n a_{ij} < 1.$$

On a le résultat rassurant :

**Théorème 1.2.** Si l'économie est productive alors  $I_n - A$  est inversible, et

$$X = (I_n - A)^{-1}(C)$$

est la production nécessaire pour satisfaire la demande de consommation  $C$ .

*Démonstration.* Comme  $I_n - A$  est une matrice carrée, elle est inversible si son noyau est nul. Soit  $X = (x_1, x_2, \dots, x_n)$  tel que  $(I_n - A)X = 0 \Leftrightarrow X = AX$ . On a alors pour tout  $i$ ,

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

D'où, comme les coefficients de  $A$  sont positifs par définition

$$|x_i| \leq \sum_{j=1}^n a_{ij}|x_j| \implies \sum_{i=1}^n |x_i| \leq \sum_{j=1}^n \left( \sum_{i=1}^n a_{ij} \right) |x_j| < \sum_{j=1}^n |x_j|,$$

par hypothèse sur  $A$ , si  $X \neq 0$ . Ceci est impossible et  $\ker(I_n - A) = \{0\}$ .  $\square$

## 1.4 Conclusion.

Un des intérêts de cette analyse est de permettre de prévoir les changements de productions nécessaires en cas de changement de demande de consommation en biens d'un secteur  $I_j$  donné. Ils sont simplement donnés par la  $j$ -ème colonne de  $(I_n - A)^{-1}$ . Même si l'analyse entrée-sortie n'est pas parfaite, elle fut acceptée et utilisée pour modéliser les économies de plus de 50 pays après la seconde guerre mondiale!

Le lecteur intéressé par ce sujet peut consulter Wikipedia par exemple.

## 2 Évolution de systèmes et matrices de transition

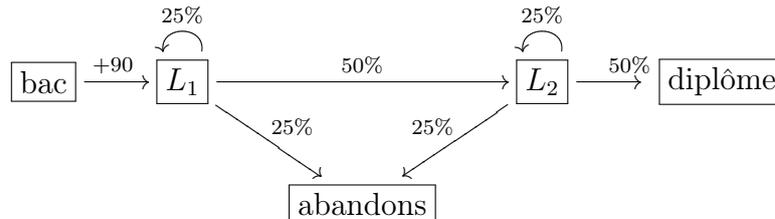
Nous allons utiliser le calcul matriciel pour modéliser l'évolution d'un système possédant un nombre fini « d'états ». Nous traitons un premier exemple pour se familiariser avec le thème.

## 2.1 Un exemple : les effectifs dans un cycle d'étude

**Le problème.** Nous voulons étudier l'évolution des effectifs des étudiants dans un cycle d'étude de deux ans  $L_1$  et  $L_2$ . On a les données suivantes :

- chaque année, 90 nouveaux étudiants arrivent en  $L_1$  ;
- 50% des étudiants de  $L_1$  passent en  $L_2$ , 25% redoublent et 25% abandonnent ;
- 50% des étudiants de  $L_2$  ont le diplôme, 25% redoublent et 25% abandonnent.

On peut résumer les données sous forme d'un graphe :



On voudrait connaître :

1. l'évolution des effectifs, se stabilisent-ils, explosent-ils ?
2. le taux de réussite global du cycle.

**Modélisation.** On regroupe les effectifs de l'année  $n$  dans un vecteur

$$X(n) = \begin{pmatrix} x_1(n) \\ x_2(n) \end{pmatrix},$$

et on écrit une relation de récurrence entre  $X(n+1)$  et  $X(n)$ . On a

$$X(n+1) = \begin{pmatrix} \frac{1}{4}x_1(n) + 90 \\ \frac{1}{2}x_1(n) + \frac{1}{4}x_2(n) \end{pmatrix} = \begin{pmatrix} 1/4 & 0 \\ 1/2 & 1/4 \end{pmatrix} X(n) + \begin{pmatrix} 90 \\ 0 \end{pmatrix}$$

c'est-à-dire

$$\left| \begin{array}{l} X(n+1) = MX(n) + E, \\ \text{avec } M = \begin{pmatrix} 1/4 & 0 \\ 1/2 & 1/4 \end{pmatrix} \text{ et } E = \begin{pmatrix} 90 \\ 0 \end{pmatrix} \text{ est le vecteur des entrées.} \end{array} \right. \quad (2)$$

C'est une relation de récurrence sur le vecteur  $X(n)$ . Cette suite est déterminée par les effectifs initiaux  $X(0)$ .

**Résolution dans le cas stable.** On peut facilement implanter le calcul de la suite  $X(n)$  avec Scilab pour différentes *conditions initiales*  $X(0)$ , et en étudier numériquement l'évolution. On peut aussi résoudre théoriquement le problème.

- On recherche d'abord les **effectifs stables possibles**. Dans ce cas, on doit avoir

$$X(n+1) = X(n) = X.$$

D'où, d'après (2)

$$X = MX + E \Leftrightarrow (I_2 - M)X = E.$$

La matrice  $I_2 - M = \begin{pmatrix} 3/4 & 0 \\ -1/2 & 3/4 \end{pmatrix}$  est échelonnée et inversible, de déterminant  $9/16$  et d'inverse

$$(I_2 - M)^{-1} = \begin{pmatrix} 4/3 & 0 \\ 8/9 & 4/3 \end{pmatrix},$$

d'après l'expression générale vue dans le chapitre précédent. On a donc

$$X = (I_2 - M)^{-1}(E) = \begin{pmatrix} 120 \\ 80 \end{pmatrix}, \quad (3)$$

c'est-à-dire que si les effectifs sont stables, on a

- 120 étudiants en  $L_1$  ;
- et 80 en  $L_2$ .
- Dans ce cas, on a de plus un taux de réussite globale

$$\frac{\text{diplômés}}{\text{entrants}} = \frac{50\% \times 80}{90} = \frac{40}{90} \simeq 44\%,$$

ce qui peut paraître élevé à première vue par rapport aux taux de réussite des années prisent séparément.<sup>1</sup>

**Évolution des effectifs en général.** Pour étudier l'évolution générale des effectifs, nous les comparons aux effectifs stables  $X = \begin{pmatrix} 120 \\ 80 \end{pmatrix}$ , en considérant la différence

$$\Delta(n) = X(n) - X.$$

Comme  $E = (I_2 - M)X$ , la récurrence (2) s'écrit

$$\begin{aligned} \Delta(n+1) &= X(n+1) - X = MX(n) + (I_2 - M)X - X \\ &= MX(n) - MX = M\Delta(n). \end{aligned}$$

On a donc par une récurrence immédiate

$$\Delta(n) = M^n \Delta(0). \quad (4)$$

Il faut calculer les puissances de  $M$ . On trouve

$$M = \frac{1}{4} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \quad M^2 = \frac{1}{16} \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}, \quad M^3 = MM^2 = \frac{1}{4^3} \begin{pmatrix} 1 & 0 \\ 6 & 1 \end{pmatrix}.$$

Ceci nous fait deviner que

$$M^n = \frac{1}{4^n} \begin{pmatrix} 1 & 0 \\ 2n & 1 \end{pmatrix},$$

1. Ce qui se passe, c'est que les étudiants de cette filière peuvent avoir leur diplôme « à l'usure », en redoublant plusieurs fois.

ce qui se vérifie facilement par récurrence.

Comme les coefficients de  $M^n$  tendent tous (très vite) vers 0 lorsque  $n \rightarrow +\infty$ , on en déduit que les composantes de  $\Delta(n) = X(n) - X = M^n \Delta(0)$  tendent vers 0 quand  $n \rightarrow 0$ .

En conclusion, on voit que les effectifs du cycle convergent vers les effectifs stables : 120 étudiants en  $L_1$  et 90 en  $L_2$ , et ceci quels que soient les effectifs initiaux.

## 2.2 Matrices de transition et évolution de systèmes

L'exemple précédent décrit l'évolution d'une population à deux états (les deux années du cycle) et dont les transitions d'un état à l'autre sont données par des probabilités. On peut modéliser systématiquement ce genre de problème par du calcul matriciel.

### Les chaînes de Markov.

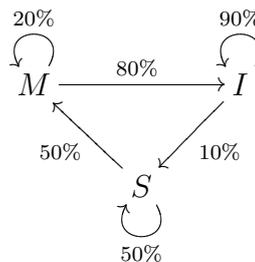
On se donne un système qui peut être dans un certain nombre d'états  $E_1, E_2, \dots, E_n$ . On suppose qu'il évolue d'un instant  $t = n$  au suivant en suivant des *probabilités de transition*.

On note  $p_{ij}$  = probabilité de passer de l'état  $E_j$  à l'état  $E_i$  entre  $t = n$  et  $t = n + 1$ .

Prenons par exemple le problème d'épidémie suivant, à trois états. On peut être malade (M), immunisé (I), ou sain mais non immunisé (S). D'une semaine à l'autre,

- 20% des malades le restent, 80% guérissent et sont immunisés ;
- 90% des immunisés le restent, 10% ne le sont plus mais restent sains (mutation du virus par exemple) ;
- 50% des personnes saines mais non immunisées le restent, 50% tombent malades.

L'ensemble des données peut se représenter dans un graphe.



C'est joli, mais pour faire des calculs et quand le nombre des états augmente, il est bien plus efficace de représenter les données sous forme matricielle.

**Définition 2.1.** La matrice de transition de la chaîne de Markov est la matrice  $M = (p_{ij})$ , où  $p_{ij}$  représente la probabilité de passer de  $E_j$  à  $E_i$  entre un instant et le suivant.

Cela donne dans notre exemple avec  $E_1 = M, E_2 = I, E_3 = S$

$$M = \begin{pmatrix} 0,2 & 0 & 0,5 \\ 0,8 & 0,9 & 0 \\ 0 & 0,1 & 0,5 \end{pmatrix}.$$

On remarque que la somme des coefficients de chaque colonne fait toujours 1, c'est-à-dire que pour tout  $j$ ,

$$\sum_{i=1}^n p_{ij} = 1.$$

C'est une propriété générale des matrices de transition lorsque la liste des états est complète. Ces sommes représentent la probabilité totale de passer d'un état donné à n'importe quel autre.

**Évolution du système.** On s'intéresse à l'évolution des différentes proportions  $x_i$  du système qui sont dans les états  $E_i$  à chaque instant. On fabrique ainsi un vecteur

$$X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}$$

Par exemple dans le cas précédent on part de  $X(0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$  avant l'arrivée de l'épidémie,

et si  $X = \begin{pmatrix} 0.5 \\ 0.2 \\ 0.3 \end{pmatrix}$ , alors 50% de la population est malade, 20% est immunisée et 30% saine mais non immunisée.<sup>2</sup>

L'évolution de ces probabilités est très simple à calculer avec la matrice de transition  $M$ .

**Théorème 2.2.** On a

$$X(t+1) = MX(t)$$

*Démonstration.* La composante  $x_i(t+1)$  de  $X(t+1)$  est la probabilité d'être dans l'état  $E_i$  à l'instant  $t+1$ . On a donc en décomposant le calcul suivant les états possibles à l'instant  $t$  :

$$\begin{aligned} x_i(t+1) &= \sum_{j=1}^n P(E_i(t+1) \cap E_j(t)) \\ &= \sum_{j=1}^n P(E_i(t+1) \text{ sachant } E_j(t)) \times P(E_j(t)) \\ &= \sum_{j=1}^n p_{ij} x_j(t) = (MX(t))_i \end{aligned}$$

2. Attention pour écrire les nombres en Scilab : le séparateur décimal est le point, et la virgule représente un changement de colonne !

□

On en déduit immédiatement par récurrence que :

**Corollaire 2.3.** *On a  $X(n) = M^n X(0)$ .*

Ces calculs s'effectuent très vite avec Scilab, et permettent donc d'observer l'évolution de l'état du système. Dans le cas du modèle d'épidémie par exemple, on observe un phénomène de convergence vers une *état stationnaire*  $X$  lorsque  $n \rightarrow +\infty$ .<sup>3</sup> De plus cet état limite est indépendant de la situation de départ. C'est un phénomène assez général.

**Le théorème de Perron-Frobenius** Si  $M$  est une matrice de transition agissant sur un système. Un état  $X$  du système est un vecteur de composantes  $x_i \geq 0$  et satisfaisant

$$\sum_{i=1}^n x_i = 1.$$

Dans ce cas, les  $x_i$  s'interprètent comme les probabilités pour le système d'être dans un état  $E_i$ , ou encore si on étudie une population, la proportion des individus dans l'état  $E_i$ .

On dit qu'un état est *stationnaire* si  $X(n) = X$ . Comme on a toujours  $X(n+1) = MX(n)$ , d'après le théorème 2.2, on a donc

$$MX = X$$

pour  $X$  stationnaire.

**Définition 2.4.** On dit que la matrice de transition est *irréductible* si il existe une puissance  $k$  telle que **tous** les coefficients de  $M^k$  sont  $> 0$ .

Par exemple dans le modèle d'épidémie ci-dessus,  $M$  est irréductible puisque

$$M^2 = \begin{pmatrix} 0.04 & 0.05 & 0.35 \\ 0.88 & 0.81 & 0.4 \\ 0.08 & 0.14 & 0.25 \end{pmatrix}.$$

L'hypothèse d'irréductibilité permet de montrer que l'évolution du système ne possède pas de « cycle » ni de « cul-de-sac ». On a alors le résultat général suivant.

**Théorème 2.5** (Théorème de Perron-Frobenius). *Si  $M$  est irréductible, alors il existe un unique état stationnaire  $X$ , et de plus, quel que soit l'état initial du système, on a*

$$X(n) \rightarrow X \text{ lorsque } n \rightarrow +\infty.$$

3. À voir en TP Scilab.

Le résultat est admis car la démonstration est du niveau L3 math. Cela ne doit pas empêcher de le connaître ni de l'utiliser !

## 2.3 Sur le PageRank de Google

Les notions précédentes ont de nombreuses applications, en biologie, statistique, etc. Une application assez récente et spectaculaire est liée au PageRank de Google.

Lorsque l'on envoie une requête à un moteur de recherche, celui-ci doit d'abord rechercher tous les sites référencés liés à votre liste de mots, et ensuite, *classer* les résultats pour les présenter. Il faut donc en quelque sorte attribuer une « note » aux documents, sans pour autant qu'aucun humain ne les ait tous lus pour juger de leur intérêt relatif ! Cette note est le *PageRank*, dont l'efficacité fut une raison du succès de Google.

L'idée est la suivante. Le réseau internet est modélisé par un graphe gigantesque. Ces sommets sont les pages référencées ( $N \simeq 25 \times 10^9$  avec les chiffres de 2007). Des liens pointent de ces pages vers d'autres pages.

On imagine maintenant que l'utilisateur clique *au hasard* sur ces liens et on essaye de déterminer l'importance d'une page donnée en fonction de la fréquence ou probabilité de tomber dessus en un grand nombre de clics. Cela ressemble aux problèmes précédents de chaîne de Markov, avec des états numérotés par les différentes pages, mais cela ne marche pas tel quel. En effet, certaines pages sont des « culs-de-sac » car elles ne contiennent pas de liens. Pour fabriquer une matrice de transition avec le réseau, on rajoute des liens virtuels un peu partout, mais de très faible poids, à l'aide d'un paramètre  $0 < q < 1$  supplémentaire.<sup>4</sup>

Supposons qu'il y ait  $N$  pages  $P_j$  référencées, et que chaque page  $P_j$  contienne  $k_j$  liens. Les fondateurs de Google, Brin et Page, ont défini la « matrice du réseau »  $G$ , de taille  $N \times N$  par

$$G_{ij} = \begin{cases} \frac{1-q}{k_j} + \frac{q}{N} & \text{si } P_j \text{ pointe vers } P_i, \\ \frac{q}{N} & \text{si } P_j \text{ ne pointe pas vers } P_i \text{ et } k_j \neq 0, \\ \frac{1}{N} & \text{si } k_j = 0. \end{cases}$$

**Proposition 2.6.** *La matrice  $G$  est une matrice de transition.*

*Démonstration.* On a pour tout  $j$ ,

$$\sum_{i=1}^N g_{ij} = \begin{cases} k_j \left( \frac{1-q}{k_j} + \frac{q}{N} \right) + (N - k_j) \frac{q}{N} = 1 & \text{si } k_j \neq 0, \\ N \times \frac{1}{N} = 1 & \text{si } k_j = 0. \end{cases}$$

4. Il semble que  $q \simeq 0,15$  soit utilisé en pratique.

□

Comme de plus, on a  $g_{ij} > 0$  pour tout  $i, j$ , la matrice  $G$  est irréductible et le théorème 2.5 s'applique. On a donc qu'il existe un état stationnaire unique  $R$  tel que quelque soit l'état  $X$  d'origine

$$G^n X \rightarrow R.$$

La composante  $R_i$  de  $R$  représente la fréquence de visite de la page  $P_i$ , *en naviguant au hasard* sur internet. C'est une mesure de « l'importance » de cette page, et est un ingrédient important du PageRank de la page  $P_i$ . Bien sûr, ce n'est pas le seul, le business ayant son mot à dire !

Les étudiants que ces notions intéressent peuvent consulter les pages Wikipedia des chaînes de Markov ou du PageRank.

### 3 La méthode des moindres carrés et quelques applications

Nous changeons de domaine pour étudier comment des techniques d'algèbre linéaire permettent d'analyser des collections de données en cherchant à les approcher le mieux possible par des fonctions simples. L'exemple le plus simple est celui de la régression linéaire.

#### 3.1 La régression linéaire

**Le cadre.** Étant données deux séries de données de même taille :  $(x_1, x_2, \dots, x_n)$  et  $(y_1, y_2, \dots, y_n)$ , on cherche si les  $y_i$  s'approchent bien par une expression affine des  $x_i$ , c'est-à-dire s'il existe des coefficients réels  $a$  et  $b$  tels que

$$y_i \simeq ax_i + b.$$

Géométriquement, il s'agit de trouver une droite affine  $D$  du plan, d'équation  $y = ax + b$ , et dont les  $n$  points  $(x_1, y_1), (x_2, y_2), (x_n, y_n)$  soient le plus proche possible.

Il faut bien sûr préciser ce que l'on entend par « le plus proche possible ». Un choix très pratique est celui des « moindres carrés ». Il s'agit de trouver  $a$  et  $b$  tels que la somme des carrés

$$\Sigma(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2 \quad (5)$$

soit la plus petite possible.

La droite qui réalise ce minimum s'appelle la *droite de régression linéaire*. Les fonctions statistiques des calculatrices calculent ces coefficients  $a$  et  $b$ . Nous allons voir une approche de ce problème qui se généralise à d'autres type de « fitting » par d'autres fonctions, et qui conduit à des calculs que Scilab effectue efficacement.

**Moindres carrés et projection orthogonale.** L'idée est de considérer les données comme deux vecteurs de  $\mathbb{R}^n$

$$X = (x_1, x_2, \dots, x_n) \quad \text{et} \quad Y = (y_1, y_2, \dots, y_n).$$

Un troisième vecteur utile est  $V = (1, 1, \dots, 1)$ . Alors le vecteur

$$Z = Y - aX - bV$$

a pour coordonnées

$$z_i = y_i - ax_i - b.$$

En particulier, la norme euclidienne au carré de  $Z$

$$\|Z\|^2 = \sum_{i=1}^n z_i^2 = \sum_{i=1}^n (y_i - ax_i - b)^2 = \Sigma(a, b) \quad (6)$$

est précisément la quantité que l'on cherche à minimiser. On remarque aussi que le vecteur  $aX + bV$  est dans le plan vectoriel  $P$  engendré par  $X$  et  $V$ . En fait,  $P$  est un plan, lorsque  $X$  n'est pas proportionnel à  $V$ , c'est-à-dire lorsque les  $x_i$  ne sont pas tous égaux. Dans ce cas dégénéré, les données sont toutes sur une droite verticale  $x = c$ .

En conclusion, dans le cas général où les  $x_i$  ne sont pas tous égaux, le problème de la régression linéaire a une interprétation géométrique claire dans  $\mathbb{R}^n$  : il s'agit de trouver dans le plan  $P$ , le vecteur  $aX + bV$  le plus proche de  $Y$ .

On va montrer que ce vecteur existe, est unique, et calculable. Il s'appelle la *projection orthogonale* de  $Y$  sur  $P$ .

### 3.2 Les projections

**Définition générale.** La définition générale d'une projection dans  $\mathbb{R}^n$  s'appuie sur la donnée de deux espaces supplémentaires :

$$\mathbb{R}^n = E \oplus F.$$

Tout vecteur  $\vec{u}$  de  $\mathbb{R}^n$  se décompose alors de manière unique en

$$\vec{u} = \vec{v} + \vec{w} \quad \text{avec} \quad \vec{v} \in E \text{ et } \vec{w} \in F.$$

**Définition 3.1.** La projection  $p$  de  $\mathbb{R}^n$  sur  $E$  parallèlement à  $F$  (ou le long de  $F$ ) est l'application définie par

$$\begin{aligned} p : \mathbb{R}^n = E \oplus F &\longrightarrow \mathbb{R}^n \\ \vec{u} = \vec{v} + \vec{w} &\longmapsto \vec{v} = p(\vec{u}) \end{aligned}$$

On a clairement les propriétés suivantes.

**Proposition 3.2.**  $p$  est une application linéaire, dont le noyau est  $\ker p = F$  et l'image  $\text{Im } p = E$ . De plus,  $p$  satisfait  $p \circ p = p$ .

**Exemples dans  $\mathbb{R}^2$  et  $\mathbb{R}^3$ .** • Soient  $D_1$  et  $D_2$  deux droites vectorielles distinctes de  $\mathbb{R}^2$ . On a alors  $\mathbb{R}^2 = D_1 \oplus D_2$ . Si  $D_1$  est engendrée par  $\vec{u}_1$ , et  $D_2$  a pour équation  $ax + by = 0$ ,

on peut écrire explicitement la projection  $p$  sur  $D_1$  le long de  $D_2$ .

$$\left| \begin{array}{l} \text{Si } \vec{v} = (x, y) \text{ et } \ell(\vec{v}) = ax + by, \text{ on a } p(\vec{v}) = \frac{\ell(\vec{v})}{\ell(\vec{u}_1)} \vec{u}_1. \end{array} \right.$$

En effet, la formule proposée est clairement une application linéaire. De plus, on a

$$p(\vec{u}_1) = \vec{u}_1 \quad \text{et} \quad p(\vec{u}_2) = \vec{0} \quad \text{si} \quad \vec{u}_2 \in D_2 = \ker \ell,$$

et donc  $p(\lambda_1 \vec{u}_1 + \vec{u}_2) = \lambda_1 \vec{u}_1$  est bien l'application cherchée.

- Dans un  $\mathbb{R}^3$ , un plan  $P$  et une droite  $D$  telle que  $P \cap D = \{\vec{0}\}$  sont supplémentaires :

$$\mathbb{R}^3 = D \oplus P.$$

Si  $D$  est engendré par  $\vec{u}_1$  et  $P$  a pour équation  $\ell(x, y, z) = ax + by + cz = 0$ , alors la formule

$$p_D(\vec{v}) = \frac{\ell(\vec{v})}{\ell(\vec{u}_1)} \vec{u}_1$$

définit de nouveau la projection sur  $D$ , parallèlement à  $P$ .

La projection  $p_P$  sur  $P$  le long de  $D$  s'obtient par

$$p_P(\vec{v}) = \vec{v} - p_D(\vec{v}) = \vec{v} - \frac{\ell(\vec{v})}{\ell(\vec{u}_1)} \vec{u}_1.$$

En effet, d'après la définition 3.1, on a de manière générale

$$p_E + p_F = \text{Id} \quad \text{si} \quad \mathbb{R}^n = E \oplus F.$$

### 3.3 L'orthogonal d'un espace et les projections orthogonales

Dans  $\mathbb{R}^n$ , si  $\vec{u} = (x_1, x_2, \dots, x_n)$  et  $\vec{v} = (y_1, y_2, \dots, y_n)$  on note

$$\langle \vec{u}, \vec{v} \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n \tag{7}$$

le produit scalaire de  $\vec{u}$  et  $\vec{v}$ . De la même façon que dans  $\mathbb{R}^2$  ou  $\mathbb{R}^3$ , on dit que les vecteurs  $\vec{u}$  et  $\vec{v}$  sont *orthogonaux* si  $\langle \vec{u}, \vec{v} \rangle = 0$ .

On peut définir l'orthogonal d'un sev  $E$  de  $\mathbb{R}^n$ .

**Définition 3.3.** Soit  $E$  un sev de  $\mathbb{R}^n$ . L'orthogonal de  $E$ , noté  $E^\perp$ , est l'ensemble des vecteurs de  $\mathbb{R}^n$  orthogonaux à tous les vecteurs de  $E$  :

$$E^\perp = \{ \vec{u} \in \mathbb{R}^n, \langle \vec{u}, \vec{v} \rangle = 0 \text{ pour tout } \vec{v} \in E \}.$$

Par linéarité du produit scalaire par rapport à  $\vec{u}$  ou  $\vec{v}$ , on a :

**Proposition 3.4.** 1.  $E^\perp$  est un sev de  $\mathbb{R}^n$ .

2. De plus, si  $E$  est engendré par  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p$ , alors

$$E^\perp = \{ \vec{u} \in \mathbb{R}^n, \langle \vec{u}, \vec{v}_1 \rangle = \langle \vec{u}, \vec{v}_2 \rangle = \dots = \langle \vec{u}, \vec{v}_p \rangle = 0 \}.$$

**Les sommes directes orthogonales.** On arrive au point important, qui permet de définir les projections orthogonales en général.

**Théorème 3.5.** Pour  $E$  sev de  $\mathbb{R}^n$ , on a toujours

$$\mathbb{R}^n = E \oplus E^\perp.$$

*Démonstration.* Soient  $\vec{u}$  dans  $\mathbb{R}^n$  et  $B = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p)$  une base de  $E$ . On cherche

$$\begin{aligned} \vec{v} &= x_1 \vec{v}_1 + x_2 \vec{v}_2 + \dots + x_p \vec{v}_p \in E \quad \text{tel que} \quad \vec{u} - \vec{v} \in E^\perp \\ &\Leftrightarrow \langle \vec{u} - \vec{v}, \vec{v}_1 \rangle = 0 = \langle \vec{u} - \vec{v}, \vec{v}_2 \rangle = \dots = \langle \vec{u} - \vec{v}, \vec{v}_p \rangle \\ &\Leftrightarrow (S) \begin{cases} \langle \vec{v}, \vec{v}_1 \rangle = \langle \vec{u}, \vec{v}_1 \rangle \\ \langle \vec{v}, \vec{v}_2 \rangle = \langle \vec{u}, \vec{v}_2 \rangle \\ \vdots \\ \langle \vec{v}, \vec{v}_p \rangle = \langle \vec{u}, \vec{v}_p \rangle \end{cases} \end{aligned} \quad (8)$$

(S) est un système linéaire à  $p$  équations et  $p$  inconnues  $x_1, x_2, \dots, x_p$ . Il est de Cramer et possède une solution unique ssi il est de rang  $p$ , c'est-à-dire si  $\vec{v} = \vec{0}$  est la seule solution du système homogène associé (pas d'inconnue non-principale). On peut aussi dire que l'application linéaire  $\mathbb{R}^p \rightarrow \mathbb{R}^p$  associée à (S) est bijective ssi son noyau est nul.

Or  $\vec{v} \in E$  satisfait

$$\langle \vec{v}, \vec{v}_1 \rangle = 0 = \langle \vec{v}, \vec{v}_2 \rangle = \dots = \langle \vec{v}, \vec{v}_p \rangle$$

ssi  $\vec{v} \in E^\perp$  par la proposition 3.4. Mais alors  $\vec{v} \in E \cap E^\perp$  est orthogonal à lui même

$$\langle \vec{v}, \vec{v} \rangle = \|\vec{v}\|^2 = 0 \Leftrightarrow \vec{v} = \vec{0}$$

ce qu'il fallait montrer. □

**Corollaire 3.6.** *En conclusion, la projection orthogonale de  $\vec{u}$  sur  $E$  est donnée par le vecteur  $\vec{v} = x_1\vec{v}_1 + x_2\vec{v}_2 + \dots + x_p\vec{v}_p \in E$  satisfaisant le système (S).*

**Projection, moindres carrés et théorème de Pythagore dans  $\mathbb{R}^n$ .** Le théorème de Pythagore s'étend naturellement dans  $\mathbb{R}^n$ .

**Théorème 3.7.** *Si  $\vec{u}$  et  $\vec{v}$  sont deux vecteurs orthogonaux de  $\mathbb{R}^n$ , alors on a*

$$\|\vec{u} + \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2.$$

*Démonstration.* On a

$$\begin{aligned} \|\vec{u} + \vec{v}\|^2 &= \langle \vec{u} + \vec{v}, \vec{u} + \vec{v} \rangle \\ &= \langle \vec{u}, \vec{u} \rangle + \langle \vec{u}, \vec{v} \rangle + \langle \vec{v}, \vec{u} \rangle + \langle \vec{v}, \vec{v} \rangle \\ &= \|\vec{u}\|^2 + \|\vec{v}\|^2 \quad \text{puisque } \langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle = 0. \end{aligned}$$

□

On revient enfin à notre problème initial de « moindres carrés » rencontré dans (6).

**Théorème 3.8.** *Soit  $E$  un sev de  $\mathbb{R}^n$  et  $\vec{u} \in \mathbb{R}^n$  quelconque. Alors la projection orthogonale  $p_E(\vec{u})$  de  $\vec{u}$  sur  $E$  est le vecteur de  $E$  le plus proche de  $\vec{u}$ , c'est-à-dire que*

$$\forall \vec{v} \in E, \text{ on a } \|\vec{v} - \vec{u}\|^2 \geq \|p_E(\vec{u}) - \vec{u}\|^2,$$

*avec égalité ssi  $\vec{v} = p_E(\vec{u})$ .*

*Démonstration.* On écrit  $\vec{v} - \vec{u} = (\vec{v} - p_E(\vec{u})) + (p_E(\vec{u}) - \vec{u})$ , avec

$$\vec{v} - p_E(\vec{u}) \in E \quad \text{orthogonal à} \quad p_E(\vec{u}) - \vec{u} \in E^\perp.$$

On a donc par Pythagore

$$\|\vec{v} - \vec{u}\|^2 = \|\vec{v} - p_E(\vec{u})\|^2 + \|p_E(\vec{u}) - \vec{u}\|^2.$$

□

### 3.4 Retour sur la régression linéaire

On rappelle que le problème de la régression linéaire est de trouver la « meilleure droite »  $y = ax + b$ , la plus proche possible, au sens des moindres carrés, des données  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

On a vu que si on note

$$X = (x_1, x_2, \dots, x_n), \quad Y = (y_1, y_2, \dots, y_n) \quad \text{et} \quad V = (1, 1, \dots, 1) \in \mathbb{R}^n,$$

les coefficients  $a$  et  $b$  sont tels que le vecteur  $aX + bV$  est la projection orthogonale de  $Y$  sur le plan  $P$  engendré par  $X$  et  $V$ . On suppose ici que les  $x_i$  ne sont pas tous égaux pour que  $X$  et  $V$  ne soient pas liés. Sinon les données sont sur une droite verticale.

D'après le corollaire 3.6, ces coefficients  $a$  et  $b$  sont les solutions du système linéaire à deux équations

$$\begin{cases} \langle aX + bV, X \rangle = \langle Y, X \rangle \\ \langle aX + bV, V \rangle = \langle Y, V \rangle \end{cases} \Leftrightarrow \begin{cases} a\|X\|^2 + b\langle V, X \rangle = \langle Y, X \rangle \\ a\langle X, V \rangle + b\|V\|^2 = \langle Y, V \rangle \end{cases}$$

avec

$$\|X\|^2 = \sum x_i^2, \quad \langle X, V \rangle = \sum x_i, \quad \|V\|^2 = n, \quad \langle Y, X \rangle = \sum x_i y_i \quad \text{et} \quad \langle Y, V \rangle = \sum y_i.$$

On a donc

$$\begin{aligned} \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} a \\ b \end{pmatrix} &= \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}. \end{aligned}$$

Cela résout le problème! Il n'est pas nécessaire d'apprendre ces formules par coeur car nous verrons que Scilab fait simplement ces calculs.

Nous mentionnons d'abord quelques situations d'ajustement de courbes qui se ramènent à la régression linéaire.

**Variantes de la régression linéaire.** On se donne toujours deux séries de données

$$X = (x_1, x_2, \dots, x_n) \quad \text{et} \quad Y = (y_1, y_2, \dots, y_n).$$

avec  $x_i$  pas tous égaux. Selon le modèle utilisé, on peut chercher d'autres courbes que des droites pour chercher à lier les données  $x_i$  aux  $y_i$ .

- **Loi puissance.** Si on cherche des coefficients  $a$  et  $b$  tels que

$$y_i \simeq a(x_i)^b,$$

alors on peut utiliser une régression linéaire avec les données  $\ln y_i$  et  $\ln x_i$  :

$$\ln y_i \simeq \ln a + b \ln x_i$$

- **Loi exponentielle.** La loi exponentielle intervient souvent en physique. On peut chercher des coefficients  $a$  et  $b$  tels que

$$y_i \simeq ae^{bx_i}$$

à l'aide d'une régression linéaire entre  $\ln y_i$  et  $x_i$  :

$$\ln y_i \simeq \ln a + bx_i.$$

### 3.5 Comment faire les calculs avec Scilab ?

En fait, Scilab résout automatiquement les problèmes de moindres carrés !

Si  $A$  est une matrice de type  $n \times p$  de rang  $p$  et  $Y \in \mathbb{R}^n$  est présenté en colonne, alors la commande

```
-->A\Y // attention au symbole \ utilisé, c'est un 'backslash'
```

renvoie en vecteur  $X$  de  $\mathbb{R}^p$  tel que  $\|AX - Y\|^2$  soit le plus petit possible. En d'autres termes,  $AX$  est la projection orthogonale de  $Y$  sur l'image de  $A$ .

Une variante de la commande, qui fonctionne aussi sur les matrices non injectives, est d'utiliser la fonction `pinv(A)` (pseudo-inverse). On a alors

```
-->pinv(A)*Y
```

renvoie un vecteur  $X$  de  $\mathbb{R}^p$  orthogonal à  $\ker A$  et tel que  $AX$  soit orthogonal à l'image de  $A$ .

Dans le cas de la régression linéaire, on cherche  $a$  et  $b$  tels que  $aX + bV$  soit la projection de  $Y$  sur le plan  $P$  engendré par  $X$  et  $V$ . Il suffit donc d'écrire les données  $X$  et  $Y$  en colonnes (la commande Scilab « ' » échange les colonnes et les lignes d'une matrice), ainsi que le vecteur  $V = (1, 1, \dots, 1)$  en colonne

```
-->V=ones(n,1) // pour avoir une colonne de 1 de hauteur n
```

Alors la matrice  $A$  définie par

```
-->A=[X,V]
```

a deux colonnes  $X$  et  $V$ , et donc  $P$  pour image. Les coefficients  $a$  et  $b$  s'obtiennent avec

```
-->A\Y
```

tandis que la suite de valeurs estimées  $\tilde{y}_i = ax_i + b$  sur la droite de régression est simplement le vecteur  $aX + bV$ , c'est-à-dire

```
-->A*(A\Y)
```

### 3.6 Encore plus de « data fitting » !

La méthode présentée de moindres carrés s'applique tout aussi bien dans des problèmes d'ajustement de courbes, de surfaces, dépendant de plus de paramètres. En voici quelques exemples.

- **Approximation quadratique.** Un modèle, ou l'allure des données  $(x_i, y_i)$ , peut nous conduire à chercher une parabole approchante plutôt qu'une droite. C'est-à-dire que l'on cherche  $a$ ,  $b$  et  $c$  tels que

$$y_i \simeq ax_i^2 + bx_i + c$$

au sens des moindres carrés.

Cela reste un problème linéaire en  $a$ ,  $b$ ,  $c$ , ou l'on projette le vecteur  $Y$  sur le sev engendré par  $X^2 := (x_1^2, x_2^2, \dots, x_n^2)$  (marche en Scilab),  $X$  et  $V$ . Il suffit de nouveau de fabriquer une matrice  $A$  dont les trois colonnes sont  $X^2$ ,  $X$  et  $V$ , puis de taper

-->A\Y

pour avoir les coefficients  $a$ ,  $b$  et  $c$  (en colonne), et

-->A\*ans // ans (answer) reprend le résultat du calcul précédent

pour avoir les points  $\tilde{y}_i = ax_i^2 + bx_i + c$  sur la parabole cherchée.

• **Approximation sinusoïdale.** Beaucoup de problèmes physiques sont périodiques et bien approchés par des lois sinusoïdales du type

$$y_i \simeq a + b \cos(\omega x_i) + c \sin(\omega x_i).$$

Si  $\omega$  est déjà connu, cela s'étudie de nouveau avec la technique des moindres carrés. On crée une matrice  $A$  dont les trois colonnes sont  $V$ ,  $\cos(\omega X)$  et  $\sin(\omega X)$ , et on procède comme ci-dessus.

• **Problèmes à plusieurs variables.** On peut parfois chercher à lier plus de deux variables entre elles. Par exemple, chercher des relations linéaires entre le poids  $P$ , la taille  $T$ , l'âge  $A$  et le sexe ( $S=0$  ou  $1$ ) des individus

$$P \simeq x_1 + x_2 T + x_3 A + x_4 S.$$

C'est encore un problème linéaire dans les inconnues sur lequel la méthode des moindres carrés s'applique ! On crée ici une matrice dont les colonnes sont  $V$  et les différentes données  $T$ ,  $A$ ,  $S$ , et on procède comme ci-dessus.

### 3.7 Comment évaluer la qualité d'une approximation ?

Tout cela est bien joli, mais il faut pouvoir juger de la pertinence des approximations calculées. Par exemple la méthode de régression linéaire trouve toujours une droite, même si les points  $(x_i, y_i)$  sont distribués au hasard et loin d'être alignés !

Il y a un indicateur général et simple à calculer dans la méthode des moindres carrés. On projette ici un vecteur  $Y$  sur un sev  $V$  de  $\mathbb{R}^n$  engendré par une famille d'autres données ( $V$ ,

$X$ , etc). L'approximation de  $Y$  par  $P_V(Y)$  est d'autant meilleure que l'angle  $\theta$  que fait  $Y$  et  $P_V(Y)$  dans  $\mathbb{R}^n$  est faible. Cet angle mesure la **corrélation** entre  $Y$  et  $P_V(Y)$ .

Or cet angle satisfait (grâce à Pythagore, comme dans  $\mathbb{R}^2$  et  $\mathbb{R}^3$ )

$$\cos \theta = \frac{\|P_V(Y)\|}{\|Y\|}.$$

**Définition 3.9.** Le nombre  $r = \cos \theta$  est un nombre compris entre 0 et 1, et s'appelle *coefficient de corrélation* de  $Y$  par rapport à  $V$ . Plus il est proche de 1, plus l'angle de  $Y$  avec  $P_V(Y)$  est faible et l'approximation de bonne qualité.

Avec Scilab, une fois calculé  $P_V(Y)$  à l'aide de

```
-->P=A*(A\Y)
```

on peut calculer facilement  $r^2 = \cos^2 \theta = \frac{\|P_V(Y)\|^2}{\|Y\|^2}$  en faisant

```
-->(P'*P)/(Y'*Y)
```

La valeur seuil dépend des domaines d'étude, mais si

$$r^2 = \cos^2(\theta) \geq 0,9$$

alors l'approximation de  $Y$  par  $P_V(Y)$  est souvent considérée comme très bonne. Moralement, cela signifie que « 90% » des valeurs de  $Y$  « s'expliquent » par celles de  $P_V(Y)$ . Reste « 10% » de  $Y$  orthogonal à  $V$ .

Si cet indicateur est mauvais ( $r^2 < 0,5$  par exemple), il vaut mieux ne pas trop s'appuyer sur cette approximation, et changer de loi modèle. À cet égard, il faut reconnaître que l'oeil est un outil très efficace pour deviner des lois pertinentes ! Il ne faut donc pas hésiter par commencer par représenter les différentes données, avec un « `plot` » of course.

I wish you an Happy Data Fitting with Scilab !

---

email : [michel.rumin@math.u-psud.fr](mailto:michel.rumin@math.u-psud.fr)

page web : <http://www.math.u-psud.fr/~rumin/enseignement/enseignement.html>