

Agrégation 2002-2003

Diagonalisation des auto-adjoints

**Applications aux formes quadratiques : Directions principales
Applications en Géométrie, en Statistique et en Mécanique**

Les paragraphes 1, 2, 3 donnent un bref résumé de la théorie, accompagné de références et de la mise en évidence des principales difficultés de ce chapitre. Dans les paragraphes suivants, on trouvera quelques applications de la diagonalisation des auto-adjoints, pouvant servir d'exposé dans plusieurs leçons d'algèbre, de géométrie, de statistique ou de calcul scientifique en particulier dans :

- Matrices semblables ;
- Réduction d'un endomorphisme ;
- Sous-espaces stables ; (diagonalisation des endomorphismes normaux)
(diagonalisation des auto-adjoints)
- Formes quadratiques ;
- Quadriques ; (quand ce titre existe)
- Endomorphismes remarquables d'un espace euclidien ;
- Angles et distances (entre variables et entre individus en statistiques) ;
- $O(2, \mathbb{R})$, $O(3, \mathbb{R})$; (diagonalisation par blocs des endomorphismes normaux sur \mathbb{R})
- Endomorphismes remarquables d'un espace hermitien ;
- Coniques ;
- Courbes et surfaces ; (point de vue métrique)
- Isométries ;
- Endomorphismes diagonalisables ;
- Exemples de décompositions remarquables dans $GL_p(\mathbb{R})$. (Décomposition polaire)

**1 Diagonalisation
des endomorphismes auto-adjoints,
en dimension finie**

Pour une définition intrinsèque de l'adjoint d'un endomorphisme, il est intéressant d'utiliser le dual. (Voir [RDO] page 3, page 7, puis, page 34)

Etant donné un espace muni d'une forme bilinéaire symétrique φ , on a une application linéaire naturelle $d_\varphi : y \longmapsto (d_\varphi(y) : x \mapsto \varphi(x, y))$ (d comme droite) de E dans le dual E^* . La matrice de cette application linéaire dans une base (e_i) et dans la base, associée, (e_i^*) , du dual n'est rien d'autre que la matrice de la forme bilinéaire symétrique φ dans la base (e_i) . Le vocabulaire est cohérent, puisque le

noyau de d_φ est bien ce que l'on appelle le noyau de φ en algèbre bilinéaire.

Si maintenant on a un espace euclidien E , muni de la forme bilinéaire symétrique définie positive φ , comme le noyau de φ est réduit à $\{0\}$, d_φ est injective et est donc un isomorphisme en dimension finie.

Cet isomorphisme permet de définir l'adjoint pour φ , d'une application linéaire u de E dans E à partir de la transposée u^t de E^* dans E^* : $u^* = d_\varphi^{-1} \circ u^t \circ d_\varphi$. (Voir [RDO] page 34). Autrement dit, u^* rend le diagramme suivant commutatif.

$$\begin{array}{ccc} E & \xrightarrow{u^*} & E \\ d_\varphi \downarrow & & \downarrow d_\varphi \\ E^* & \xrightarrow{u^t} & E^* \end{array}$$

On vérifiera, facilement, que cette définition équivaut à celle présentée dans les manuels de premier cycle :

$$\forall x, \forall y, \varphi(x, u^*(y)) = \varphi(u(x), y)$$

On peut donner une définition analogue dans le cas hermitien, mais il faut faire attention à l'endroit où l'on place la conjugaison. On a le choix entre antilinéarité de d_φ ou structure conjuguée du dual. Ce n'est pas présenté de la même manière selon les livres. Il vaut mieux s'en tenir à un seul livre, de préférence récent !

On trouve juste après, dans [RDO] page 35, un résultat qui sert pour l'orthogonalisation simultanée de deux formes quadratiques. C'est la correspondance entre les formes bilinéaires symétriques ψ sur E et les endomorphismes u , φ -symétriques, c'est à dire φ -auto-adjoints, de E .

$$\begin{array}{ccc} \psi & \longmapsto & d_\varphi^{-1} \circ d_\psi \\ ((x, y) \mapsto \varphi(x, u(y))) & \longleftarrow & u \end{array}$$

Il y a deux points de vue pour prouver que les endomorphismes auto-adjoints sont diagonalisables dans une base orthonormée, soit directement (par exemple dans [Gob]), soit en recherchant les extrema de $\varphi(x, u(x))$ sur la sphère unité $\varphi(x, x) = 1$ (par exemple dans [Gour]). Nous allons voir, plus loin, le lien entre les deux approches.

Pour un point de vue matriciel, voir le livre, très synthétique, [Ser] page 29, qui commence par la diagonalisation des matrices normales sur \mathbb{C} dans une base orthonormée (avec matrice de passage unitaire). Puis, il caractérise les différents exemples par leurs valeurs propres.

- Hermitiens, valeurs propres réelles.
- Unitaires, valeurs propres de module 1.
- Anti-hermitiens, valeurs propres imaginaires pures.

Il en déduit la diagonalisation par blocs dans le cas réel avec une matrice de passage orthogonale.

- Symétriques, diagonalisables à valeurs propres réelles.

- Orthogonaux, bloc de taille 1 avec +1 ou -1 ou bloc de taille 2, matrice de rotation.
- Anti-symétriques, bloc de taille 2 (voir [Grif] page 341 exercice 3).

Attention : Il faut que la base soit orthonormée pour que la matrice de u^* soit la transposée de celle de u . Sinon la matrice de d_φ , c'est à dire celle de φ , et son inverse interviennent dans le calcul. $M(u^*) = M(\varphi)^{-1} \cdot M(u)^t \cdot M(\varphi)$.

Les deux auto-adjoints associés à une application linéaire

Soient (E, φ) et (F, ψ) deux espaces euclidiens et $f : E \rightarrow F$ une application linéaire. $f^* : F \rightarrow E$ se définit de la même manière que précédemment : $f^* = d_\varphi^{-1} \circ f^t \circ d_\psi$. Autrement dit, f^* rend le diagramme suivant commutatif.

$$\begin{array}{ccc} F & \xrightarrow{f^*} & E \\ d_\psi \downarrow & & \downarrow d_\varphi \\ F^* & \xrightarrow{f^t} & E^* \end{array}$$

Cette définition équivaut à légalité

$$\forall x \in F, \forall y \in E, \varphi(x, f^*(y)) = \psi(f(x), y)$$

On vérifie aisément que $u = f^* \circ f$ est φ -autoadjoint et que $v = f \circ f^*$ est ψ -autoadjoint. On utilise souvent, dans la pratique, l'un ou l'autre de ces deux endomorphismes auto-adjoints associés à f .

Ceci est utilisé par exemple dans la décomposition polaire avec $E = F$. ([Grif] page 300 exercice 25) ou dans l'analyse en composantes principales, en statistique (voir 8).

Une remarque est intéressante dans cette situation : Les valeurs propres non nulles de u et v sont les mêmes avec la même multiplicité. (Voir [LMP] page 20).

En effet, si λ est non nul, alors $f \circ f^*$ est un automorphisme de $\text{Ker}(f \circ f^* - \lambda I)$ et $f^* \circ f$ est un automorphisme de $\text{Ker}(f^* \circ f - \lambda I)$.

De plus, il est facile de vérifier que

$$f(\text{Ker}(f^* \circ f - \lambda I)) \subset \text{Ker}(f \circ f^* - \lambda I) \text{ et}$$

$$f^*(\text{Ker}(f \circ f^* - \lambda I)) \subset \text{Ker}(f^* \circ f - \lambda I).$$

On en déduit que f et f^* induisent des isomorphismes entre $\text{Ker}(f^* \circ f - \lambda I)$ et $\text{Ker}(f \circ f^* - \lambda I)$.

Donc si λ est valeur propre non nulle de $u = f^* \circ f$, elle est également valeur propre de $v = f \circ f^*$, avec la même multiplicité.

D'un point de vue matriciel, dans des bases orthonormées, pour une matrice rectangulaire X , dans $M_{n,p}(\mathbb{R})$, avec $p < n$, on choisit de diagonaliser la matrice carrée symétrique de plus petite taille, $X^t \cdot X$, obtenue par ce procédé. De plus, si $(V_i)_{i \in I}$ est une base de vecteurs propres pour cette matrice, associée aux différentes valeurs propres non nulles, $(XV_i)_{i \in I}$ sera une base de vecteurs propres pour $X \cdot X^t$ associée aux mêmes valeurs propres non nulles. Cette remarque est utilisée en statistique, pour l'analyse des données. Voir 8.

Petits rappels sur les changements de base et les matrices de passage.

Soit (e_i) et (e'_i) deux bases de \mathbb{R}^p . P est la matrice de Id dans les bases $\mathbb{R}^p \xrightarrow{Id} \mathbb{R}^p$.
 $(e'_i) \quad (e_i)$

P contient, en colonne, les coordonnées des vecteurs de la nouvelle base (e'_i) exprimées dans l'ancienne base (e_i) .

Regardons l'effet sur les coordonnées.

si $\vec{V} = \sum_{i=1}^p x_i e_i = \sum_{i=1}^p x'_i e'_i$, $Id(\vec{V}) = \vec{V}$ se traduit par

$$P \begin{pmatrix} x'_1 \\ \vdots \\ x'_p \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

Donc les nouvelles coordonnées s'obtiennent en fonction des anciennes en lisant les lignes de P^{-1} . Autrement dit, le changement de base de e_i^* à e'_i^* a pour matrice $(P^{-1})^t$

Utilité d'avoir une matrice de passage orthogonale

Le fait, dans le cas réel, que la matrice de passage P soit orthogonale présente plusieurs intérêts.

- On peut interpréter le changement de base dans deux domaines totalement différents : les endomorphismes linéaires d'une part et les formes bilinéaires symétriques ou quadratiques d'autre part.
- L'application correspondant à P dans une base fixée, orthonormale, est une isométrie. C'est utile pour distinguer une ellipse d'un cercle, par exemple.
- Le changement de base sur les formes coordonnées dans le dual, pour passer de la base e_i^* à la base e'_i^* a pour matrice la transposée de P^{-1} , mais ici, $(P^{-1})^t = P$. On a directement, à partir du "changement de variable", les colonnes de la matrice de passage dans \mathbb{R}^p de la base (e_i) à la base (e'_i) . Cette remarque servira pour l'analyse en composantes principales, où le changement de base se fera sur les coordonnées, dans le dual.

2 Orthogonalisation ("diagonalisation") simultanée

On se donne deux formes quadratiques q_1 et q_2 sur \mathbb{R}^p , associées respectivement à deux formes bilinéaires symétriques φ_1 et φ_2 et à deux matrices symétriques M_1 et M_2 dans la base canonique de \mathbb{R}^p . On suppose φ_1 définie positive, munissant \mathbb{R}^p d'une structure euclidienne.

Théorème : *Il existe une base orthonormée pour φ_1 , orthogonale pour φ_2 .*

ingrédients de la preuve :

Considérons l'endomorphisme $u = d_{\varphi_1}^{-1} \circ d_{\varphi_2}$ associée à la matrice $M_1^{-1} M_2$.

u est φ_1 -symétrique, c'est à dire, φ_1 -auto-adjoint. La base recherchée est une base orthonormée de vecteurs propres pour u , obtenue en utilisant, à l'intérieur de

chaque sous-espace propre, un procédé d'orthonormalisation (Gauss ou Gram-Schmidt). Voir [RDO] ou [LFA1] page 417.

Pour la recherche des valeurs propres et vecteurs propres de u , on peut remarquer que $M_1^{-1}M_2(V) = \lambda V$ équivaut à $(M_2 - \lambda M_1)V = 0$. Ceci évite d'inverser M_1 et les logiciels proposent, dans leur menu, ce genre de calcul. Voir `eigenvalues(A,C)`, en Maple.

En diagonalisant u on peut obtenir une base de vecteurs propres orthonormée pour φ_1 et orthogonale pour φ_2 . La matrice de passage P est φ_1 -orthogonale mais elle n'est pas orthogonale pour le produit scalaire usuel. $P^t \neq P^{-1}$.

La traduction matricielle du théorème est que dans la base de vecteurs propres la matrice de φ_1 est Id et celle de φ_2 est diagonale. On peut ranger les vecteurs propres de telle sorte qu'ils correspondent aux valeurs propres rangées dans un ordre décroissant.

Attention aux démonstrations "rapides" :

M_2 bien qu'étant une matrice symétrique, ne définit pas un endomorphisme auto-adjoint pour la structure euclidienne définie par φ_1 .

Dans beaucoup de livres ([Gou] page 241, [Gob] page 301) cette démonstration peut entraîner des difficultés au moment de la mise en oeuvre concrète du théorème. En effet, soit on se place d'emblée dans une base déjà orthonormée pour φ_1 , soit on commence par modifier la forme quadratique par la méthode de Gauss-Gram-Schmidt pour obtenir une base orthonormée pour φ_1 . Ce premier changement de base est uniquement dans le domaine de l'algèbre bilinéaire et la matrice de passage P_1 est "quelconque" : $P_1^t \neq P_1^{-1}$. On ne considère aucun endomorphisme auto-adjoint dans cette première partie. On utilise seulement la méthode de Gauss-Gram Schmidt.

Après cette étape on revient dans le chapitre de la diagonalisation des endomorphismes auto-adjoints. La matrice de φ_1 est Id et celle de φ_2 est symétrique. Cette matrice symétrique est la matrice de $u = d_{\varphi_1}^{-1} \circ d_{\varphi_2}$ et on la diagonalise dans une base orthonormée.

Un exemple simple des deux itinéraires possibles de calcul, avec des matrices 2x2, permet de comprendre les deux manières de présenter la démonstration.

Considérez dans \mathbb{R}^2 , deux formes quadratiques définies par leur matrice respective dans la base canonique. $M(q_1) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ $M(q_2) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Vérifier que la forme q_1 est définie positive et déterminez en suivant les différentes démonstrations des livres, une base orthonormale pour q_1 et orthogonale pour q_2 . Les différentes matrices de passage utilisées sont-elles orthogonales ?

Voir aussi le problème de Math généré 2001, où cette question d'orthogonalisation simultanée intervient.

3 Directions principales, Extrema du rapport de 2 formes quadratiques

Soient q_1 une forme quadratique définie positive et q_2 une forme quadratique sur \mathbb{R}^p . On notera φ_1 et φ_2 , les formes bilinéaires symétriques correspondantes.

Considérons le rapport $\frac{q_2(x)}{q_1(x)}$. Ce rapport est constant dans une direction donnée puisque $\frac{q_2(\lambda x)}{q_1(\lambda x)} = \frac{\lambda^2 q_2(x)}{\lambda^2 q_1(x)} = \frac{q_2(x)}{q_1(x)}$.

Soient $\lambda_1 \cdots \lambda_p$, les valeurs propres, supposées distinctes, rangées dans l'ordre décroissant, de l'endomorphisme auto-adjoint $u = d_{\varphi_1}^{-1} \circ d_{\varphi_2}$, associé à q_1 et q_2 . Dans la pratique, pour des formes prises au hasard, on est presque sur, au sens des probabilités, de tomber sur des valeurs propres distinctes. Cette hypothèse est donc peu restrictive.

Soit $\{V_1 \dots V_p\}$ une base de vecteurs propres de u , orthonormée pour φ_1 . Alors si $x = \sum x_i V_i$

$$\lambda_p = \frac{\lambda_p \sum x_i^2}{\sum x_i^2} \leq \frac{q_2(x)}{q_1(x)} = \frac{\sum \lambda_i x_i^2}{\sum x_i^2} \leq \frac{\lambda_1 \sum x_i^2}{\sum x_i^2} = \lambda_1$$

Ainsi le rapport $\frac{q_2}{q_1}$ est maximum dans la direction de V_1 et minimum dans la direction de V_p

- le maximum étant $\lambda_1 = \frac{q_2(V_1)}{q_1(V_1)}$
- le minimum étant $\lambda_p = \frac{q_2(V_p)}{q_1(V_p)}$

Le rapport $\frac{q_2}{q_1}$ étant constant sur $Vect(V) - \{0\}$ on peut le calculer pour des

vecteurs normés pour q_1 . On s'intéresse alors à $\sup_{V \mid q_1(V)=1} q_2(V) = \sup_{V \neq 0} \frac{q_2(V)}{q_1(V)}$

On peut considérer, pour rendre plus fine l'analyse, des extrema intermédiaires.

$F_1 = Vect(V_1)$ réalise le maximum λ_1 de $\frac{q_2}{q_1}$ sur \mathbb{R}^p

Ensuite on s'intéresse à l'orthogonal, pour φ_1 et φ_2 , de $F_1 : F_1^\perp = Vect(V_2 \cdots V_p)$
 $F_2 = Vect(V_2)$ réalise le maximum λ_2 de $\frac{q_2}{q_1}$ sur F_1^\perp

Puis on considère $(F_1 \oplus F_2)^\perp$

$F_3 = Vect(V_3)$ réalise le maximum λ_3 de $\frac{q_2}{q_1}$ sur $(F_1 \oplus F_2)^\perp$ et ainsi de suite.

Lorsque les valeurs propres sont distinctes, on obtient ainsi ce que l'on appelle les **directions principales**.

Dans le cas où q_2 est également définie positive, on peut intervertir les rôles. Comme $d_{\varphi_2}^{-1} \circ d_{\varphi_1}$ et $d_{\varphi_1}^{-1} \circ d_{\varphi_2}$ ont les valeurs propres inverses l'un de l'autre avec les mêmes vecteurs propres, les directions principales sont les mêmes, orthogonales pour φ_1 et pour φ_2 .

NB : On peut aussi inverser le cours et utiliser des arguments d'extrema pour diagonaliser u . (Voir [Gour] page 240 ou [LMP] 1.2.3 page 34 ou [Sap] 8.2.2).

4 Directions principales d'une quadrique, directions principales d'un ellipsoïde

Pour une quadrique quelconque définie par une forme quadratique q_2 sur \mathbb{R}^p et l'équation $q_2(x) = 1$, les directions principales seront par définition celles qui correspondent aux extrema successifs du rapport $\frac{q_2(x)}{q_1(x)}$, où $q_1(x) = \sum x_i^2$ est la forme égale à la norme usuelle au carré. Dans le cas d'un ellipsoïde centré, les deux formes étudiées étant définies positives, on peut considérer les directions principales correspondant au rapport inverse $\frac{q_1(x)}{q_2(x)}$. Ce sont les mêmes directions, les extrema étant rangés dans l'ordre inverse.

Exemple : l'ellipse $x^2 + 2xy + 3y^2 = 1$ dans \mathbb{R}^2 .

$$M(q_1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad M(q_2) = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$$

Calculons les valeurs propres λ_1 et λ_2 de $M(q_1)^{-1}M(q_2) = M(q_2)$. Le polynôme caractéristique de $M(q_2)$: $\lambda^2 - 4\lambda + 2$ a pour racines : $2 \pm \sqrt{2}$.

Les deux extrema de la norme usuelle quand on parcourt l'ellipse, extrema du rapport inverse $\frac{q_1(x)}{q_2(x)}$ sont donc :

$$\text{longueur du demi grand axe : } a = \sqrt{\frac{1}{\lambda_1}} = \sqrt{\frac{1}{2 - \sqrt{2}}}$$

$$\text{longueur du demi petit axe : } b = \sqrt{\frac{1}{\lambda_2}} = \sqrt{\frac{1}{2 + \sqrt{2}}}$$

Les vecteurs propres correspondants donnent les directions du grand axe et du petit axe, orthogonales pour φ_1 et pour φ_2 .

On peut aussi obtenir les directions des axes d'une conique sans diagonaliser, par un changement de coordonnées, en faisant tourner les axes initiaux d'un angle θ .

$$\begin{aligned} x &= \cos \theta x' - \sin \theta y' \\ y &= \sin \theta x' + \cos \theta y' \end{aligned}$$

On choisit θ pour obtenir une équation réduite de la conique dans une base orthonormée pour le produit scalaire, orthogonale pour q_2 . Pour cela on élimine les termes en $x'y'$. Pour cet exemple on trouve $\theta = \frac{3\pi}{8} \text{ mod } \frac{\pi}{2}$.

5 Rayons conjugués d'un ellipsoïde

Les notations pour ce paragraphe sont celles de la référence : [Gob] page 310, pour faciliter la lecture.

Dans \mathbb{R}^3 , soit Q_0 la forme quadratique associée au produit scalaire usuel. C'est la norme usuelle au carré. Soit q un ellipsoïde défini par l'équation $Q(\overrightarrow{OM}) = 1$, où Q est une forme définie positive.

Un système de rayons conjugués est une base $(\overrightarrow{OA_1}, \overrightarrow{OA_2}, \overrightarrow{OA_3})$, orthonormée pour Q . Pour trouver des rayons conjugués, il suffit d'appliquer l'algorithme de Gauss-Gram-Schmidt pour Q , à une base quelconque. On en obtient une infinité. Nous allons démontrer que le volume du parallélépipède engendré par les trois rayons conjugués et la quantité $OA_1^2 + OA_2^2 + OA_3^2$ sont des constantes, indépendantes du système de rayons conjugués. C'est un théorème d'Appolonius. (Voir [Ber] 15.6)

Désignons par φ et φ_0 les formes bilinéaires symétriques associées à Q et Q_0 et d_φ et d_{φ_0} les applications linéaires correspondantes, de matrices M et $M_0 = Id$, respectivement, dans une base usuelle orthonormée pour Q_0 .

Enfin posons $u = d_\varphi^{-1} \circ d_{\varphi_0}$, de matrice $M^{-1}M_0 = M^{-1}$.

Le polynôme caractéristique de u et donc, tous ses coefficients, sont invariants par changement de base, en particulier la trace et le déterminant de u .

On remarque que u^{-1} a pour matrice M , dans la base usuelle. Les valeurs propres de u sont donc les inverses de celles de M :

$$\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \frac{1}{\lambda_3}$$

Dans une base de vecteurs propres de u , Q -orthonormée et Q_0 -orthogonale, la matrice de Q est donc Id et celle de Q_0 est diagonale.

Quand on passe de cette base à une base de rayons conjugués Q -orthonormée, $(\overrightarrow{OA_1}, \overrightarrow{OA_2}, \overrightarrow{OA_3})$, la matrice de Q devient $P^t \cdot Id \cdot P = Id$. La matrice de passage P est donc orthogonale. Le changement de base se fait pour les formes bilinéaires avec P^t , mais aussi pour l'application linéaire u puisque $P^{-1} = P^t$.

La matrice M'_0 de la forme bilinéaire Q_0 dans cette nouvelle base est $\text{Gram}(\overrightarrow{OA_1}, \overrightarrow{OA_2}, \overrightarrow{OA_3})$. Celle de l'application linéaire u est aussi $M'_0 = Id^{-1}M'_0$. La diagonale de cette matrice est OA_1^2, OA_2^2, OA_3^2 et donc, $OA_1^2 + OA_2^2 + OA_3^2$ est la trace de u , indépendante de la base choisie, c'est à dire du système de rayons conjugués. C'est

$$\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3}$$

Si l'on considère le déterminant de u , déterminant de $\text{Gram}(\overrightarrow{OA_1}, \overrightarrow{OA_2}, \overrightarrow{OA_3})$, il est égal à $\frac{1}{\lambda_1 \lambda_2 \lambda_3}$, mais aussi au carré du volume du parallélépipède engendré par les trois rayons conjugués. Celui-ci est donc indépendant du choix de système de rayons conjugués. (Voir [Ber] 15.6)

Goblot fait une démonstration purement matricielle de ce résultat. Il est possible de s'en contenter. Mais il est intéressant de voir, comme ici, avec les directions principales et les rayons conjugués d'un ellipsoïde, les deux applications possibles de l'orthogonalisation simultanée quand les formes sont toutes deux définies positives et de voir que les matrices que l'on diagonalise sont inverses l'une de l'autre. Leurs valeurs propres sont inverses les unes des autres et leurs vecteurs propres coïncident.

6 Cercles sur les quadriques

Les notations pour ce paragraphe sont celles de la référence : [Gob] page 308, pour faciliter la lecture.

Soit Q_o la forme quadratique associée au produit scalaire et Q la forme quadratique associée à une quadrique d'équation $Q = 1$

On cherche des plans qui rencontrent la quadrique selon un cercle de rayon r . Les vecteurs V de ces plans sont tels que si $Q(V) = 1$ alors $Q_o(V) = r$. Ceci nous ramène à chercher V tel que $(Q_o - rQ)(V) = 0$, c'est à dire, des plans contenus dans le cône isotrope de $Q_o - rQ$. L'existence de tels plans exige que le rang de $Q_o - rQ$ soit inférieur ou égal à 2.

Soit M la matrice symétrique correspondant à Q . Pour résoudre ce problème, choisissons une base orthogonale simultanément pour Q et Q_o , orthonormée pour Q_o . Soit V_1, V_2, V_3 des vecteurs propres de M , orthonormée pour Q_o , associés respectivement aux valeurs propres : $\lambda_1 \geq \lambda_2 \geq \lambda_3$.

Dans la base V_1, V_2, V_3 les matrices de Q_o et Q deviennent :

$$M'_0 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \text{ et } M' = P^t.M.P = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}$$

et l'équation $(Q_o - rQ)(V) = 0$ où $V = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$, devient

$$(1 - \lambda_1 r)X^2 + (1 - \lambda_2 r)Y^2 + (1 - \lambda_3 r)Z^2 = 0$$

avec $1 - \lambda_1 r \leq 1 - \lambda_2 r \leq 1 - \lambda_3 r$.

Comme le rang de cette forme doit être strictement plus petit que 3, nécessairement l'un des $1 - \lambda_i r$ est nul. De plus, pour avoir un plan de vecteurs isotropes, il nous faut, soit " des + et des -", soit un plan contenu dans le noyau. Les seuls signes possibles pour $1 - \lambda_1 r, 1 - \lambda_2 r, 1 - \lambda_3 r$ sont alors :

- $(0, 0, +)$, qui nous donne le plan $Z = 0$, et un cercle de rayon $r = \frac{1}{\lambda_1} = \frac{1}{\lambda_2}$, dans le cas où $\lambda_1 = \lambda_2$ sont non nuls ;
- $(-, 0, 0)$, qui nous donne le plan $X = 0$, et un cercle de rayon $r = \frac{1}{\lambda_2} = \frac{1}{\lambda_3}$, dans le cas où $\lambda_2 = \lambda_3$ sont non nuls,
- $(-, 0, +)$ correspondant à $r = \frac{1}{\lambda_2}$, qui nous donne deux plans, donc deux cercles, à condition toutefois que $\lambda_2 \neq 0$. Dans ce cas,

$$Q_o - rQ = \left(1 - \frac{\lambda_3}{\lambda_2}\right) Z^2 - \left(\frac{\lambda_1}{\lambda_2} - 1\right) X^2.$$

On pose $\alpha^2 = \lambda_2 - \lambda_3 \geq 0$ et $\beta^2 = \lambda_1 - \lambda_2 \geq 0$. L'équation $\lambda_2(Q_o - rQ) = 0$

se factorise en donnant deux équations de plans

$$(\alpha Z + \beta X) = 0, \quad (\alpha Z - \beta X) = 0$$

Ces 2 plans vectoriels ne peuvent coïncider car $\beta \neq 0$, puisque $\lambda_1 \neq \lambda_2$.

Cas particulier : Pour $\lambda_i = 0$, on ne peut obtenir $\lambda_i r = 1$. De ce fait, si $\lambda_2 = 0$, il n'y a pas de solution .

7 Directions et courbures principales d'une surface

Voir [Aud] et [LFA3] et le cours 2002 de M.Bauer.

Soit une surface de \mathbb{R}^3 , régulière en M , définie par un paramétrage :

$$f : (u, v) \rightarrow f(u, v)$$

On a alors une base "naturelle" du plan tangent $T_M : \left\{ \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\}$. [Aud] page 250.

On définit sur le plan tangent T_M deux formes quadratiques. q_1 est la première forme fondamentale, celle correspondant au produit scalaire de \mathbb{R}^3 , restreint à T_M , espace tangent en M .

La matrice de q_1 dans la base de T_M est

$$M_1 = \begin{pmatrix} \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \right\rangle & \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\rangle \\ \left\langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial u} \right\rangle & \left\langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right\rangle \end{pmatrix} \quad \text{[LFA3] page 476}$$

La deuxième forme fondamentale, q_2 , (notée aussi II comme deux) "mesure" le changement de direction du vecteur unitaire normal N par rapport au plan tangent.

$\Pi_M(X, Y) = \langle -T_p n(X), Y \rangle$ (notation n au lieu de N , [Aud] page 258. Ce livre ne donne pas les calculs matriciels qui figurent dans [LFA3]).

Si l'on pose $N = \frac{\left(\frac{\partial f}{\partial u} \wedge \frac{\partial f}{\partial v} \right)}{\left\| \frac{\partial f}{\partial u} \wedge \frac{\partial f}{\partial v} \right\|}$, la matrice de II dans la base $\left\{ \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\}$ est :

$$M_2 = \begin{pmatrix} \left\langle N, \frac{\partial^2 f}{\partial u^2} \right\rangle & \left\langle N, \frac{\partial^2 f}{\partial u \partial v} \right\rangle \\ \left\langle N, \frac{\partial^2 f}{\partial u \partial v} \right\rangle & \left\langle N, \frac{\partial^2 f}{\partial v^2} \right\rangle \end{pmatrix}$$

En diagonalisant $M_1^{-1} M_2$, si les valeurs propres sont distinctes, on obtient les directions principales. Si $\lambda_1 > \lambda_2$ sont les valeurs propres et V_1, V_2 les vecteurs

propres correspondants, V_1 correspond à la direction où la variation du vecteur normal est maximale. On obtient la courbure maximale, λ_1 , de la surface en M . C'est la courbure, en M , de la coupe, dans \mathbb{R}^3 , de la surface par le plan $Vect(N, V_1)$. La direction de V_2 donne l'autre courbure principale λ_2 dans le plan $Vect(N, V_2)$. Le produit, $\lambda_1\lambda_2$, s'appelle la courbure de Gauss. C'est le déterminant de la forme q_2 , invariant par un changement de base isométrique pour q_1 , c'est à dire, pour la distance usuelle de \mathbb{R}^3 .

8 En statistique, analyse en composantes principales

Pour comprendre puis, démontrer les résultats de ce chapitre, voir dans l'ordre : [EscPag] chapitre 9, [LMP] Chapitre 1, section 1.2 (et 1.1), [Sap] chapitre 8. On dispose de n individus chacun de poids p_i ($\sum_{i=1}^n p_i = 1$) sur lesquels on peut observer p variables. L'individu i donne $(x_{i,1} \cdots x_{i,p})$. On en déduit :

$$\begin{aligned} \bar{x}_j &= \sum_i p_i x_{i,j}, && \text{la moyenne empirique pondérée;} \\ \sigma_j^2 &= \sum_i p_i (x_{i,j} - \bar{x}_j)^2, && \text{la variance empirique pondérée;} \\ c_{jj'} &= \sum_i p_i (x_{i,j} - \bar{x}_j)(x_{i,j'} - \bar{x}_{j'}), && \text{la covariance entre les variables } j \text{ et } j'; \\ r_{jj'} &= \sum_i p_i \frac{(x_{i,j} - \bar{x}_j)}{\sigma_j} \frac{(x_{i,j'} - \bar{x}_{j'})}{\sigma_{j'}}, && \text{la corrélation entre les variables } j \text{ et } j'. \end{aligned}$$

Pour simplifier, dans [LMP], on suppose $p_i = \frac{1}{n}$. Tous les individus ont le même poids. Pour que ces données soient linéaires, on commence par les centrer. Ces observations centrées sont enregistrées dans la matrice

$$X_0 = \begin{pmatrix} x_{1,1} - \bar{x}_1 & \cdots & x_{1,p} - \bar{x}_p \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_{i,1} - \bar{x}_1 & x_{i,j} - \bar{x}_j & x_{i,p} - \bar{x}_p \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_{n,1} - \bar{x}_1 & \cdots & x_{n,p} - \bar{x}_p \end{pmatrix}$$

Cette matrice définit une application linéaire $f : \mathbb{R}^p \rightarrow \mathbb{R}^n$ où \mathbb{R}^p est l'espace des variables, coordonnées des individus, et \mathbb{R}^n , l'espace des individus dans lequel on dessinera les vecteurs "variables". La i -ème des n formes coordonnées de ces vecteurs correspond à l'individu i . De manière naturelle on choisit de munir ces variables, formes coordonnées sur \mathbb{R}^p d'un poids correspondant à la distance

”écart type”, donné par la forme quadratique q_1 , associée à la matrice

$$\begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{pmatrix}$$

De même on munit la forme coordonnée de l'individu i dans \mathbb{R}^n d'un poids correspondant au poids relatif attribué à cet individu sondé, dans la population considérée. Ce poids est donné par la forme quadratique q'_1 , associée à la matrice

$$\begin{pmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{pmatrix}$$

Toute la suite dépend de ces deux choix. Si l'on change la norme sur les variables ou si l'on change la pondération des individus, le résultat final de l'analyse des données sera totalement différent.

Les premiers changements de base par la méthode de Gauss permettent d'une part, de donner le même poids à toutes les variables en les normant et d'autre part de tenir compte du poids de chaque individu .

On s'intéresse alors aux variables centrées réduites. La matrice de passage pour les variables est

$$P_1 = \begin{pmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_p \end{pmatrix}$$

Dans la nouvelle base de variables, base du dual, $M(q_1) = I$. P_1 est non orthogonale, c'est à dire, $P_1^t \neq P_1^{-1}$

Pour les individus, la matrice de passage est

$$Q_1 = \begin{pmatrix} 1/\sqrt{p_1} & & \\ & \ddots & \\ & & 1/\sqrt{p_n} \end{pmatrix}$$

P_1 et Q_1 ne sont pas orthogonales, c'est à dire, $P_1^t \neq P_1^{-1}$ et $Q_1^t \neq Q_1^{-1}$

(Se ramener à des bases orthonormées, c'est la première étape, de la méthode du Gourdon ou du Goblots). On peut résumer les changements de coordonnées, changements de base sur les espaces duaux, par le diagramme.

$$\begin{array}{ccc} \mathbb{R}^p & \xrightarrow{f} & \mathbb{R}^n \\ & X & \\ Id \downarrow P_1 & & Q_1 \downarrow Id \\ \mathbb{R}^p & \xrightarrow[X_0]{f} & \mathbb{R}^n \end{array}$$

où $X = Q_1^{-1} X_0 P_1$

$$X = \begin{pmatrix} \frac{\sqrt{p_1}(x_{1,1} - \bar{x}_1)}{\sigma_1} & \dots & \frac{\sqrt{p_1}(x_{1,p} - \bar{x}_p)}{\sigma_1} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \frac{\sqrt{p_i}(x_{i,1} - \bar{x}_1)}{\sigma_1} & \frac{\sqrt{p_i}(x_{i,j} - \bar{x}_j)}{\sigma_j} & \frac{\sqrt{p_i}(x_{i,p} - \bar{x}_p)}{\sigma_p} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \frac{\sqrt{p_n}(x_{n,1} - \bar{x}_1)}{\sigma_1} & \dots & \frac{\sqrt{p_n}(x_{n,p} - \bar{x}_p)}{\sigma_p} \end{pmatrix} \quad \text{avec } p \ll n.$$

C'est la nouvelle matrice choisie pour représenter les observations centrées réduites sur les n individus, dont la pondération a été "normée". Voir [LMP] page 36.

Considérons alors l'automorphisme q_1 -auto-adjoint $f^* \circ f$. Sa matrice dans les bases orthonormées est $X^t.X$. C'est la matrice carrée $p \times p$, de covariance empirique des variables centrées réduites, matrice de corrélation empirique, qui définit une forme quadratique q_2 sur \mathbb{R}^p , "mesurant" la covariance entre les variables.

$$M(q_2) = X^t.X$$

C'est dans \mathbb{R}^p , dont le dual est muni d'une base orthonormée pour la distance, associée à q_1 choisie sur les variables, que l'on va représenter les individus. Plus précisément, chaque ligne de la matrice X représente dans \mathbb{R}^p un individu. Chaque colonne de la matrice X est une variable, une forme coordonnée. La distance entre les individus, dont parle [LMP] page 25, est, en fait, définie ici au moyen de q_1 , sur les coordonnées, dans le dual. La matrice M de [LMP] définissant la distance entre les individus est donc $M = M(q_1)^{-1}$.

Remarquons que trouver la direction maximisant le rapport $\frac{q_2}{q_1}$ équivaut à trouver la direction de projection dans l'espace des variables \mathbb{R}^p selon laquelle le nuage des individus sera le plus "étalé" possible, aura la plus grande inertie. En effet, si on représente l'individu i par un point M_i de coordonnées la i -ème ligne de X et si H_i est sa projection dans \mathbb{R}^p dans la direction de u , vecteur unitaire, les n composantes de Xu donnent les mesures algébriques des vecteurs $\overrightarrow{OH_i}$ sur la droite orientée par u . Ainsi, on a

$$q_2(u) = u^t X^t.X u = \sum_i OH_i^2$$

La deuxième étape consiste à diagonaliser la matrice de q_2 , matrice de corrélation $X^t.X$ dans une base déjà orthonormée pour q_1 . La seconde matrice de passage P_2 est telle que $P_2^t = P_2^{-1}$. C'est une isométrie de \mathbb{R}^p qui ne change pas la distance entre les individus.

$$P_2^t.I.P_2 = P_2^{-1}.I.P_2 = I$$

et

$$P_2^t . X^t . X . P_2 = P_2^{-1} . X^t . X . P_2 = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}, \text{ avec } \lambda_1 \geq \lambda_2 \cdots \geq \lambda_p.$$

Dans la base $\{U_1 \cdots U_p\}$ de vecteurs propres normés, on a une corrélation maximale, dans la direction de U_1 , pour des variables comparables, puisque centrées et réduites. Puis ensuite, si l'on cherche une autre information, non corrélée à celle déjà trouvée, on choisit une direction non corrélée à $Vect(U_1)$, c'est à dire, orthogonale à U_1 pour q_2 , dans $Vect(U_2, \dots, U_p)$. Compte tenu de l'orthogonalisation simultanée, U_2 est aussi orthogonal à U_1 pour le produit scalaire usuel. On obtient la deuxième direction principale, U_2 . Dans ce contexte, les directions principales sont appelées axes factoriels.

Le second changement de base effectué dans le dual de \mathbb{R}^p peut se traduire par une nouvelle matrice $X P_2$, décrivant nos n individus par les p colonnes des nouvelles variables. Plus précisément, si U_α est la α -ième colonne de P_2 , $X U_\alpha$ est la liste des coordonnées de nos individus sur l'axe factoriel de la α -ème variable principale : $Vect(U_\alpha)$. La variance de cette variable est $U_\alpha^t . X^t . X . U_\alpha = q_2(U_\alpha) = \lambda_\alpha$. On peut réduire cette variable principale, en la normant pour q_2 , en considérant $\frac{X . U_\alpha}{\sqrt{\lambda_\alpha}}$.

Variables et individus

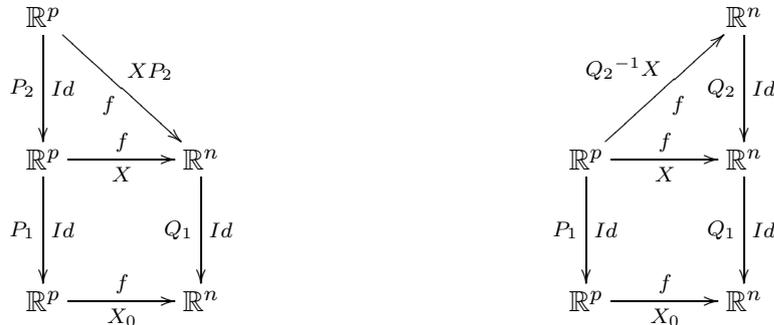
Dans \mathbb{R}^p , on représente les n individus en normant les variables (base duale), composantes de ces individus. Les deux premières variables sont les plus significatives pour le problème étudié. C'est, sur une projection dans le plan $Vect(U_1, U_2)$, que les variables, formes coordonnées, seront les plus corrélées et que la projection du nuage des individus sera la plus étalée.

La matrice X , contenant les p données centrées réduites, sur une population "normée" de n individus est de taille (n, p) . C'est la matrice de $f^* \circ f$, matrice de corrélation $X^t . X$, que l'on diagonalise, dans $M_{p,p}(\mathbb{R})$, dans l'espace des variables. Si les U_i dans \mathbb{R}^p , sont vecteurs propres associés aux valeurs propres non nulles, on en déduit, pour l'autre matrice symétrique, $X . X^t$, que l'on peut considérer dans $M_{n,n}(\mathbb{R})$, (voir 1), les vecteurs propres $X U_i$. On trouve ainsi, dans l'espace des individus, les deux premières directions principales $\frac{X U_1}{\sqrt{\lambda_1}}$ et $\frac{X U_2}{\sqrt{\lambda_2}}$, vecteurs propres normés, correspondant aux deux mêmes plus grandes valeurs propres λ_1 , respectivement λ_2 , que dans l'espace des variables. La matrice de changement de base Q_2 a pour premières colonnes les $\frac{X U_i}{\sqrt{\lambda_i}}$. (voir [LMP] remarque de bas de page 20).

Pour cette deuxième étape dans les changements de coordonnées,

dans l'espace des variables
on change les variables
pas les individus,

dans l'espace des individus
on change les individus
pas les variables.



Dans \mathbb{R}^n , espace des individus, on représente les p vecteurs variables, données initiales centrées réduites, par leurs n coordonnées sur les n individus "normées pour la pondération". Si l'on note $(Y_1 \dots Y_p)$ la base des variables centrées réduites, les coordonnées des vecteurs variables sont $(f(Y_1) \dots f(Y_p))$, colonne de X . La matrice de corrélation est la matrice de Gramm des vecteurs variables. De là, les vecteurs variables sont de norme 1 et la corrélation entre deux variables est le cosinus de l'angle que forment les deux vecteurs représentatifs de ces variables dans \mathbb{R}^n .

Les vecteurs variables deviennent, avec le changement isométrique de variables dans \mathbb{R}^n , les colonnes de $Q_2^{-1}X$ ou les lignes de $X^t.Q_2$. les premières colonnes de $X^t.Q_2$ sont les $X^t \cdot \frac{XU_i}{\sqrt{\lambda_i}} = \sqrt{\lambda_i}U_i$ correspondants aux $\lambda_i \neq 0$. Cela donne, en ligne, les premières coordonnées des p vecteurs variables.

Les angles entre leurs projections dans le plan principal donnent une représentation "optimale" dans un plan (sur le papier) de la "proximité" des variables. On lit les coordonnées de ces p projections dans les p lignes de la matrice $(\sqrt{\lambda_1}U_1, \sqrt{\lambda_2}U_2)$. Pour les dessins voir [LMP] page 46.

9 Mécanique :

Axe d'inertie d'un solide, en mouvement autour de son centre de gravité

Autre idée à travailler dans [Gob] page 311.

Références

[Aud] M.Audin, Géométrie, Belin

[Ber] M.Berger, Géométrie

[EscPag] B.Escofier, J.Pagès, Initiation aux traitements statistiques, collection DIDACT STATISTIQUE, PUR (Presse Universitaire de Rennes).

[Gob] R.Goblot, Algèbre linéaire, Masson.

- [Gour] X.Gourdon, Mathématiques pour M', Ellipses.
- [Grif] J.Grifone, Algèbre linéaire, Cepadues.(bon livre de premier cycle classique. Quelques erreurs. Remettre les vecteurs à la verticale).
- [LMP] L.Lebart, A.Morineau, M.Piron, Statistique exploratoire multidimensionnelle, Dunod.
- [LFA1] Lelong-Ferrand, J.M.Arnaudies, t1 Algèbre, Dunod.
- [LFA3] Lelong-Ferrand, J.M.Arnaudies, t3 Géométrie, Dunod.
- [RDO] E.Ramis, G.Deschamps, J.Odoux, Math Spé 2, Masson.
- [Sap] Saporta, Probabilités Analyse des données et statistiques, TECHNIP.
- [Ser] D.Serre, Les Matrices, Dunod. (Très clair et concis sur la réduction des endomorphismes normaux, contient par ailleurs, Perron-Frobenius et module sur A principal. Attention, erreur en haut de la page 3).