REVUE DE STATISTIQUE APPLIQUÉE

A. KOBILINSKY

Ordre entre formes quadratiques. Application à l'optimalité de sous-espaces en analyse des données

Revue de statistique appliquée, tome 27, nº 1 (1979), p. 45-54

<http://www.numdam.org/item?id=RSA_1979__27_1_45_0>

ORDRE ENTRE FORMES QUADRATIQUES. APPLICATION A L'OPTIMALITÉ DE SOUS-ESPACES EN ANALYSE DES DONNÉES

A. KOBILINSKY

INRA, Laboratoire de Biométrie du CNRA,
Etoile de Choisy, Route de Saint-Cyr, 78000 VERSAILLES

RESUME

Après avoir rappelé les principaux théorèmes sur les inégalités entre formes quadratiques, nous en déduisons les propriétés d'optimalité des sous-espaces principaux obtenus en Analyse en Composantes Principales (A.C.P).

Essentiellement géométrique, l'exposé diffère notablement de celui de OKAMOTO (1969). Il met en évidence la complète symétrie entre les critères relatifs aux individus et ceux relatifs aux variables et établit les relations d'ordre qui existent entre ces différents critères. Les résultats, qui font apparaître clairement les deux formes quadratiques en jeu, sont applicables directement aux analyses de données qui se ramènent à l'A.C.P. : A.C.P. normée, Analyse Factorielle des Correspondances, Discriminante, Canonique éventuellement généralisée à plus de deux groupes de variables par la méthode de CAROLL (1968).

INTRODUCTION

L'analyse en Composantes Principales (A.C.P.) a été introduite par PEARSON (1901), qui recherchait une image de dimension réduite d'un nuage de points multidimensionnel. Une telle recherche implique le choix d'un critère de qualité de la représentation, permettant de comparer les images que l'on peut obtenir. Ainsi, le sous espace le mieux ajusté avec le critère de PEARSON est celui qui minimise la somme des carrés des distances aux points du nuage.

Il s'avère que le sous-espace le mieux ajusté au sens de PEARSON est en fait optimal pour plusieurs autres critères. OKAMOTO (1969) a fait une étude synthétique de ces critères qu'il classe en trois catégories, selon qu'ils s'énoncent en terme de variation, de perte d'information ou de corrélation.

Nous reprenons ici cette étude, mais sur un plan beaucoup plus géométrique, qui :

i) fait ressortir l'entière symétrie entre critères relatifs aux individus et critères relatifs aux variables ;

ii) permet d'établir des relations d'ordre entre les différents critères ;

iii) s'applique directement aux analyses de données qui se ramènent à l'A.C.P.

Les propriétés d'optimalité de l'A.C.P. sont déduites de deux théorèmes essentiels sur les inégalités entre formes quadratiques : le théorème de GAUSS-MARKOV, énoncé sous une forme géométrique voisine de celle de CAILLIEZ et PAGES, et un théorème permettant de passer de l'ordre habituel sur les formes quadratiques à un ordre sur les spectres.

Mots-clés : **Analyse des données, Composantes principales, Optimalité, Formes quadratiques, Ordre, Gauss - Markov.**

L'application à deux A.C.P. particulières – l'analyse canonique de plus de deux groupes de variables, telle qu'elle est définie par CAROLL, et l'analyse discriminante – de certaines propriétés d'optimalité illustre l'intérêt qu'il y a à faire apparaître une forme quadratique positive quelconque M pour définir la distance entre individus.

1. ORDRE SUR LES FORMES QUADRATIQUES

1.1. Notations – Définitions

Nous supposerons connu dans ce qui suit le contenu du chapitre VIII du livre de CAILLIEZ et PAGES (1976).

E est un espace vectoriel réel de dimension finie p, muni d'une forme quadratique M définie positive (espace préhilbertien séparé). M est identifié à une application linéaire de E dans son dual E* par la relation :

$$M(\underline{u}, \underline{u}) = \langle M\underline{u}, \underline{u} \rangle .$$

De même une forme quadratique V (non nécessairement définie positive) sur E* sera identifiée à une application linéaire de E* dans E :

$$V(\underline{v}, \underline{v}) = \langle \underline{v}, V\underline{v} \rangle .$$

Définition 1

Ordre partiel sur les formes quadratiques V_1 et V_2 étant deux formes quadratiques, V_1 est dite inférieure à V_2 ($V_1 \leqslant V_2$) si la forme quadratique $V_2 - V_1$ est semi-définie positive. On vérifie aisément que la relation $\leqslant$ est bien une relation d'ordre.

1.2. Ordre sur les spectres

Soient $\underline{u}_1, \ldots, \underline{u}_p$ une base M-orthonormée de vecteurs propres de V_1M et $\underline{v}_1, \ldots, \underline{v}_p$ une base M-orthonormée de vecteurs propres de V_2M.

On notera λ_i la valeur propre de $\underline{u}_i$

$$\mu_i \text{ celle de } \underline{v}_i .$$

On supposera que les spectres $\lambda_1, \ldots, \lambda_p$ de V_1M et $\mu_1, \ldots, \mu_p$ de V_2M sont ordonnés par valeurs décroissantes :

$$\lambda_1 \geqslant \cdots \geqslant \lambda_p , \quad \mu_1 \geqslant \cdots \geqslant \mu_p$$

Théorème 1

$V_1 \leqslant V_2 \Rightarrow \lambda_i \leqslant \mu_i$ pour $i = 1, \ldots, p$ et quelle que soit la métrique M.

Démonstration

L'espace engendré par $\underline{v}_1, \ldots, \underline{v}_i$ de dimension i et l'espace engendré par $\underline{u}_i, \ldots, \underline{u}_p$ de dimension $p - i + 1$ ont un vecteur commun $\underline{u}$ non nul.

$$\underline{u} = \alpha_1 \underline{}_1 + \cdots + \alpha_i \underline{u}_i \qquad (1)$$
$$\underline{u} = \beta_i \underline{u}_i + \cdots + \beta_p \underline{u}_p \qquad (2)$$

On peut supposer $\underline{u}$ normé pour M. On a alors :

$$\alpha_1^2 + \cdots + \alpha_i^2 = 1$$
$$\beta_i^2 + \cdots + \beta_p^2 = 1$$

En utilisant l'égalité (1), on voit que :

$$V_1(M\underline{u}, M\underline{u}) = \langle M\underline{u}, V_1 M\underline{u} \rangle = M(\underline{u}, V_1 M\underline{u}) = \lambda_1 \alpha_1^2 + \cdots + \lambda_i \alpha_i^2 \geqslant \lambda_i$$

On démontre de même à partir de (2) que :

$$V_2(M\underline{u}, M\underline{u}) \leqslant \mu_i$$

On a donc : $\lambda_i \leqslant V_1(M\underline{u}, M\underline{u}) \leqslant V_2(M\underline{u}, M\underline{u}) \leqslant \mu_i$.

L'implication en sens inverse : $\lambda_i \leqslant \mu_i$, $i = 1, \ldots, p \Rightarrow V_1 \leqslant V_2$ est fausse. Cela résulte de la proposition ci-dessous et du fait du $AV_1A'M$ et V_1M ont même spectre quand A est M-orthogonale ($A\underline{u}_i$ est vecteur propre de $AV_1A'M$ avec la valeur propre λ_i).

Proposition

Si A est une application linéaire M-orthogonale (*), l'inégalité $AV_1A' \leqslant V_1$ implique l'égalité $AV_1A' = V_1$. Cette égalité a lieu si A est l'identité et plus précisément si et seulement si A est invariante sur chaque sous-espace propre de V_1M.

Note : A est dite invariante sur un sous-espace F si AF = F.

L'invariance de A sur les sous-espaces propres de V_1M implique que AV_1A' et V_1 coincident sur la base $MA\underline{u}_1, \ldots, MA\underline{u}_p$ de E*.

Réciproprement, si A n'est pas invariante et que $\underline{u}_k$ est le premier vecteur propre "non invariant" de V_1M, on montre que :

$$V_1(\underline{v}, \underline{v}) < AV_1A'(\underline{v}, \underline{v}) \quad \text{où} \quad \underline{v} = MA\underline{u}_k .$$

Si l'implication du théorème 1 ne peut être remplacée par une équivalence, on a cependant le :

Théorème 2

Les assertions suivantes, relatives aux formes quadratiques V_1 et V_2 sont équivalentes :

i) $\lambda_i \leqslant \mu_i$ pour $i = 1, \ldots, p$,

ii) il existe une application linéaire M-orthogonale A telle que $AV_1A' \leqslant V_2$.

Démonstration

ii) $\Rightarrow$ i) : résulte du théorème 1 et de la remarque sur l'identité des spectres de V_1M et $AV_1A'M$ lorsque A est M-orthogonale.

(*) Par définition, A vérifie $M(A\underline{u}, A\underline{u}) = M(\underline{u}, \underline{u})$, ou encore $A'MA = M$.

i) $\Rightarrow$ ii) : A étant définie par $A\underline{u}_i = \underline{v}_i$, tout vecteur $\underline{v}$ de E* peut s'écrire sous la forme $\underline{v} = MA\underline{u}$ où $\underline{u} = \alpha_1 \underline{u}_1 + \cdots + \alpha_p \underline{u}_p$. On a alors :

$$AV_1A'(\underline{v}, \underline{v}) = \sum_{i=1}^{p} \lambda_i \alpha_i^2 \leqslant \sum_{i=1}^{p} \mu_i \alpha_i^2 = V_2(\underline{v}, \underline{v})$$

Définition 2 (préordre partiel sur les formes quadratiques)

V_1 et V_2 étant deux formes quadratiques, V_1 sera dite M-inférieure à V_2 ($V_1 \underset{M}{\lesssim} V_2$) si il existe une application linéaire A M-orthogonale telle que $AV_1A' \leqslant V_2$.

Définition 3 (ordre sur les spectres)

Si ayant ordonné par valeurs décroissantes les valeurs propres des endomorphismes A et B, chaque valeur propre de A est inférieure à la valeur propre de même rang de B, le spectre de A est dit inférieur au spectre de B : $Sp(A) \leqslant Sp(B)$.

Avec ces définitions, le théorème 2 peut être énoncé sous la forme :

$$Sp(V_1 M) \leqslant Sp(V_2 M) \Leftrightarrow V_1 \underset{M}{\lesssim} V_2 .$$

2. APPLICATION AUX FORMES QUADRATIQUES OBTENUES PAR PROJECTION

Dans ce qui suit, E désigne toujours un espace vectoriel réel de dimension p.

2.1. Théorème de Gauss-Markov

Théorème 3 (Gauss-Markov)

Soit A un projecteur sur un sous-espace de E, V une forme quadratique définie positive sur E*. Les assertions suivantes sont équivalentes :

i) A est V^{-1} orthogonal

ii) $V \geqslant AVA'$

iii) $BVB' \geqslant AVA'$ pour tout autre projecteur B ayant même espace image que A.

Démonstration

i) $\Rightarrow$ ii) : si A est V^{-1} orthogonal, A' est V-orthogonal et on a l'égalité : $V = AVA' + (I - A)V(I - A)'$ qui implique que $V \geqslant AVA'$.

ii) $\Rightarrow$ iii) : $V \geqslant AVA' \Rightarrow BVB' \geqslant BAVA'B' = AVA'$.

iii) $\Rightarrow$ i) : soit G l'espace image de A. Prenons pour B le projecteur V^{-1} orthogonal sur G. Posant $A_1 = I - A$, $B_1 = I - B$, l'inégalité iii) peut s'écrire sous la forme :

$$\| \underline{u} - B_1' \underline{u} \|_V^2 \geqslant \| \underline{u} - A_1' \underline{u} \|_V^2 . \qquad (1)$$

Or A'_1 et B'_1 ont même espace image, à savoir l'orthogonal $G^{\perp}$ de G dans E*. B étant V^{-1} orthogonal, B'_1 est V-orthogonal. L'inégalité (1) ci-dessus montre alors que $A'_1 \underline{u}$ est le point de $G^{\perp}$ le plus proche de $\underline{u}$ pour la métrique V. A'_1 est donc V-orthogonal et par suite A est V^{-1} orthogonal.

Remarques

1) le théorème implique en particulier que la forme quadratique BVB' où B est un projecteur n'est comparable à V que si B est V^{-1} orthogonal.

2) la V^{-1} orthogonalité de A est équivalente à la V-orthogonalité de A'.

3) E étant le dual de E*, le théorème peut être réénoncé avec une forme quadratique définie positive M sur E.

2.2. Commutativité du spectre

Si $A : E \rightarrow F$ et $B : F \rightarrow E$ sont deux applications linéaires, les produits de composition AB et BA ont même spectre (à la valeur 0 près si les dimensions de F et E diffèrent) : Sp (AB) = Sp (BA).

2.3. Encadrement de Sp $(PVP'M)$

a) Notations

Dans ce qui suit, nous supposerons que M et V sont des formes quadratiques définies positives respectivement sur E et E*.

Soit P le projecteur M-orthogonal sur le sous-espace G de E. P' est donc le projecteur M^{-1} orthogonal sur MG.

Soient alors Q le projecteur V^{-1} orthogonal sur G
et Q_1 le projecteur V^{-1} orthogonal sur VMG.

Remarque : il y a une symétrie complète entre E, M, P, Q d'une part et E*, V, Q'_1, P' de l'autre. Cette symétrie sera utilisée dans la démonstration du théorème 4 qui suit.

b) Majoration de Sp $(PVP'M)$

Théorème 4

Avec les notations ci-dessus, on a les inégalités :

$$\mathrm{Sp}\,(QVQ'M) \leqslant \mathrm{Sp}\,(PVP'M) \leqslant \mathrm{Sp}\,(Q_1VQ'_1M) \leqslant \mathrm{Sp}\,(VM)$$

Démonstration

Appliquant le théorème 3 avec B = P, A = Q, on obtient l'inégalité $QVQ' \leqslant PVP'$ d'où l'on déduit l'inégalité : $\mathrm{Sp}\,(QVQ'M) \leqslant \mathrm{Sp}\,(PVP'M)$.

De façon symétrique, on a : $P'MP \leqslant Q'_1MQ_1$ d'où l'on déduit en utilisant la commutativité du spectre :

$$\mathrm{Sp}\,(PVP'M) = \mathrm{Sp}\,(P'MPV) \leqslant \mathrm{Sp}\,(Q'_1MQ_1V) = \mathrm{Sp}\,(Q_1VQ'_1M)$$

La dernière inégalité : $\mathrm{Sp}\,(Q_1VQ'_1M) \leqslant \mathrm{Sp}\,(VM)$ s'obtient immédiatement en utilisant le théorème de Gauss-Markov et le théorème 1.

c) **Minoration de** Sp (PVP'M)

Le spectre de PVP'M est égal à celui de P'MV. Or :

Lemme : Les applications P'MV et $Q_1' V^{-1} M^{-1}$ restreintes à MG sont inverses l'une de l'autre.

En effet, si $\underline{u}$ et $\underline{v}$ sont deux vecteurs de MG, on a :

$$V(\underline{u}, \underline{v}) = M^{-1}(\underline{u}, MV\underline{v}) = M^{-1}(\underline{u}, P'MV\underline{v}) = V(\underline{u}, V^{-1}M^{-1}P'MV\underline{v}) = \\ = V(\underline{u}, Q_1'V^{-1}M^{-1}P'MV\underline{v})$$

d'où $\underline{v} = Q_1'V^{-1}M^{-1}P'MV\underline{v}$.

On déduit de ce lemme le théorème suivant :

Théorème 5

Soient $\lambda_1(P) \geqslant \cdots \geqslant \lambda_r(P)$ les valeurs propres non nulles de PVP'M, $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ celles de VM. On a :

$$\begin{array}{ccc} \lambda_1(P) \leqslant \lambda_1 & & \lambda_p \leqslant \lambda_r(P) \\ \vdots & \text{et} & \vdots \\ \lambda_r(P) \leqslant \lambda_r & & \lambda_{p-r+1} \leqslant \lambda_1(P) \end{array}$$

Démonstration

Le premier système d'inégalités n'est autre que l'inégalité Sp(PVP'M) $\leqslant$ Sp (VM) qui apparaît dans le théorème 4.

Remplaçant P, V et M respectivement par Q_1', V^{-1} et M^{-1} dans cette inégalité, on obtient l'inégalité : $\mathrm{Sp}(Q_1'V^{-1}Q_1M^{-1}) \leqslant \mathrm{Sp}(V^{-1}M^{-1})$. Or les valeurs propres non nulles de $Q_1'V^{-1}Q_1'M^{-1} = Q_1'V^{-1}M^{-1}$ sont d'après le lemme les inverses des valeurs propres non nulles de P'MV qui a même spectre que PVP'M. On en déduit le second système d'inégalités.

Remarque : si V n'est pas définie positive, le théorème 5 reste vrai. La démonstration est alors analogue à celle du théorème 1.

3. APPLICATIONS A L'ANALYSE EN COMPOSANTES PRINCIPALES

3.1. Notations

Elles sont toujours identiques à celles de CAILLIEZ et PAGES, chap. III. $X(p, n) = (\underline{x}_1, \ldots, \underline{x}_n)$ désigne donc la matrice des données, $\underline{x}_i$ étant le vecteur de $E = R^p$ associé au i-ième individu.

M est la forme quadratique définie positive sur E définissant la distance entre les individus.

Chaque individu est muni d'un poids p_i. L'espace $F = R^n$ qui contient les caractères (ou variables) $\underline{x}^1, \ldots, \underline{x}^p$ (colonnes de X') est alors muni de la métrique euclidienne des poids D_p.

X est supposée centrée : $\sum_i p_i \, \underline{x}_i = 0$

E* et F* sont munis respectivement des formes quadratiques $V = XD_pX'$ et $W = X'MX$ figurées sur le schéma de dualité ci-dessous :

$$\begin{tikzcd}
E \arrow[d, shift right, "M"'] & F^* \arrow[l, "X"'] \arrow[d, shift right, "W"'] \\
E^* \arrow[u, shift right, "V"'] \arrow[r, "X'"'] & F \arrow[u, shift right, "D_p"']
\end{tikzcd}$$

De plus, $\underline{u}_1, \ldots, \underline{u}_p$ désigne une base M-orthonormée de vecteurs propres de VM ayant $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ comme valeur propres respectives.

Le sous espace engendré par $\underline{u}_1, \ldots, \underline{u}_r$ est appelé sous-espace principal de dimension r.

De même, si A est un projecteur de rang r de E, on désigne par

$$\lambda_1(A) \geqslant \cdots \geqslant \lambda_r(A)$$

les r plus grandes valeurs propres de AVA'M, et par $\underline{u}_1(A), \ldots, \underline{u}_r(A)$ un système M-orthonormé de vecteurs propres de AVA'M associés respectivement à $\lambda_1(A), \ldots, \lambda_r(A)$.

3.2. Optimalité sur les individus

Les propriétés énoncées ci-dessous sont des corollaires immédiats des théorèmes qui précèdent :

a) P étant un projecteur M-orthogonal de rang r, on sait que l'inertie du nuage des $P\underline{x}_i$: tr(PVP'M), est maximale lorsque P est le projecteur M-orthogonal sur le sous-espace principal de dimension r. Le théorème 5 montre que le critère de dispersion "inertie" peut être remplacé par n'importe quelle fonction $f(\lambda_1(P), \ldots, \lambda_r(P))$ croissante en chaque argument. Le produit : $\lambda_1(P), \ldots, \lambda_r(P)$, la somme des carrés : $\lambda_1^2(P) + \cdots + \lambda_p^2(P)$, sont des exemples de fonctions de ce type.

b) Soient N (resp. N_0) le nuage obtenu par projection M-orthogonale sur un sous-espace G de dimension r (resp. le sous-espace principal de dimension r). Si l'on superpose les nuages N et N_0 au moyen de l'isométrie qui amène $\underline{u}_i(P)$ sur $\underline{u}_i$, il résulte du théorème 5 que N_0 est plus allongé que N(*) dans toutes les directions.

c) Enfin, on peut définir l'A.C.P. comme la recherche du projecteur M-orthogonal P tel que l'espace image de P soit le plus proche possible du nuage des $\underline{x}_i$. On utilisera comme indice de proximité un indice de dispersion du nuage des

----

(*) L'allongement dans une direction est mesuré par l'inertie de la projection orthogonale du nuage dans cette direction.

$\underline{x}_i - P\underline{x}_i$. Appliquant le théorème 5 au projecteur $(I - P)$ de rang $p - r$, on voit que :

$$\lambda_1 (I - P) \geqslant \lambda_{r+1}$$
$$\vdots$$
$$\lambda_{p-r}(I - P) \geqslant \lambda_p$$

Ceci montre qu'avec un indice de proximité qui est une fonction de $\lambda_1 (I - P), \ldots, \lambda_r (I - P)$, croissante en chaque argument, l'espace de dimension r engendré par $\underline{u}_1, \ldots, \underline{u}_r$ est le plus proche du nuage des $\underline{x}_i$.

3.3. Optimalité sur les variables

Les propriétés étudiées en 3.2 sont relatives à un projecteur P M-orthogonal. Elles s'interprètent donc bien en terme de projection d'un nuage d'individus. Nous nous intéressons dans ce paragraphe aux projections D_p-orthogonales, donc aux liaisons entre variables.

Soit donc $\underline{y}^1, \ldots, \underline{y}^r$ des variables engendrant un espace de dimension r, Q le projecteur sur cet espace. $Q\underline{x}$ est donc la variable expliquée par la régression de $\underline{x}$ sur $\underline{y}^1, \ldots, \underline{y}^r$. Notons alors V_e la matrice de covariance de $\underline{x}^1, \ldots, \underline{x}^p$ expliquée par la régression sur $\underline{y}^1, \ldots, \underline{y}^r$. V_e est la matrice de l'application linéaire $XQ'D_pQX'$.

Recherche du meilleur r-uplet de prédicteurs de $x^1, \ldots, x^p$.

On peut supposer que $\underline{y}^1, \ldots, \underline{y}^r$ engendrent un sous espace X'H de X'E. En effet dans le cas contraire, on obtient une prédiction au moins aussi bonne (au sens de l'ordre $\leqslant$ défini en 1) en remplaçant $\underline{y}^1, \ldots, \underline{y}^r$ par leurs projection sur X'E.

Si Q_1' est le projecteur V-orthogonal sur H, il est immédiat que $X'Q_1' = QX'$. On a donc :

$$V_eM = XQ'D_pQ\,X'M = Q_1VQ_1'M$$

d'où la proposition :

Proposition

V_e étant la matrice de covariance expliquée par la régression de $\underline{x}^1, \ldots, \underline{x}^p$ sur un système $\underline{y}^1, \ldots, \underline{y}^r$ de r prédicteurs, $\mathrm{Sp}(V_eM)$ est maximum quand $\underline{y}^1, \ldots, \underline{y}^r$ sont les r premières composantes principales de $\underline{x}^1, \ldots, \underline{x}^p$.

Note : On a une propriété analogue en remplaçant V_e par la matrice de covariance résiduelle $V_r = V - V_e$ et maximum par minimum.

3.4. Expression des traces de $PVP'M$ et $XQ'D_p\,QX'M$

Dans le cas particulier où la fonction croissante du spectre utilisée pour définir l'optimalité est la trace, il est intéressant de développer cette trace : on retrouve ainsi des propriétés classiques :

$$\mathrm{tr}(PVP'M) = \mathrm{tr}(D_p X'P'MPX) = \sum_{i=1}^{n} p_i \| P\underline{x}_i \|_M^2$$

$$\mathrm{tr}(XQ'D_p Q X'M) = \sum_{i,j} m_{ij} D_p(Q\underline{x}^i, Q\underline{x}^j)$$

(m_{ij}) étant la matrice de M : $m_{ij} = \langle Me_i, e_j \rangle$

Cas de l'Analyse Factorielle des Correspondances sur tableau disjonctif complet (Burt) et de la canonique généralisée (Caroll).

Dans ces deux cas, les variables sont partitionnées en k paquets (1 paquet = les indicatrices associées aux modalités d'une variable en AFC). A ce découpage correspond un partitionnement de X' en k sous-matrices : $X' = (X'_1, X'_2, \ldots, X'_k)$.

V_{jj} est la matrice de covariance (forme quadratique d'inertie) correspondant au j-ième paquet : $V_{jj} = X_j D_p X'_j$.

La métrique M sur les individus est définie par la matrice bloc diagonale :

$$\begin{bmatrix} V_{11}^{-1} & & \\ & \ldots & \\ & & V_{kk}^{-1} \end{bmatrix}$$

On a alors :

$$\mathrm{tr}(XQ'D_p Q X'M) = \sum_{j=1}^{k} \mathrm{tr}(X_j Q' D_p Q X'_j V_{jj}^{-1}) = \sum_{j=1}^{k} \mathrm{tr}(X'_j V_{jj}^{-1} X_j D_p Q) =$$

$$= \sum_{j=1}^{k} \mathrm{tr}(Q_j Q)$$

où $Q_j = X'_j V_{jj}^{-1} X_j D_p$ est le projecteur D_p orthogonal sur l'espace H_j engendré par les caractères du j-ième paquet, Q étant le projecteur sur un sous-espace H de dimension r de F.

Dans l'expression $\mathrm{tr}(Q_j Q)$, on reconnait la somme des carrés des corrélations canoniques entre les espaces H_j et H. On peut donc énoncer :

Proposition

Si l'on mesure la proximité entre les sous-espaces H et H_j de F par la somme $d^2(H, H_j)$ des carrés des coefficients de corrélation canonique entre H et H_j, la somme $\sum_{j=1}^{k} d^2(H, H_j)$ est maximum quand H est engendré par les r premières composantes principales.

3.5. Cas particulier de l'analyse discriminante

L'analyse discriminante peut être considérée comme une analyse en composantes principales, où les matrices de M et V sont respectivement l'inverse de la matrice de covariance résiduelle (intra) W et la matrice inter B :

$$M = W^{-1}$$
$$V = B$$

Si P' désigne le projecteur W-orthogonal sur un sous espace H de dimension r de E^*, la restriction à H de W et B est PWP' et PBP'.

L'inverse de l'application linéaire PWP' restreinte à H et WH est $P'MP$. Les facteurs discriminants de l'analyse restreinte à H s'obtiennent donc par diagonalisation de : $PVP'P'MP = PVP'M$.

Le théorème 5 peut donc s'énoncer sous la forme : le plus discriminant des espaces de facteurs de dimension r est l'espace engendré par les r premiers facteurs discriminants, la discrimination étant mesurée par une fonction croissante en chaque argument des valeurs propres de l'analyse discriminante sur ces r facteurs.

REFERENCES

[1] CAILLIEZ F., et PAGES J.P. – *Introduction à l'analyse des données.* Paris, Smash (éd.), p. 616, 1976.

[2] CAROLL J.D. – A generalisation of canonical correlation analysis to three or more sets of variables. *Proc. 76 th Conv. Amer. Psych. Assoc.,* 227-228, 1968.

[3] OKAMOTO M. – Optimality of Principal Components. In *Multivariate Analysis,* 2 (P.R. Krishnaiah, éd.), 673-685. Academic Press, New-York, 1969.