

# CAHIERS DU BURO

A. CROQUETTE

A. KOBILINSKY

J. P. PAGES

Y. SCHEKTMAN

**Optimalité en analyse factorielle Vers de nouveaux critères et de nouvelles stratégies. Première partie : Décomposition de la matrice de variance. Préordre entre formes quadratiques et optimalité**

*Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche*, tome 43-44 (1984), p. 3-41

[http://www.numdam.org/item?id=BURO\\_1984\\_\\_43-44\\_\\_3\\_0](http://www.numdam.org/item?id=BURO_1984__43-44__3_0)

© Institut Henri Poincaré — Institut de statistique de l'université de Paris, 1984, tous droits réservés.

L'accès aux archives de la revue « Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## OPTIMALITE EN ANALYSE FACTORIELLE

Vers de nouveaux critères et de nouvelles stratégies

A. CROQUETTE<sup>1</sup>, A. KOBILINSKY<sup>2</sup>, J.P. PAGES<sup>3</sup>, Y. SCHEKTMAN<sup>1</sup>

### PREMIÈRE PARTIE :

DECOMPOSITION DE LA MATRICE DE VARIANCE.  
PREORDRE ENTRE FORMES QUADRATIQUES ET OPTIMALITE.

<sup>1</sup> *Laboratoire Modèles et logiciels d'analyse de données.  
Univ. Paul Sabatier. Toulouse - GRECO-CNRS 59.*

<sup>2</sup> *INRA-CNRA. Laboratoire de Biométrie.  
Route de St Cyr 78000 Versailles.*

<sup>3</sup> *Laboratoire de Statistique et d'Etudes Economiques et Sociales.  
Centre d'Etudes Nucléaires de Fontenay-aux-Roses.*

On ne peut inventer de nouvelles stratégies pour rechercher des axes ou des sous-espaces optimaux sans avoir au préalable passé en revue, d'une part, les principales propriétés d'optimalité des analyses factorielles, en les présentant de façon synthétique et en exploitant la dualité, et analysé et regroupé en familles, d'autre part, les critères qui permettent de retrouver axes et sous-espaces principaux. C'est à un tel examen de l'analyse factorielle, sous l'angle de l'optimalité, qu'est convié le lecteur dans cet article.

En analyse factorielle, rechercher un sous-espace optimal revient à rechercher une décomposition optimale de la matrice de variance. C'est la raison pour laquelle, dans une première partie, on se consacre aux décompositions des formes quadratiques. On examine, dans un premier temps, les différents types de décomposition que l'on rencontre en analyse factorielle et plus généralement en analyse de données multidimensionnelles ; puis l'optimalité des décompositions est envisagée en faisant intervenir les deux préordres les plus naturels entre formes quadratiques . C'est en les regardant sous l'angle de ces deux préordres que l'on peut commencer à classer les critères permettant de retrouver les sous-espaces principaux.

Dans une deuxième partie, ayant analysé le domaine de variation du spectre des opérateurs qui caractérisent les sous-espaces et celui des systèmes de coefficients de Rayleigh associés aux bases orthogonales, on passe en revue et on analyse les propriétés de certaines familles de critères permettant de retrouver soit les sous-espaces principaux soit les bases engendrant ces sous-espaces.

SOMMAIRE DE LA PREMIERE PARTIE

	Pages
1. DECOMPOSITION DES MATRICES SYMETRIQUES POSITIVES DE TYPE $V = V_1 + V_2$ . DUALITE ET ORTHOGONALITE ENTRE SOUS-ESPACES	9
1.1 Analyse en composantes principales	10
1.2 Régression et analyse canonique	11
1.3 Analyse factorielle au sens de Spearman et théorie des images	13
2. AJUSTEMENT OPTIMAUX DE RANG FIXE INDUITS PAR LES INEGALITES USUELLES ENTRE FORMES QUADRATIQUES	20
2.1 Inégalités basées sur l'inclusion des ellipsoïdes	20
2.1.1 Ajustement par en dessous optimal de rang q	21
2.1.2 Application à la régression	25
2.1.3 Application à l'estimation des scores en analyse factorielle au sens de Spearman	26
2.2 Inégalités basées sur la comparaison des spectres	31
2.2.1 Ajustement par en dessous optimal de rang q d'une forme positive	33
2.2.2 Le cas d'une forme quelconque	37
Bibliographie	40

En analyse factorielle on est pratiquement toujours conduit à une décomposition de la matrice de variance  $V$  du type :

$$V = V_1 + V_2$$

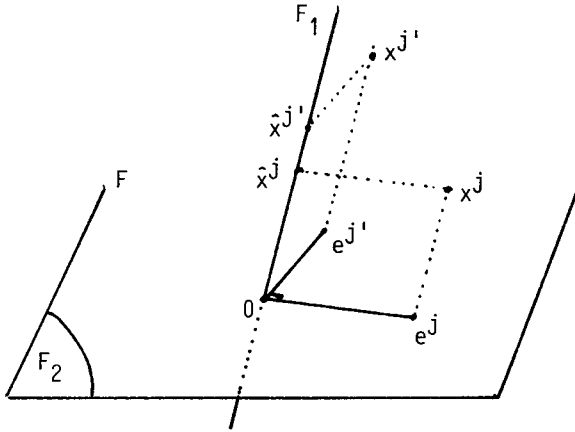
.  $V_1$  , qui peut être assimilée à une matrice de variance "inter", est considérée comme l'approximation de la matrice  $V$  fournie par les facteurs

.  $V_2$  , qui peut être assimilée à une matrice de variance "intra", est la matrice de variance résiduelle.

Alors qu'avec C. Spearman et L. Thurstone on recherche une matrice  $V_1$  , de rang petit, telle que  $V_2$  soit suffisamment proche d'une matrice diagonale, en analyse en composantes principales, avec K. Pearson et H. Hotelling, on recherche une matrice  $V_1$  , de rang petit, qui permette de reconstruire suffisamment  $V$  .

Les deux types de décomposition correspondent à des problèmes géométriques de nature différente :

. en analyse factorielle au sens de Spearman, si  $V_2$  est strictement diagonale, le rang de  $V_1$  étant égal à  $q$ , il correspond à la décomposition de  $V$  une décomposition de l'espace des caractères  $F$  en sous-espaces supplémentaires orthogonaux  $F_1$  , de dimension  $q$  , et  $F_2 = F_1^\perp$  de dimension  $n - q$  , vérifiant la propriété : si on projette les caractères  $x_j$  sur  $F_2$  , les résidus  $e_j$  obtenus sont deux à deux orthogonaux.



La décomposition de la matrice de variance  $V$  correspond donc aux  $p$  décompositions  $x^j = \hat{x}^j + e^j$ , la matrice de variance  $V_2$  des résidus  $e^j$  étant diagonale.

Le sous-espace  $F_1$  n'est pas, en général, contenu dans le sous-espace engendré par les caractères  $x^j$ ; si c'est le cas, au moins  $q$  des résidus  $e^j$  sont nuls. Aussi, en général, à la décomposition en somme directe de  $F$  n'est pas associée par dualité une décomposition en somme directe de l'espace des individus  $E$ .

Un exemple classique d'une telle décomposition, pour  $q=1$ , est celui de la décomposition de la matrice des corrélations  $R$  dont tous les coefficients non diagonaux sont positifs et égaux à  $r$ :

$$R = r jj' + (1-r) I_p$$

où  $j$  est le vecteur (colonne) dont les  $p$  coordonnées sont égales à 1.

Les projections  $\hat{x}^j$  des caractères sur la droite  $F_1$  sont ici égales à un même vecteur de variance  $r$ , les résidus  $e^j$  étant tous de variance  $(1-r)$ .

On peut faire de l'analyse factorielle au sens de Spearman sans poser, comme le faisaient les psychomètres du début du siècle, un modè-

le : on ne postule plus l'existence d'un sous-espace unique de rang  $q$  ; on se contente de rechercher parmi les sous-espaces de dimension  $q$  ,  $q$  étant choisi petit par rapport au nombre de variables  $p$  , celui qui optimise un critère de diagonalité. Si ce critère est "rendre maximum le déterminant de la matrice des corrélations résiduelles" (Howe 1955), il est remarquable de constater que l'on obtient, pour décomposer la matrice de covariance, le même algorithme que celui issu de la méthode du maximum de vraisemblance (Lawley et Maxwell 1963) qui, lui, fait référence à une population de Laplace-Gauss respectant rigoureusement le modèle factoriel.

. En analyse en composantes principales, à la décomposition retenue de  $V$ ,  $V_1$  étant de rang  $q$ , correspond une décomposition de l'espace des individus  $E$  en deux sous-espaces supplémentaires orthogonaux  $E_1$  , de dimension  $q$  , et  $E_2 = E_1^\perp$  , de dimension  $p - q$  . Le sous-espace  $E_1$  est, au sens de l'inertie, le sous-espace de dimension  $q$  le plus proche du nuage engendré par les points individus  $x_i$  .

A cette décomposition de  $E$  correspond, dans l'espace des caractères  $F$ , la décomposition duale du sous-espace engendré par les caractères  $x^j$  . Ceux-ci sont encore décomposés sous la forme  $x^j = \hat{x}^j + e^j$  , mais la matrice de covariance  $V_2$  des résidus n'est plus diagonale ; par contre ici la somme des rangs de  $V_1$  et de  $V_2$  est égale au rang de  $V$ .

Dans l'exemple du modèle à corrélations égales, la matrice de corrélation  $R$  se décompose en somme de deux matrices  $V_1$  et  $V_2$  respectivement de rang 1 et  $p - 1$  ; avec :

$$V_1 = \frac{1 + (p - 1)r}{p} jj' \quad ; \quad V_2 = (1 - r)I_p + \frac{r - 1}{p} jj' .$$

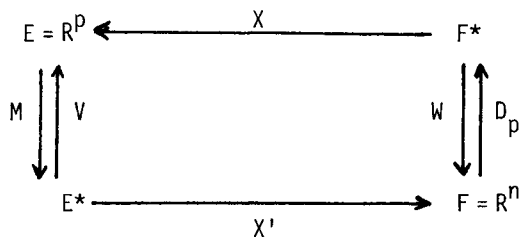
Dans  $E$  le sous-espace propre  $E_1$  de  $R$  de valeur propre  $1 + (p - 1)r$  est engendré par le vecteur  $j$  ; la valeur propre  $(1 - r)$ , d'ordre de multiplicité  $(p - 1)$ , admet l'hyperplan  $E_2 = E_1^\perp$  pour sous-espace propre. Dans  $F$  la décomposition duale du sous-espace engendré par les caractères  $x^j$  fournit les décompositions  $x^j = \hat{x}^j + e^j$  ; avec :  $\hat{x}^j = \frac{1}{p} \sum x^j$  . L'approximation  $\hat{x}^j$  , qui est homothétique à la première composante

principale si  $r$  est positif, est de variance  $\frac{1+(p-1)r}{p}$  ; les corrélations entre les résidus  $e_j$ , de variance  $\frac{(p-1)(1-r)}{p}$ , sont égales à  $\frac{1}{1-p}$ .

Aux deux grands types d'analyse factorielle correspondent donc des recherches de sous-espaces optimaux au sens de critères différents et conduisant à des décompositions de la forme quadratique  $V$  de natures différentes. Dans cette première partie, ayant dans un premier chapitre passé en revue les différentes décompositions de forme quadratique que l'on rencontre en analyse linéaire, on s'intéresse, dans un deuxième chapitre, au problème général de l'ajustement d'une forme quadratique par une autre de rang inférieur fixé.

1. DECOMPOSITION DES MATRICES SYMETRIQUES POSITIVES DE TYPE  $V = V_1 + V_2$ .  
DUALITE ET ORTHOGONALITE ENTRE SOUS-ESPACES

En analyse factorielle la donnée d'un tableau "individus x caractères"  $X$  à  $p$  lignes (caractères  $x^j$ ) et  $n$  colonnes (individus  $x_i$ ) est inséparable de la donnée de deux métriques : la métrique  $M$  permet de mesurer la proximité entre les individus-colonnes ; la métrique diagonale des poids  $D_p$  (distance en moyenne quadratique) permet d'exprimer les cosinus d'angle entre caractères-lignes en terme de corrélation. Le tableau  $X$  étant supposé centré, au triplet  $(X, M, D_p)$  est alors associé le schéma de dualité (Cailliez et Pagès 1976).





Dans ce schéma  $V$  et  $W$  désignent les (pseudo) métriques induites sur les duaux  $E^*$  et  $F^*$  par  $D_p$  et  $M$  respectivement :

$$V = X D_p X' \quad ; \quad W = X' M X \quad .$$

La matrice associée à  $V$ ,  $X$  étant centré, n'est autre que la matrice de variance entre variables  $x^j$ .

### 1.1 Analyse en composantes principales

L'analyse en composantes principales effectuée sur le triplet  $(X, M, D_p)$  fournit simultanément des décompositions de la matrice de variance  $V = X D_p X'$  et de la matrice  $M$ .

Si les axes principaux  $\Delta_{u_i}$ , de valeurs propres  $\lambda_i$ , sont engendrés par les vecteurs  $M$ -orthogonaux  $u_i$ , le système des facteurs principaux  $v_i = M u_i$  étant alors  $M^{-1}$ -orthonormé, on a simultanément :

$$V = \sum \lambda_i u_i u_i' \quad ; \quad M = \sum v_i v_i' \quad .$$

D'où, si on ne retient que les  $q$  premiers axes principaux, les deux décompositions :

$$(1) \quad V = V_1 + V_2 \quad ; \quad (2) \quad M = M_1 + M_2$$

Les deux décompositions vérifient :

$$(3) \quad \text{rang } V = \text{rang } V_1 + \text{rang } V_2 \quad ; \quad (4) \quad \text{rang } M = \text{rang } M_1 + \text{rang } M_2$$

Remarquons que l'on a de plus :

$$\begin{aligned} MVM &= \sum \lambda_i v_i v_i' \quad ; \quad VMV = \sum \lambda_i^2 u_i u_i' \\ V^{-1} &= \sum \frac{1}{\lambda_i} v_i v_i' \quad ; \quad M^{-1} = \sum u_i u_i' \quad . \end{aligned}$$

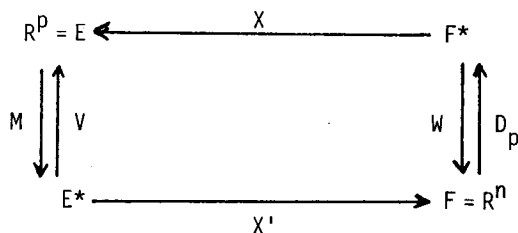
Si l'on retient les  $q$  premiers axes principaux, les décompositions associées à ces formes quadratiques vérifient au niveau des rangs des propriétés analogues à (3) et (4).

Les conditions (1) et (3) sont caractéristiques des décompositions de  $E$  en sous-espaces supplémentaires  $E_1$  et  $E_2$   $V^{-1}$ -orthogonaux ; les conditions (2) et (4) le sont de même pour des sous-espaces  $M$ -orthogonaux. Les quatre conditions sont caractéristiques des décompositions en somme directe bâties à partir de sous-espaces engendrés par des axes principaux.

Si  $E_1$  et  $E_2$  sont des sous-espaces principaux supplémentaires  $M$ -orthogonaux (ils sont alors  $V^{-1}$ -orthogonaux), les projecteurs  $A$  et  $(I-A)$  associés à la décomposition  $E = E_1 + E_2$  sont à la fois  $M$  et  $V^{-1}$ -symétriques ( $A$  vérifie par exemple :  $A'M = MA$  ;  $A'V^{-1} = V^{-1}A$ ) ; on a simultanément :

$$V_1 = AVA' ; V_2 = (I-A)V(I-A') ; M_1 = A'MA ; M_2 = (I-A')M(I-A) .$$

Si on se place dans l'espace des caractères  $F$ , à la décomposition  $E = E_1 + E_2$  de  $E$  correspond, par dualité, la décomposition en somme directe  $F = F_1 + F_2$  de  $F$ , où  $F_1$  est le sous-espace engendré par les  $q$  premières composantes principales et  $F_2$  son supplémentaire  $D_p$ -orthogonal.



Notons  $B$  et  $I-B$  les projecteurs  $D_p$ -symétriques associés à la décomposition  $F = F_1 + F_2$ . Aux décompositions (1) et (2) des formes quadratiques  $M$  et  $V$  correspondent par dualité les décompositions :

$$W = W_1 + W_2 \quad ; \quad D_p = P_1 + P_2 \quad ,$$

des formes quadratiques  $W$  et  $D_p$  . On a :

$$W_1 = B W B' \quad ; \quad W_2 = (I-B) W (I-B') \quad ; \quad P_1 = B' D_p B \quad ; \quad P_2 = (I-B') D_p (I-B) \quad .$$

Les projecteurs  $A$  et  $B$  sont reliés par la formule :

$$A = X B' D_p B X' V^{-1} = X D_p B X' V^{-1} \quad .$$

Il existe une infinité de métriques  $\hat{M}$  -et  $M_1$  appartient à cette infinité- telles que les axes principaux du triplet  $(X, \hat{M}, D_p)$  coïncident avec ceux du triplet  $(X, M, D_p)$ . Ces métriques de façon générale s'écrivent :

$$\hat{M} = \sum \alpha_i v_i v_i' \quad ; \quad \alpha_i \geq 0 \quad .$$

Les axes principaux sont obtenus dans le même ordre si et seulement si :

$$\lambda_i \alpha_i \geq \lambda_{i+1} \alpha_{i+1} \quad ; \quad i = 1, \dots, p-1 \quad .$$

Il existe, de façon symétrique, une infinité de matrices  $\hat{V}$  qui permettent,  $M$  étant fixé, de retrouver les axes principaux et cela dans le même ordre.

Ces remarques sur les classes de formes quadratiques  $\hat{M}$  et  $\hat{V}$  suggèrent déjà que bien des critères permettront de retrouver les axes principaux comme solutions d'un problème d'optimisation, si on ne garde des valeurs propres que l'ordre dans lequel elles se rangent.

### 1.2 Régression et analyse canonique

Bien des décompositions de formes quadratiques de type  $V = V_1 + V_2$ , que l'on rencontre en analyse de données, ne vérifient pas l'égalité :  $\text{rang } V = \text{rang } V_1 + \text{rang } V_2$  . Ces décompositions ne sont pas alors associées à une décomposition de l'espace  $E$  en sous-espaces  $V^{-1}$ -orthogonaux.



p variables  $x^j$  (matrice de variance V) et les q composantes factorielles. L'analyse fine de la variance expliquée  $V_1 = V - D$  peut donc s'effectuer en diagonalisant  $V_1$   $V_2^{-1} = (V - D)D^{-1}$  (optique de Fisher). Si on adopte les mêmes notations qu'en analyse en composantes principales, les q axes canoniques  $\Delta_{u_i}$  vérifient donc :

$$(V - D)D^{-1} u_i = \lambda_i u_i \quad ; \quad \|u_i\|_{D^{-1}} = \lambda_i$$

$$\text{et :} \quad V_1 = V - D = \sum u_i u_i'$$

Remarquons que les q vecteurs  $u_i$ , qui forment un système  $D^{-1}$ -orthonormé, forment aussi un système  $V^{-1}$ -orthogonal ; si  $\delta_{ii'}$  désigne le symbole de Kronecker :

$$V^{-1}(u_i, u_{i'}) = \frac{\lambda_i}{\lambda_i + 1} \delta_{ii'}$$

Mais en général, quand on postule l'existence d'un modèle à q facteurs, on ne connaît pas la décomposition correspondante de la variance totale. Pour l'obtenir on peut s'inspirer de la solution particulière proposée par Lawley (Lawley et Maxwell 1963) à la méthode du maximum de vraisemblance, solution identique, nous l'avons vu, à celle à laquelle on aboutit quand, mettant de côté toute considération statistique, on cherche à rendre maximum le déterminant de la matrice des corrélations résiduelles (Howe 1955) ; on s'appuie alors sur le système d'équations :

$$(5) \quad (V - D) D^{-1} u_i = \lambda_i u_i \quad ; \quad \|u_i\|_{D^{-1}} = \lambda_i$$

$$(6) \quad V_1 = \sum_{i=1}^q u_i u_i' \quad ; \quad V_2 = V - V_1$$

$$(7) \quad D = \text{diag } V_2 \quad .$$

Ce système, dans lequel V désigne l'estimateur sans biais de la matrice de variance, est résolu par itérations, en général sans trop de problèmes, à partir d'une première approximation de la matrice des "spécificités" D. Cette première approximation de D fait très souvent

référence à la "théorie des images" de Louis Guttman (L. Guttman 1953 et R.P. Mac Donald 1975) ; il en est ainsi, aussi, des méthodes qui permettent d'éviter les itérations.

L. Guttman a proposé, avec la théorie des images, d'étudier la structure des corrélations en effectuant les régressions entre chacune des variables et toutes les autres. Le tableau des données centré-réduit  $X$ , de matrice de corrélation  $R$ , se trouve alors décomposé en deux tableaux : le tableau expliqué par les régressions  $X_1$  ; appelé image partielle, et le tableau des résidus  $X_2$ , appelé anti-image partielle (on parle d'image et d'anti-image quand on se situe dans la population et non pas dans un échantillon).

Les éléments diagonaux de la matrice  $I - S^2$ , où :  $S^2 := [\text{diag } R^{-1}]^{-1}$ , étant égaux aux carrés des coefficients de corrélation multiple, le tableau des résidus s'obtient directement par la formule :

$$(8) \quad X_2 = S^2 R^{-1} X \quad .$$

Les matrices de variance, associées aux tableaux  $X_1$  et  $X_2$  respectivement, s'écrivent alors :

$$(9) \quad V_1 = (R - S^2)R^{-1}(R - S^2) = R - 2 S^2 + S^2 R^{-1} S^2$$

$$(10) \quad V_2 = S^2 R^{-1} S^2 \quad .$$

On a donc la relation :

$$(11) \quad V = R = V_1 + V_2 - 2 S^2 (R^{-1} S^2 - I) \quad .$$

On comprend mieux l'intérêt de la théorie des images quand on se remet en mémoire les questions posées par les disciples de Spearman au début de ce siècle. Compte tenu de l'importance accordée à l'interprétation des facteurs, il était normal que ceux-ci s'interrogent sur la stabilité des résultats fournis par les techniques mises au point. Cette stabilité dépend de l'influence du choix des variables sur la

structure mise en évidence. Le nombre  $p$  de variables choisies étant suffisamment grand, pour que le modèle soit valide, il faut qu'en particulier les propriétés suivantes soient vérifiées :

- . le minimum  $q$  de  $V_1$ , permettant d'obtenir théoriquement une matrice de variance  $V_2$  diagonale dans la décomposition  $R = V_1 + V_2$ , est indépendant de  $p$ .

- . la décomposition de  $V$  précédente est la seule de rang  $q$  de ce type (unicité).

- . le sous-espace engendré par les  $q$  composantes factorielles est indépendant de  $p$ .

D'après la troisième propriété : quel que soit le couple  $(x,y)$  de variables de l'univers considéré, la corrélation partielle entre  $x$  et  $y$  "sachant les composantes factorielles" doit être nulle. Si on dispose de variables "supplémentaires" on dispose donc d'un moyen simple pour juger de la validité du modèle factoriel.

L'univers des variables auquel il est fait référence est plus insaisissable que la population d'où sont tirés les individus : comment être sûr que les variables tirées respectent la structure de l'univers ? C'est-à-dire : permettant de les extraire, renferment-elles toute l'information sur les facteurs ? Supposer que sur cet univers un modèle à  $q$  facteurs s'applique rigoureusement est un acte de foi qui ne peut être conforté qu'en mettant le modèle à l'épreuve sur des jeux de variables de plus en plus grands. Aussi, mettant de côté le problème de l'unicité, qui semble mineur si  $q$  est pris petit devant  $p$ , la seule question à laquelle pouvait tenter de répondre le psychomètre d'alors était la suivante : les variables étant supposées tirées d'un univers, si le modèle à  $q$  facteurs s'applique rigoureusement sur cet univers, que dire des décompositions de la matrice de corrélation  $R$  quand on fait croître le nombre de variables tirées jusqu'à l'infini ?

M. Roff (1936) cité par R.P. Mac Donald (1975) a montré le premier que la part de variance d'une variable expliquée par les compo-

santes factorielles (communauté), sous l'hypothèse du modèle à  $q$  facteurs, était supérieure au carré du coefficient de corrélation multiple de cette variable avec l'ensemble des autres variables. Si on considère les deux décompositions d'une variable quelconque  $x$ , prise parmi les  $p$ , que fournissent respectivement théorie des images et modèle factoriel :

$$x = x_1 + e_1 \text{ (théorie des images)} ; x = x_2 + e_2 \text{ (modèle factoriel)}.$$

On peut écrire en effet :

$$\|e_1\|^2 = \|x - x_1\|^2 = \|x - x_2 + x_2 - x_1\|^2 = \|e_2 + x_2 - x_1\|^2 .$$

Le modèle factoriel exige l'orthogonalité entre  $e_2$  et  $x_2$  d'une part, et les  $(p-1)$  variables autres que  $x$  d'autre part ;  $e_2$  est donc orthogonal à la fois à  $x_2$  et  $x_1$  ; on a donc bien :

$$\|e_2\|^2 \leq \|e_2\|^2 + \|x_2 - x_1\|^2 = \|e_1\|^2 .$$

Le nombre  $q$  de facteurs restant fixé, L. Guttman a alors montré que, si l'on fait tendre le nombre de variables  $p$  vers l'infini, sous certaines conditions qui paraissent réalistes, l'inégalité précédente tend à devenir une égalité. S'il y a égalité la décomposition qu'offre la théorie des images coïncide avec la décomposition induite par le modèle factoriel ; aussi, sous l'hypothèse de l'existence d'un modèle de rang  $q$ , peut-on écrire, si  $p$  est pris suffisamment grand, que la matrice obtenue en sommant les matrices de variance associées au tableau image et au tableau anti-image est proche de  $R$ . L'équation (11) conduisant à l'identité :  $R^{-1} S^2 = I$ , montre alors que  $R^{-1}$  peut être considérée comme pratiquement diagonale, tout en ne pouvant,  $R$  ne l'étant pas, l'être tout à fait.

Si on reprend par exemple la matrice de corrélation :

$R = r jj' + (1-r) I$ , où toutes les corrélations sont égales à  $r$ , la matrice  $R^{-1}$  s'écrit :



$$R^{-1} = \alpha jj' + \beta I$$

avec :  $\alpha = -\frac{r}{(1-r)((p-1)r+1)}$  et  $\beta = \frac{1}{1-r}$

Il est clair ici, lorsque  $p$  tend vers l'infini, le coefficient hors diagonal  $\alpha$  tendant vers zéro et les éléments diagonaux tendant donc vers  $\beta$ , que la matrice  $R^{-1}$  va vers une forme diagonale. Mais il faut remarquer que la somme des carrés des  $p(p-1)$  cosinus construits à partir des éléments non diagonaux de  $R^{-1}$  tend alors, non pas vers zéro (critère classique de diagonalité), mais vers une limite finie toujours égale à 1. Si à la limite  $R^{-1}$  n'est pas strictement diagonale, elle l'est en tout cas considérablement plus que  $R$ .

Ce sont ces remarques sur le comportement à l'infini de la structure des corrélations qui ont conduit K. Joreskog (1963) puis H. Kaiser (1970) à proposer des stratégies qui se révèlent efficaces pour mener à bien les calculs en analyse factorielle au sens de Spearman.

La stratégie de Joreskog repose sur l'hypothèse que la matrice diagonale des spécificités  $D$ , limite des  $V_2$  dans les itérations impliquées par les équations (5), (6), (7), est du type  $t S^2$ , où  $t$  est un nombre compris entre 0 et 1 ; pour obtenir la décomposition cherchée de la matrice de corrélation,  $t$  n'étant pas connu a priori, on diagonalise la matrice  $R S^{-2}$  dont les  $(p-q)$  dernières valeurs propres sont égales à  $t$ , si le modèle est de rang  $q$  et si l'hypothèse de Joreskog se trouve vérifiée. Pratiquement  $q$  et  $t$  sont estimés simultanément ; le rang  $q$  est choisi de façon que les dernières valeurs propres de  $R S^{-2}$  puissent être considérées comme égales, leur moyenne fournissant une estimation de  $t$ . La méthode préconisée par Joreskog revient à effectuer l'analyse en composantes principales du triplet  $(X, M, D_p)$ , où la matrice associée à la métrique  $M$  est égale à  $\text{diag } V^{-1}$ , si  $X$  est simplement centré et à  $\text{diag } R^{-1} = S^{-2}$ , si  $X$  est de plus réduit. Cette métrique est appelée en France : métrique de Joreskog.

La stratégie de Kaiser est encore plus directement inspirée de la théorie des images. Si on suppose que la matrice de variance  $V_1$  de

la décomposition factorielle peut être assimilée à la matrice de variance du tableau image, la matrice  $V_2$  étant approchée par la matrice  $S^2$ , considérée comme une bonne approximation de la matrice de variance du tableau anti-image (si  $R^{-1}$  est suffisamment diagonale :  $S^2 \neq S^2 R^{-1} S^2$ ), il est logique de diagonaliser :  $(R - S^2)R^{-1}(R - S^2)S^{-2}$ , pour mener à bien l'analyse factorielle (optique de Fisher). Cette façon d'opérer proposée par Kaiser est en réalité très voisine de celle introduite par Joreskog ; en effet :

$$\text{si : } R S^{-2} u = \lambda u \text{ ; alors : } (R - S^2)R^{-1}(R - S^2)S^{-2} u = \left(\lambda + \frac{1}{\lambda} - 2\right) u \text{ .}$$

Les deux analyses conduisent donc aux mêmes vecteurs propres qui peuvent apparaître, il est vrai, dans des ordres différents.

C'est la méthode de Kaiser qui semble à l'heure actuelle avoir la préférence ; elle a en particulier une place de choix dans un logiciel comme le BMDP.

Il est à noter que le modèle de Spearman peut se généraliser au cas où les variables sont regroupées en paquets ; la matrice de variance des résidus est alors supposée, non plus diagonale, mais diagonale par blocs. Si on utilise en première approximation, comme précédemment, les résultats que fournit la théorie des images quand on travaille par blocs, on obtiendra sans difficulté la solution. L'anti-image  $X_2$  dans la théorie des images "par blocs" sera obtenue à l'aide d'une formule analogue à (8) :

$$X_2 = S_B^2 R^{-1} X \text{ ; avec : } S_B^2 = [\text{diag-bloc}(R^{-1})]^{-1}$$

où,  $R$  désignant la matrice de corrélation,  $\text{diag-bloc}(R^{-1})$  est la matrice diagonale "par blocs", issue de  $R^{-1}$ , correspondant à la répartition des variables en paquets.

De cette généralisation on peut déduire une façon d'envisager l'analyse factorielle sur variables qualitatives au sens de Spearman. Dans ce cas, si  $X$  est le tableau des indicatrices des modalités des

variables qualitatives, le rôle de  $R$  est joué, non par le tableau de Burt  $X D_p X'$  qui n'est pas inversible, mais par la matrice  $Z D_p Z'$  où  $Z$  se déduit de  $X$  par centrage et suppression d'une colonne par bloc associé à chaque variable. La diagonalisation de :  $S_b^{-2} R$  dans l'optique Joreskog, ou de :  $S_b^{-2}(R - S_b^{-2})R^{-1}(R - S_b^{-2})$  dans l'optique Kaiser, donne alors les facteurs dont l'image par  $Z$  fournit les composantes factorielles recherchées.

## 2. AJUSTEMENTS OPTIMAUX DE RANG FIXE INDUITS PAR LES INÉGALITÉS USUELLES ENTRE FORMES QUADRATIQUES

La régression et les analyses factorielles conduisent donc à la décomposition d'une forme quadratique semi-définie positive  $V$  en somme de deux formes quadratiques semi-définies positives  $V_1$  et  $V_2$ . Dans ces décompositions  $V_1$  peut être considérée comme un ajustement de  $V$ . Dans ce chapitre on s'intéressera, de façon générale, au meilleur ajustement de  $V$  par une forme quadratique inférieure à  $V$ ; l'espace vectoriel  $E$  étant supposé de dimension  $p$ , les formes quadratiques considérées seront assimilées à des applications de  $E^*$  dans  $E$ .

Deux types d'inégalités peuvent être exploitées pour comparer les formes quadratiques semi-définies positives, dont l'étude renvoie à la géométrie des ellipsoïdes : la première traduit l'inclusion des ellipsoïdes ; la seconde exploite les différences dans les longueurs de leurs grands axes.

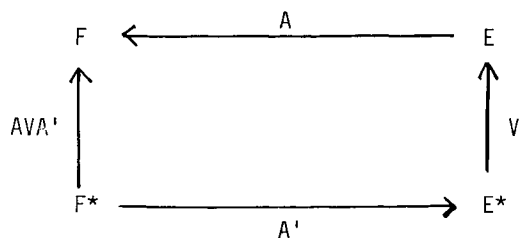
### 2.1 Inégalité basée sur l'inclusion des ellipsoïdes

$V_1$  et  $V_2$  sont deux formes quadratiques définies sur un même espace  $E^*$ .

Définition 1 : Ordre  $O_1$

$V_1$  est dite inférieure à  $V_2$ , au sens de l'ordre  $O_1$ , si la forme quadratique  $V_2 - V_1$  est semi-définie positive. On écrit alors :  $V_1 \leq V_2$ .

Si  $A$  est une application linéaire de  $E$  dans  $F$  :



On démontre sans difficulté le lemme suivant :

Lemme 1 : Si l'application linéaire  $A$  est injective les deux propriétés suivantes sont équivalentes :

- (a)  $V_1 \leq V_2$
- (b)  $A V_1 A' \leq A V_2 A'$  .

Remarque : si l'application  $A$  n'est pas injective on a seulement :

$$(a) \Rightarrow (b) .$$

### 2.1.1 Ajustement par en-dessous optimal de rang $q$

Le théorème bien connu suivant (Cailliez et Pagès 1976), pouvant être considéré comme à la base des théorèmes de Gauss-Markov et de Cochran, joue un rôle important en analyse des données.

Théorème 1 : Considérons la décomposition  $V = V_1 + V_2$  de la forme quadratique  $V$  ; notons  $A$  le projecteur  $V^{-1}$ -symétrique défini sur  $E$ , ayant même image que  $V_1$ . Si  $V$  est non dégénérée ou si  $V$  est semi-définie positive, les trois propositions suivantes sont équivalentes :

- (a)  $\text{rang } V = \text{rang } V_1 + \text{rang } V_2$
- (b)  $V_1 = A V A'$
- (c)  $V_2 = (I - A) V (I - A)'$  .

Sous la condition (a) du théorème 1,  $V_1$  et  $V_2$  sont semi-définies positives dès que la forme  $V = V_1 + V_2$  l'est ; le lemme suivant, corollaire de ce théorème, se démontre alors sans difficulté :

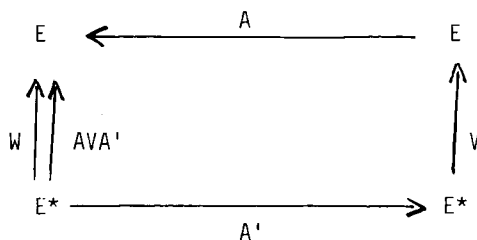
Lemme 2 : Soit  $V$  une forme quadratique définie positive ; soit  $B$  un projecteur quelconque défini sur  $E$  et  $A$  le projecteur  $V^{-1}$ -symétrique ayant même image que  $B$  ; on a les deux propriétés :

$$(a) \quad A V A' < B V B'$$

$$(b) \quad B V B' \leq V \Rightarrow B = A \quad .$$

Ce lemme précise donc en particulier, si  $V$  est une matrice de covariance, que la variance obtenue par projection est minimum (théorème de Gauss-Markov) quand le projecteur  $A$  est  $V^{-1}$ -symétrique (projection  $V^{-1}$ -orthogonale de l'ellipsoïde de dispersion).

Ce théorème suivant, qui découle des lemmes 1 et 2, permet de simplifier la recherche de la forme quadratique de rang  $q$ , inférieure à  $V$ , ajustant au mieux  $V$ .



Théorème 2 : Si la forme quadratique semi-définie positive  $W$ , de rang  $q$ , est inférieure à  $V$  (supposée de rang  $p$ ), il existe un projecteur unique  $A$ , de rang  $q$ , tel que :

$$W \leq A V A' < V$$

$A$  est le projecteur  $V^{-1}$  symétrique ayant même image que  $W$ .

La forme  $A V A'$ , qui fournit un ajustement de rang  $q$  "par en-dessous" maximal de  $V$  peut donc être appelé ajustement sous dominant, de rang  $q$ , de  $V$ .

Démonstration :

(1) Si  $B$  est un projecteur de même rang  $q$  que  $W$ , sous l'hypothèse  $W \leq V$ , on a l'équivalence :

$$W \leq B V B' \Leftrightarrow \text{image } W = \text{image } B .$$

en effet :

- si  $\text{image } W = \text{image } B$  , le projecteur  $B'$  ayant même noyau que  $W$  :

$$W(x) = W(B'x) = B W B'(x) \quad ; \quad \text{pour tout } x \text{ de } E^* .$$

On en déduit :  $W = B W B'$  ; et d'après le lemme 1 :

$$W < V \Rightarrow B W B' = W \leq B V B'$$

- si  $W \leq B V B'$  :

$\text{ker } B' \subset \text{Ker } W$  ; or  $\text{rang } B = \text{rang } W$  ; donc :  $\text{Ker } B' = \text{Ker } W$  .

(2) D'après le lemme 2, parmi les projecteurs de rang  $q$  , seuls les projecteurs  $V^{-1}$ -symétriques  $A$  vérifient la propriété :  $A V A' \leq V$  .

Si on considère le projecteur  $A$  ,  $V^{-1}$ -symétrique et de même image que  $W$  , on a donc, si  $W < V$  :

$$\text{d'après (a) : } \quad W \leq A V A' \quad ;$$

$$\text{d'après (b) : } \quad A V A' < V \quad ;$$

et  $A$  est le seul projecteur, de rang  $q$ , vérifiant ces deux propriétés.

La remarque :  $A V A' V^{-1} = A V V^{-1} A = A^2 = A$  , conduit à la proposition suivante comme corollaire du théorème 2 ; elle donne un moyen simple pour contrôler que l'on dispose d'un ajustement sous-dominant d'une forme quadratique  $V$  .

Proposition 1 :

Considérons la forme quadratique semi-définie positive  $W$  de rang  $q$  inférieure à  $V$  :  $W$  est un ajustement sous-dominant de rang  $q$  de  $V$  si et seulement si  $A = W V^{-1}$  est un projecteur  $V^{-1}$ -symétrique.

Le théorème 2 a une interprétation géométrique immédiate dans  $E$  :

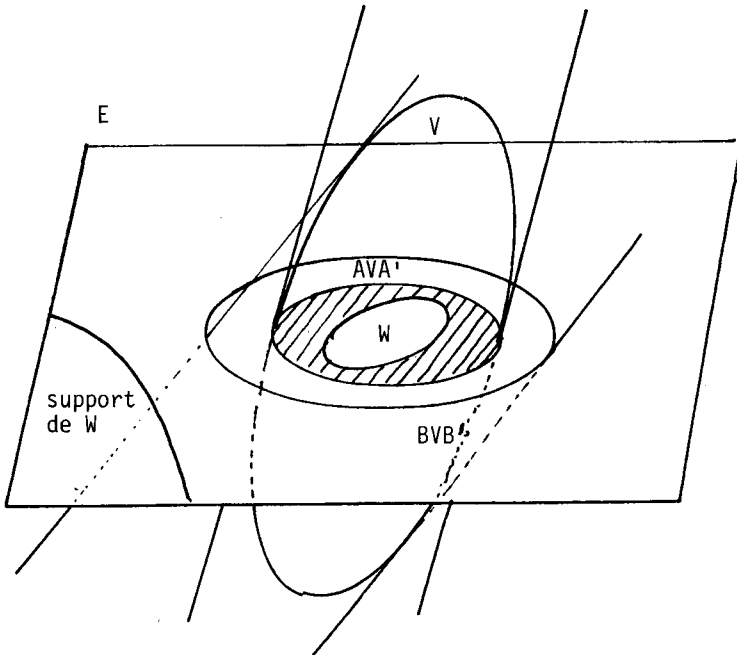


Figure 1

Représentation des ellipsoïdes de concentration associés à  $W$ ,  $AVA'$  et  $V$  .

Si on se réfère à la figure 1 :

- . l'identité des supports de  $W$ ,  $AVA'$  et  $BVB'$  traduit la propriété d'identité des images
- . les ellipsoïdes  $AVA'$  et  $BVB'$  sont obtenus par projection de l'ellipsoïde  $V$  (ombres portées)
- . la propriété "être inclus dans" pour un ellipsoïde traduit la propriété "être inférieure à" pour une forme quadratique.
- . la direction de projection qui permet d'obtenir l'ellipsoïde  $AVA'$  est définie par le supplémentaire  $V^{-1}$ -orthogonal du support de  $W$ ; cette direction coïncide avec celle des sous-espaces tangents à l'ellipsoïde  $V$  aux points d'intersection avec le support de  $W$ .

Remarque : à l'ajustement  $AVA'$  sous dominant de  $V$  correspond l'ajustement  $A'V^{-1}A$  sous dominant de  $V^{-1}$ .

2.1.2 Application à la régression

Le théorème 2 s'applique en particulier dans le cas de la régression (Cf. 1.2). Si on considère la décomposition :

$$V_{11} = V_{12} V_{22}^{-1} V_{21} + (V_{11} - V_{12} V_{22}^{-1} V_{21}) ;$$

l'ajustement  $V_{12} V_{22}^{-1} V_{21}$  de rang  $q$  (variables explicatives  $y^j$ ) de la matrice de variance  $V_{11}$  de rang  $p$  (variables à expliquer  $x^j$ ) vérifie :

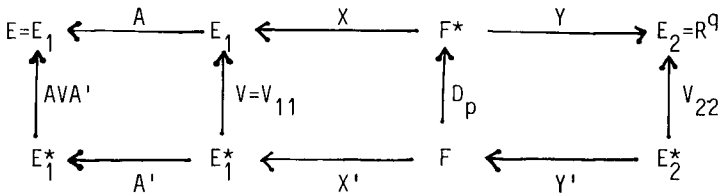
$$V_{12} V_{22}^{-1} V_{21} \leq V_{11} .$$

D'après ce qui précède il existe un projecteur unique  $A$ , de rang  $q$ , tel que :

$$V_{12} V_{22}^{-1} V_{21} \leq AV_{11}A' \leq V_{11} .$$

$A$  est le projecteur  $V_{11}^{-1}$ -symétrique ayant même image que  $V_{12}$  (image  $V_{12} = \text{image } V_{12} V_{22}^{-1} V_{21}$ ).

Rappelons, en dressant le schéma de dualité, quels sont les espaces considérés :



L'espace  $E$  coïncidant ici avec l'espace  $E_1$  de dimension  $p$  où les individus sont représentés par les vecteurs  $x_i$ , la proposition suivante précise ce qu'est la matrice  $AVA'$  ( $V = V_{11}$ ) :



Proposition 2 :

Si les vecteurs  $\hat{y}^j$  (tableau  $\hat{Y}$ ) désignent dans  $F$  les projections des variables  $y^j$  (tableau  $Y$ ) sur le sous-espace engendré par les variables  $x^j$  (tableau  $X$ ),  $AVA'$  n'est autre que la matrice de variance expliquée que l'on obtient dans la régression de  $X$  sur  $\hat{Y}$ .

Démonstration

$\hat{V} = V_{12} (V_{21} V_{11}^{-1} V_{12})^{-1} V_{21}$  est la matrice de variance expliquée par la régression sur  $Y$ ; on a :  $\hat{V} \leq V = V_{11}$ . L'application  $\hat{V} V^{-1}$  est le projecteur  $V^{-1}$ -symétrique  $A$  ayant même image que  $V_{12}$ ; on a donc d'après la proposition 1 :

$$V_{12} V_{22}^{-1} V_{21} \leq AVA' = \hat{V} < V .$$

On voit donc, en particulier, que reconstituer dans  $F$  par régression  $p$  variables  $x^j$  à partir de  $q \leq p$  combinaisons linéaires  $c^j = X'v_j$  de ces variables conduit à approcher par en-dessous de façon maximale la matrice de variance  $V$  par une forme quadratique  $V_1 = AVA' = \hat{V}$ . L'application  $A = V_1 V^{-1}$  et alors le projecteur  $V^{-1}$ -symétrique sur le sous-espace engendré par les vecteurs  $V v_j$  (image de  $V_{12}$ ). Cette opération correspond à la formule de reconstitution des données en analyse en composantes principales.

2.1.3 Application à l'estimation des scores en analyse factorielle au sens de Spearman

En analyse factorielle au sens de Spearman les composantes factorielles  $c^k$ , qui conduisent à l'approximation  $W$  de  $V$ , ne sont pas des combinaisons linéaires des variables  $x^j$ ; les covariances entre les composantes et les  $x^j$  (saturations) sont directement calculées sans passer par le calcul des coordonnées sur les axes factoriels (scores). On ne sait pas directement calculer ces scores et différentes manières de les approcher ont été proposées dans la littérature (Lawley et Maxwell 1963).

Se plaçant dans l'espace des variables  $F$ , reprenons les formules de reconstitution des variables dans le modèle de Spearman à  $q$  facteurs. A la décomposition :  $F = F_1 + F_2$ , que postule le modèle, correspond, si le sous-espace  $F_1$  de dimension  $q$  est muni de la base orthonormée des vecteurs  $c^k$ , les formules de reconstitution :

$$(12) x^j = \hat{x}^j + e^j = \sum_{k=1}^q u_k^j c^k + e^j ; \quad j = 1, \dots, p .$$

. les coordonnées de la projection  $\hat{x}^j$  (tableau  $\hat{X}$ ) de la variable  $x^j$  sur le sous-espace  $F_1$ , dans la base des  $c^k$ , sont les "saturations"  $u_k^j$ .

. les variables  $e^j$  représentent l'erreur faite dans la reconstitution des  $x^j$ ; elles admettent la matrice diagonale  $D$  pour matrice de covariance.

On pourra supposer par la suite que les vecteurs  $c^k$ , qui a priori sont définis à une transformation orthogonale près, coïncident avec les composantes principales normées à 1 du triplet  $(\hat{X}, D^{-1}, D_p)$ ; ce choix, qui correspond à ce que l'on appellera "la solution de Lawley" (Lawley et Maxwell 1963), est cohérent avec les équations introduites au paragraphe 1.3; il impose, dans l'espace des individus  $E$ , l'orthogonalité au sens de  $D^{-1}$  des vecteurs  $u_k$  de coordonnées  $u_k^j$ . Les carrés des normes des vecteurs  $u_k$  (axes canoniques) seront pris égaux aux valeurs propres  $\lambda_j$  de  $(V - D)D^{-1}$ .

Si on note  $U$  la matrice  $(p \times q)$  des coefficients de régression  $u_k^j$ , qui ne sont autres que les covariances entre les variables  $x^j$  et les composantes factorielles  $c^k$ , aux formules (12) correspond la décomposition de la matrice de variance :

$$V = UU' + D ;$$

dans cette décomposition :  $W = UU'$  est l'approximation par en-dessous de rang  $q$  qu'offre le modèle factoriel.

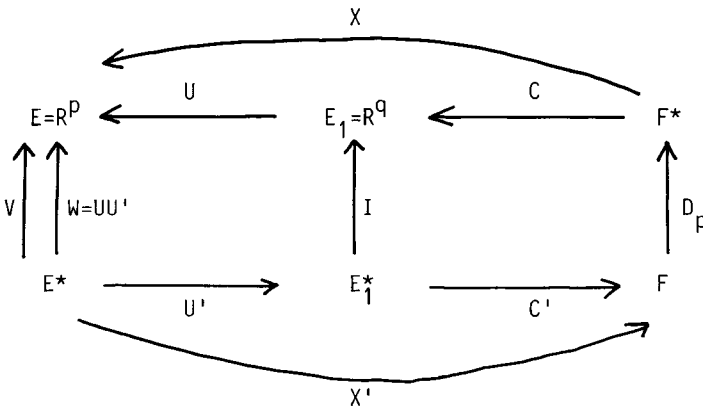
Transposées dans l'espace E où les individus sont représentés par les points  $x_i$ , les formules (12) s'écrivent :

$$(13) \quad x_i = U c_i + e_i \quad ; \quad i = 1, \dots, n .$$

Dans chacune de ces formules :

- . le vecteur  $c_i$  est le vecteur des scores de l'individu  $i$  .
- . le vecteur  $e_i$  représente l'erreur faite dans la reconstitution des  $x_i$  .

Si  $X$  ( $p \times n$ ) et  $C$  ( $q \times n$ ) sont les tableaux dont les colonnes sont les vecteurs  $x_i$  et  $c_i$  respectivement, on peut dresser le schéma de dualité :



L'approximation du tableau  $X$  fourni par le modèle factoriel est :  
 $\hat{X} = U C$  .

Opérant dans l'espace des caractères  $F$ , Thomson (Thomson 1951), pour estimer les scores, a proposé de remplacer les composantes factorielles  $c^k$  par leurs projections  $\hat{c}^k$  sur le sous-espace  $X'(E^*)$  engendré par les variables  $x^j$ . Aux approximations :

$$\hat{c}^k = X' V^{-1} X D_p c^k ,$$

correspond alors, puisque :  $U = X D_p C'$  n'est autre que la matrice

des covariances entre les variables  $x^j$  et  $c^k$ , le tableau :

$$(14) \quad \hat{C} = U'V^{-1}X \quad .$$

La base des vecteurs propres de  $(V-D)D^{-1}$  étant à la fois  $D^{-1}$  et  $V^{-1}$ -orthogonale, la matrice de variance  $U'V^{-1}U$  associée au tableau  $\hat{C}$  n'est autre, dans le cas de la solution de Lawley, que la matrice dia-

gonale des  $\frac{\lambda_i}{1+\lambda_i}$  :

$$(15) \quad U'V^{-1}U = D_{\lambda/1+\lambda} \quad (\text{Lawley}) \quad .$$

La reconstitution  $\hat{X}$  de  $X$  par régression des  $x^j$  sur les approximations  $\hat{c}^k$  conduit à approcher  $V$ , non plus par  $W$ , mais par l'ajustement sous dominant  $AVA'$ , de rang  $q$ , supérieur à  $W$  :

$$W \leqslant AVA' < V \quad ;$$

d'après ce qui précède  $A$  est le projecteur  $V^{-1}$ -symétrique qui a même image que  $W$ .

D'après (14) et (15) :  $\hat{U} = X D_p \hat{C}' (\hat{C} D_p \hat{C}')^{-1}$  matrice  $(p \times q)$  des coefficients de régression des  $x^j$  par rapport aux  $\hat{c}^k$ , s'écrit :

$$(16) \quad \hat{U} = U(U'V^{-1}U)^{-1} \quad .$$

On en déduit :

$$(17) \quad \hat{X} = \hat{U} \hat{C} = U(U'V^{-1}U)^{-1} U'V^{-1}X \quad .$$

L'image de  $W$  coïncidant avec celle de  $U$ , le projecteur :  $A = U(U'V^{-1}U)^{-1} U'V^{-1}$ , est bien le projecteur  $V^{-1}$ -symétrique sur l'image de  $W$  ; on a de plus :

$$(18) \quad AVA' = U(U'V^{-1}U)^{-1}U' \quad .$$

Au lieu d'opérer dans F comme Thomson, opérons dans E par "calibration" : si on considère la formule (13), il semble naturel d'estimer les coefficients inconnus  $c_i^j$  en projetant chacun des vecteurs  $x_i$  sur le sous-espace engendré par l'image de U . On effectuera cette régression, se souvenant du théorème de Gauss-Markov, en munissant l'espace des individus E de la métrique  $D^{-1}$  .

Remarquons qu'il est équivalent de projeter  $D^{-1}$  ou  $V^{-1}$ -orthogonalement sur l'image de U qui est engendrée par les q premiers vecteurs propres de  $(V-D)D^{-1}$  ; l'approximation de X obtenue par ces régressions dans E coïncide donc avec  $\hat{X}$  :

$$\hat{X} = A X = U \tilde{C} = \hat{U} \hat{C}$$

où  $\tilde{C}$  est l'approximation des scores obtenue dans cette deuxième approche :

$$(19) \quad \tilde{C} = (U'V^{-1}U)^{-1}U'V^{-1}X .$$

Les deux approximations des scores  $\hat{C}$  et  $\tilde{C}$  sont reliées par la formule

$$(20) \quad \tilde{C} = (U'V^{-1}U)^{-1} \hat{C} ;$$

la matrice de variance associée au tableau  $\tilde{C}$  est l'inverse de celle associée au tableau  $\hat{C}$  .

Les deux approches introduites pour estimer les scores aboutissent donc au même ajustement sous dominant AVA' , de rang q, de V ; engendrant les mêmes sous-espaces, les composantes factorielles  $\hat{c}^k$  et  $\tilde{c}^k$  sont homothétiques dans le cas de la solution de Lawley :

$$\tilde{C} = \frac{D_{1+\lambda}}{\lambda} \hat{C} \quad (\text{Lawley}) .$$

La deuxième approche présentée ici (calibration) est une variante de celle envisagée par Bartlett (Bartlett 1938) .

2.2 Inégalité basée sur la comparaison des spectres

L'ordre précédemment introduit a permis de définir une famille de formes quadratiques, constituée par les formes maximales, de rang  $q$  donné, inférieures à  $V$ . Ces ajustements sous dominants de  $V$  sont du type  $AVA'$  où  $A$  est opérateur  $V^{-1}$ -symétrique de rang  $q$ . Pour poursuivre la recherche de la forme  $W$  positive inférieure à  $V$  de rang fixé la plus proche de  $V$ , on introduit un préordre plus grossier que l'ordre précédent qui permet de comparer les  $AVA'$ .

$M$  désigne comme auparavant une métrique sur  $E$  et  $V_1$  et  $V_2$  deux formes quadratiques positives sur  $E^*$ . Les valeurs propres, rangées par valeurs décroissantes, de  $VM$  définissent son spectre noté : spectre  $(VM)$ .

Définition 2 :  $M$ -préordre  $O_2$

$V_1$  est dite  $M$ -inférieure à  $V_2$  s'il existe une transformation  $M$ -orthogonale  $U$  définie sur  $E$  telle que  $UV_1U'$  soit inférieure à  $V_2$  au sens de l'ordre  $O_1$ . On écrit alors  $V_1 \leq_M V_2$ .

Dire que  $V_1$  est  $M$ -inférieure à  $V_2$  revient à dire qu'il existe une  $M$ -isométrie  $U$  permettant d'amener l'ellipsoïde de concentration de  $V_1$  à l'intérieur de celui de  $V_2$ .

Quel que soit  $M$ , on a évidemment l'implication

$$V_1 \leq V_2 \Rightarrow V_1 \leq_M V_2 .$$

Elle implique que le  $M$ -préordre  $O_2$  est plus grossier que l'ordre  $O_1$ .

Le théorème qui suit montre en quoi le préordre  $O_2$  est basé sur la comparaison des longueurs, dans la métrique  $M$ , des axes des ellipsoïdes de concentration associés à  $V_1$  et  $V_2$ . Ces longueurs sont égales aux racines des valeurs propres de  $V_1M$  et  $V_2M$ , supposées rangées par valeurs décroissantes.

Théorème 3 :  $V_1$  est M-inférieure à  $V_2$  si et seulement si chaque valeur propre de  $V_1M$  est inférieure à la valeur propre de même rang de  $V_2M$ . On dit alors que le spectre de  $V_1M$  est inférieur au spectre de  $V_2M$  et on écrit:  $\text{spectre}(V_1M) \leq \text{Spectre}(V_2M)$ .

Sachant que si U est M-orthogonale :

$$\text{spectre}(UVU'M) = \text{spectre}(VM),$$

ce théorème découle immédiatement du lemme suivant, qui est énoncé en s'affranchissant de l'hypothèse de positivité de  $V_1$  et  $V_2$ . Ce lemme sera réutilisé dans la seconde partie du paragraphe sur les formes quadratiques quelconques.

Lemme 3 : Si,  $V_1$  et  $V_2$  étant deux formes quadratiques quelconques, on a l'inégalité :  $V_1 \leq V_2$ , chaque valeur propre de  $V_1M$  est inférieure à la valeur propre de même rang de  $V_2M$ .

Ce lemme est démontré dans Bellman (1960), cité par Okamoto (1969), à partir du théorème bien connu en min-max de Courant-Fisher (Rao 1973). Nous allons en donner une démonstration géométrique plus intuitive (A. Kobilinsky 1979).

Le sous-espace engendré par les  $i$  premiers vecteurs propres de  $V_1M$  et le sous-espace engendré par les  $p-i+1$  derniers vecteurs propres de  $V_2M$  ont nécessairement un vecteur  $u$  non nul en commun, car la somme de leurs dimensions  $(p+1)$  est strictement supérieure à la dimension de  $E$ . Parce que  $u$  est combinaison linéaire des  $i$  premiers vecteurs propres de  $V_1M$ , le coefficient de Rayleigh  $V_1(Mu, Mu)/M(u, u)$ , égal lorsque  $V_1$  est positif au moment d'inertie par rapport à l'hyperplan M-orthogonal à  $u$ , est supérieur à la  $i^{\text{ème}}$  valeur propre de  $V_1M$ . De façon analogue, on déduit de l'appartenance de  $u$  au sous-espace engendré par les  $p-i+1$  derniers vecteurs propres de  $V_2M$  que  $V_2(Mu, Mu)/M(u, u)$  est inférieur à la  $i^{\text{ème}}$  valeur propre de  $V_2M$ . L'inégalité entre les deux valeurs propres résulte alors,  $V_1$  étant inférieur à  $V_2$  au sens de l'ordre  $O_1$ , de l'inégalité :

$$V_1(Mu, Mu)/M(u, u) \leq V_2(Mu, Mu)/M(u, u) .$$

Cet argument géométrique est le même que celui utilisé dans Cailliez et Pagès (1976), pour démontrer que les sous-espaces principaux peuvent être construits pas à pas.

La proposition suivante complète le lemme 3 en cas d'égalité des valeurs propres :

Proposition 3 :

Si  $V_1M$  et  $V_2M$  ont même trace, donc en particulier s'ils ont même spectre, l'inégalité  $V_1 \leq V_2$  implique :  $V_1 = V_2$  .

Démonstration :

Si  $V_1 \leq V_2$  , les valeurs propres de  $(V_2 - V_1)M$  sont positives. Comme leur somme  $\text{tr}((V_2 - V_1)M)$  est nulle, elles sont toutes nulles et l'on a :  $V_2 - V_1 = 0$  .

2.2.1 Ajustement par en-dessous optimal de rang q d'une forme positive

Le lemme 3 montre en particulier que les valeurs propres de  $AVA'M$  , où  $AVA'$  est un ajustement sous dominant de  $V$  , sont inférieures aux valeurs propres de même rang de  $VM$  . Le lemme suivant permet d'introduire les théorèmes donnant les  $AVA'$  sous dominants de rang q, qui sont soit M-maximaux, c'est-à-dire maximaux au sens du M-préordre  $O_2$  , soit minimaux pour ce même préordre.

Lemme 4 : Si B est le projecteur M-symétrique ayant même image que le projecteur  $V^{-1}$ -symétrique A , on a :

$$\begin{aligned} \text{spectre } (AVA'M) &\leq \text{spectre } (BVB'M) \leq \text{spectre } (VM) \\ \text{spectre } (A'V^{-1}A M^{-1}) &\leq \text{spectre } (B'V^{-1}B M^{-1}) \leq \text{spectre } (V^{-1} M^{-1}) \end{aligned}$$

En effet, d'après le lemme 2 (Théorème de Gauss-Markov) :

$$AVA' \leq BVB' .$$



On a donc d'après le lemme 3 :

$$\text{spectre } (AVA'M) \leq \text{spectre } (BVB'M) .$$

On a de plus :

$$AVA' < V \quad ; \quad B'MB < M \quad ;$$

de la deuxième inégalité, on déduit en utilisant la commutativité du spectre :

$$\text{spectre } (BVB'M) = \text{spectre } (B'MBV) \leq \text{spectre } (MV) = \text{spectre } (VM) .$$

La première série d'inégalités énoncée dans le lemme est donc démontrée ; la seconde s'obtient en remplaçant  $V$  et  $M$  par leurs inverses et  $A$  et  $B$  par leurs transposés.

Le théorème suivant donne les  $AVA'$  sous-dominants  $M$ -maximaux de rang  $q$  :

Théorème 4 : La forme  $AVA'$  sous dominante de rang  $q$  est maximale pour le préordre  $O_2$  si et seulement si  $A$  est le projecteur  $V^{-1}$ -symétrique sur l'un des sous-espaces principaux de dimension  $q$  . Les valeurs propres non nulles de  $AVA'M$  coïncident alors avec les  $q$  premières valeurs propres de  $VM$ . La forme  $AVA'$  n'est pas unique s'il y a égalité des  $q$  et  $(q+1)^{\text{ième}}$  valeurs propres.

Pour démontrer le théorème 4, rappelons que la propriété d'être à la fois  $M$  et  $V^{-1}$ -symétrique est caractéristique des projecteurs associés à des décompositions en somme directe de sous-espaces engendrés par des axes principaux.

Si  $A$  est projecteur  $V^{-1}$ -symétrique sur un sous-espace principal, il est donc  $M$ -symétrique et les  $q$  valeurs propres non nulles de  $AVA'M$  coïncident avec les  $q$  premières valeurs propres de  $VM$  . D'après le lemme 4, le spectre de  $AVA'M$  est alors maximal.

Réciproquement, si  $AVA'$  est  $M$ -maximal, les valeurs propres non nulles de  $AVA'M$  sont égales aux  $q$  premières valeurs propres de  $VM$ . Par suite, si  $B$  désigne le projecteur  $M$ -symétrique de même image que  $A$ , on a simultanément :

$$\text{spectre}(AVA'M) = \text{spectre}(BVB'M)$$

$$AVA' \leq BVB'$$

et d'après la proposition 3 :  $AVA' = BVB'$ .

D'après le théorème de Gauss-Markov, de l'inégalité :

$$AVA' = BVB' < V, \quad \text{on déduit : } A = B.$$

$A$  est donc bien, étant à la fois  $M$  et  $V^{-1}$ -symétrique le  $V^{-1}$ -projecteur sur un sous-espace principal de dimension  $q$ .

Le théorème 4 admet le corollaire :

Corollaire :

Si  $W$  est une forme quadratique positive de rang  $q$ , on a l'implication :

$$W \leq V \Rightarrow W \underset{M}{\leq} AVA' \leq V$$

où  $A$  désigne le  $M$ -projecteur sur le sous-espace principal de dimension  $q$ .

Pour trouver les  $AVA'$  minimaux au sens du  $M$  préordre  $O_2$ , nous allons nous appuyer sur le lemme 4 et le théorème 4. Pour exploiter la deuxième série d'inégalités du lemme 4, remarquons que dans ce lemme :

- . les valeurs propres de  $V^{-1}M^{-1}$  sont les inverses de celles de  $VM$

- .  $AVA'M = AVM$
- . spectre  $(B'V^{-1}BM^{-1}) = \text{spectre}(BM^{-1}B'V^{-1})$
- .  $B M^{-1} B'V^{-1} = B M^{-1} V^{-1}$  .

Le lemme suivant fournit une inverse généralisée particulière pour AVM :

Lemme 5 : Si B est le projecteur M-symétrique ayant même image que le projecteur  $V^{-1}$ -symétrique A , les restrictions sur l'image de A des applications AVM et  $B M^{-1} V^{-1}$  sont inverses l'une de l'autre. Les valeurs propres non nulles de l'une sont les inverses des valeurs propres non nulles de l'autre.

En effet :

$$A V M B M^{-1} V^{-1} A = A V B' V^{-1} A = V A' B' V^{-1} A = V A' V^{-1} A = A$$

et les vecteurs propres associés à des valeurs propres non nulles de AVM et  $B M^{-1} V^{-1}$  sont dans l'image de A .

Théorème 5 : La forme  $AVA'$  sous dominante de rang q est minimale pour le préordre  $\mathcal{O}_2$  si et seulement si A est le projecteur  $V^{-1}$ -symétrique sur l'un des sous-espaces engendrés par q vecteurs propres associés aux q dernières valeurs propres de VM .

Démonstration :

Le spectre de  $AVA'M$  est minimal quand le spectre de  $B'V^{-1}BM^{-1}$  est maximal, c'est-à-dire quand l'image de B' est un sous-espace principal de dimension q de  $V^{-1}M^{-1}$  . Ceci se produit lorsque l'image de A, égale à celle de B , est un sous-espace engendré par des vecteurs propres de VM associés aux q dernières valeurs propres.

Le théorème 5 permet de poser le problème de l'approximation de la forme quadratique V par une forme inférieure de rang q de la façon suivante :

Corollaire :

La forme quadratique  $W$  de rang  $q$  inférieure à  $V$  rendant minimale la différence  $V - W$  au sens du  $M$ -préordre  $0_2$ , coïncide avec la forme  $AVA'$ , où  $A$  désigne le  $M$ -projecteur sur le sous-espace principal de dimension  $q$ .

2.2.2 Le cas d'une forme quelconque

La forme  $W$  définie dans le corollaire précédent est-elle encore optimale au sens du spectre lorsqu'on s'affranchit de la contrainte  $W \leq V$ ? Pour répondre à cette question, on va considérer le problème un peu plus général de l'approximation d'une forme quadratique  $V$  inversible quelconque par une forme quadratique  $W$ , telle que  $WM$  ait au plus  $q^+$  valeurs propres strictement positives et  $q^-$  strictement négatives. Pour comparer des formes quelconques, nous allons d'abord introduire un ordre  $0_2'$ , généralisation naturelle de l'ordre  $0_2$ .

Si  $\lambda_1, \dots, \lambda_p$  sont les valeurs propres associées aux vecteurs propres  $u_1, \dots, u_p$  de  $VM$ , on a :  $V = \sum \lambda_i u_i u_i'$ . On en déduit la décomposition suivante de  $V$  :  $V = V^+ - V^-$

$$\text{avec } V^+ = \sum_{\lambda_i > 0} \lambda_i u_i u_i' \quad \text{et} \quad V^- = - \sum_{\lambda_i < 0} \lambda_i u_i u_i' .$$

$V^+$  et  $V^-$ , qui sont toutes deux positives, seront dites parties positives et négatives de  $V$  pour la métrique  $M$ .

Définition 3 :  $M$ -préordre  $0_2'$

$V_1$  et  $V_2$  étant deux formes quadratiques, on dit que  $V_1$  est  $M$ -inférieure à  $V_2$  au sens du  $M$ -préordre  $0_2'$  et on écrit :

$$V_1 \underset{M}{\leq} V_2 ,$$

si l'on a simultanément :  $V_1^+ \underset{M}{\leq} V_2^+$  et  $V_1^- \underset{M}{\leq} V_2^-$ , où  $V_1^+$  et  $V_1^-$

sont les parties positives et négatives de  $V$  pour la métrique  $M$ .

On a alors la proposition ci-dessous, analogue du corollaire du théorème 5 pour les formes quadratiques quelconques :

Proposition 4 :

La forme quadratique  $W$ , telle que  $W^+$  soit de rang  $q^+$  et  $W^-$  de rang  $q^-$ , qui rend minimale  $V-W$  au sens du  $M$ -préordre  $O_2'$  coïncide avec la forme  $AVA'$ , où  $A$  désigne l'opérateur de  $V^-$  projection sur l'espace engendré par la réunion des vecteurs propres de  $VM$  correspondant aux  $q^+$  plus grandes et  $q^-$  plus petites valeurs propres.

Démonstration :

Notons  $R$  la différence  $V-W$ ,  $R^+$  et  $R^-$  ses parties positives et négatives pour la métrique  $M$ .

$$\text{On a : } V = W + R = W^+ - W^- + R^+ - R^- .$$

Si l'on pose :  $V_1 = V + W^- + R^-$ , on a  $V_1 = W^+ + R^+$ .

La forme quadratique  $W^+$  de rang inférieure à  $q^+$  est donc inférieure à  $V_1$ . Le corollaire du théorème 5 montre alors que :

$$\text{spectre}(R^+ M) = \text{spectre}((V_1 - W^+)M) \leq \text{spectre}((V_1 - A_1 V_1 A_1') M) ,$$

où  $A_1$  est le  $V_1^{-1}$  projecteur sur le sous-espace principal de  $V_1 M$  de dimension  $q^+$ . En terme de valeurs propres, si l'on note  $\lambda_1(U), \dots, \lambda_p(U)$  les valeurs propres ordonnées par valeurs décroissantes d'un opérateur  $U$ , cette inégalité entre spectres s'écrit :

$$\lambda_i(R^+ M) \geq \lambda_{i+q^+}(V_1 M) .$$

Par ailleurs,  $V_1$  est supérieur à  $V$  au sens de l'ordre  $O_1$  : d'après le théorème 3, on a donc :

$$\lambda_{i+q^+}(V_1 M) \geq \lambda_{i+q^+}(VM) .$$

En combinant les deux inégalités précédentes, on déduit :

$$(V - W)^+ = R^+ \underset{M}{\geq} (V - AVA')^+$$

où A est le  $V^{-1}$  projecteur sur le sous-espace propre de dimension  $q^+$  de VM . On montre de même que  $(V - W)^- = V_r^- \underset{M}{\geq} (V - AVA')^-$  .

BIBLIOGRAPHIE

- [ 1 ] M.S. BARTLETT (1938). Methods of estimating mental factors. Nature, 141, p. 609-610.
- [ 2 ] R. BELLMAN (1960). Introduction to matrix analysis. Mac Graw Hill, New-York.
- [ 3 ] F. CAILLIEZ, J.P. PAGES (1976). Introduction à l'analyse des données. SMASH (9 rue Duban 75016 PARIS).
- [ 4 ] L. GUTTMAN (1953). Image theory for the structure of quantitative variate. Psychometrika, vol. 18, n°4, décembre 1953, p. 277-296.
- [ 5 ] H. HOTELLING (1933). Analysis of a complex of statistical variables into principal components. The journal of educational psychology, vol. 24, p. 417-441 et 498-520.
- [ 6 ] W.G. HOWE (1955). Some contributions to factor analysis. USAEC Rep. ORNL - 1919.
- [ 7 ] K.G. JORESKOG (1963). Statistical estimation in factor analysis : a new technique and its foundations Almqvist and Wiksell, Stockholm.
- [ 8 ] H.F. KAISER (1970). A second generation little Jiffy. Psychometrika, vol. 35, n°4, décembre 1970; p. 401-415.
- [ 9 ] A. KOBILINSKY (1979). Ordre entre formes quadratiques. Application à l'optimalité de sous-espaces en analyse des données. Revue de statistique appliquée, vol. XXVII, n° 1, p. 45-54.
- [ 10 ] D.N. LAWLEY, A.E. MAXWELL (1963). Factor analysis as a statistical method. London Butterworths.
- [ 11 ] R.P. MAC DONALD (1975). Descriptive axioms for common factor theory, image theory and component theory. Psychometrika, vol. 40, n° 2, juin 1975, p. 137-152.
- [ 12 ] M. OKAMOTO (1969). Optimality of principal components. Multivariate Analysis n° 2, P.R. Krishnaiah ed., p. 673-685, Academic Press, New-York.
- [ 13 ] K. PEARSON (1901). On lines and planes of closest fit to systems of points in space. Phil. Mag. n° 2 (6è serie) p. 557-572.

- [14] C.R. RAO (1973). Linear statistical inference and its applications John Wiley, New-York.
- [15] M. ROFF (1936). Some properties of the communality in multiple factor theory. Psychometrika, vol. 1, p. 1-6.
- [16] C. SPEARMAN (1904). General intelligence objectively determined and measured. American Journal of Psychology, vol. 15, p. 201-292.
- [17] G.H. THOMSON (1951). The factorial analysis of human ability. London University Press.
- [18] L. THURSTONE (1947). Multiple factor analysis. University of Chicago Press.