

Zeitschrift:	Schweizerische Zeitschrift für Soziologie = Revue suisse de sociologie = Swiss journal of sociology
Herausgeber:	Schweizerische Gesellschaft für Soziologie
Band:	2 (1976)
Heft:	3
Artikel:	Une nouvelle technique d'exploration statistique : l'analyse simultanée des variables et des unités
Autor:	Schellhorn, Jean-Pierre
DOI:	https://doi.org/10.5169/seals-815172

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 29.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Une nouvelle technique d'exploration statistique:
L'ANALYSE SIMULTANEE DES VARIABLES ET DES UNITES.

Jean-Pierre Schellhorn

RESUME

L'analyse en composantes principales (ACP), l'analyse des corrélations canoniques (ACC) et l'analyse factorielle des correspondances (AFC) sont de plus en plus utilisées en sciences sociales. Après avoir indiqué que ces trois techniques sont des cas spéciaux d'une seule et même méthode, l'analyse par transformations simultanées, cet article présente en grand détail un quatrième cas spécial de la méthode générale.

Partant de l'unique décomposition singulière que possède toute matrice de données (rectangulaire), on construit de ce fait une procédure "universelle". La représentation graphique associée est étudiée de façon approfondie, car c'est elle qui détermine l'intérêt pratique de cette nouvelle technique exploratoire.

Cet article est écrit pour un lecteur qui possède le bagage mathématique usuel en sciences sociales. Le résumé mathématique, dans l'appendice, complète l'exposé pour ceux qui aimeraient faire exécuter les calculs et graphiques nécessaires par un centre de calcul.

ZUSAMMENFASSUNG

Hauptkomponentenanalyse (ACP), kanonische Korrelationsrechnung (ACC) und Korrespondenzenanalyse (AFC) werden als Spezialfälle einer einzigen Methode, der sogenannten "Analyse durch gleichzeitige Transformationen" eingeführt. Der Hauptteil des Artikels wird einem vierten, neuen Spezialfall gewidmet, der auf der singulären Dekomposition einer Matrix entstammt. Da diese Dekomposition für jede (auch nicht Quadrat-)Matrix eindeutig definiert ist, ist diese neue Datenanalyse universell anwendbar. Um die Methode wirklich praktisch zu machen, wird auf die naheliegende graphische Darstellung grosses Gewicht gelegt. Im Anhang findet der Datenverarbeiter die vollständige Sammlung von mathematischen Formeln.

LA METHODE DES TRANSFORMATIONS SIMULTANÉES

- Pour partir d'un exemple aussi simple que possible, supposons que nous ayons à analyser les relations entre cinq variables qui se répartissent naturellement en deux groupes; appelons (x_1, x_2, x_3) les variables du premier groupe et (y_1, y_2) les variables du second groupe. Supposons aussi, pour fixer les idées, que le coefficient utilisé pour mesurer l'intensité de l'association entre deux quelconques de ces cinq variables soit leur covariance.

La matrice P des covariances a la forme suivante:

$$\begin{array}{c}
 \begin{matrix} & x_1 & x_2 & x_3 & y_1 & y_2 \end{matrix} \\
 P = \left[\begin{array}{cc|cc|cc}
 x_1 & b_{11} & b_{12} & b_{13} & f_{11} & f_{12} \\
 x_2 & b_{12} & b_{22} & b_{23} & f_{21} & f_{22} \\
 x_3 & b_{13} & b_{23} & b_{33} & f_{31} & f_{32} \\
 \hline
 y_1 & f_{11} & f_{21} & f_{31} & h_{11} & h_{12} \\
 y_2 & f_{12} & f_{22} & f_{32} & h_{12} & h_{22}
 \end{array} \right] = \begin{array}{|c|c|} \hline & B & F \\ \hline & & \\ \hline F' & & H \\ \hline & & \end{array}
 \end{array}$$

Par exemple: b_{22} = variance de x_2
 h_{11} = variance de y_1
 f_{21} = covariance de x_2 et y_1

La seule information disponible se réduit essentiellement à cette matrice P. Pour analyser les relations entre les deux groupes de variables, il convient d'y substituer de nouvelles variables qui fassent apparaître, sous une forme plus simple ou plus directement interprétable, la "position relative" de ces deux groupes. C'est ce que l'on fait d'une manière un peu particulière en ACC (voir le paragraphe suivant), et que l'on peut se fixer comme objectif plus général (cf. le paragraphe 3).

- Dans l'analyse en variables canoniques (ACC), on remplace les cinq variables initiales $(x_1, x_2, x_3; y_1, y_2)$ par de nouvelles variables, les deux paires de "variables canoniques" (ξ_1, η_1) , (ξ_2, η_2) et une cinquième variable ξ_3 dont la définition importe peu ici (cf. Schellhorn, 1976).

La matrice II des covariances des nouvelles variables a la forme suivante:

$$\Pi = \begin{array}{c} \begin{array}{ccccc} \xi_1 & \xi_2 & \xi_3 & \eta_1 & \eta_2 \end{array} \\ \begin{array}{c|cc|cc} \xi_1 & 1 & 0 & 0 & s_1 & 0 \\ \xi_2 & 0 & 1 & 0 & 0 & s_2 \\ \hline \xi_3 & 0 & 0 & 1 & 0 & 0 \end{array} \\ \hline \begin{array}{c|c|c} \eta_1 & s_1 & 0 \\ \hline \eta_2 & 0 & s_2 \end{array} \end{array} = \begin{array}{c} \begin{array}{c|c} I_3 & \Delta \\ \hline \Delta & \begin{array}{c|c} \square & \square \\ \square & I_2 \end{array} \end{array} \end{array}$$

Par exemple: covariance de ξ_1 et ξ_2 = 0
covariance de η_1 et η_2 = 0
variance de ξ_i = 1
variance de η_j = 1
covariance de ξ_1 et η_1 = s_1
covariance de ξ_1 et η_2 = 0

Les nombres positifs s_1 et s_2 sont appelés les deux coefficients de corrélation canonique entre le groupe (x_1, x_2, x_3) et le groupe (y_1, y_2) .

Nous exprimerons ce résultat en disant que le but de l'ACC est de transformer simultanément

$$(x_1, x_2, x_3) \text{ en } (x_1^*, x_2^*, x_3^*) \\ (y_1, y_2) \text{ en } (y_1^*, y_2^*)$$

afin de remplacer la structure compliquée (car quelconque)

$$P = \begin{array}{c} \begin{array}{c|c} B & F \\ \hline F' & H \end{array} \end{array} \text{ par la structure simple } \Pi = \begin{array}{c} \begin{array}{c|c} I_3 & \Delta \\ \hline \Delta & \begin{array}{c|c} \square & \square \\ \square & I_2 \end{array} \end{array} \end{array}.$$

Le résultat mathématique intéressant est qu'une paire de transformations simultanées (essentiellement) unique résout ce problème; c'est la paire $(x_1^* = \xi_1, x_2^* = \xi_2, x_3^* = \xi_3)$ et $(y_1^* = \eta_1, y_2^* = \eta_2)$.

3. Passons au principe général, énoncé sans démonstration (on trouvera celle-ci dans mon étude citée plus haut).

Si deux groupes de variables (x_1, \dots, x_p) et (y_1, \dots, y_q) possèdent une matrice de covariances (ou une matrice de corrélation, ou une matrice de coefficients d'association) de

forme générale

	$x_1 \dots x_p$	$y_1 \dots y_q$
x_1	B	F
\dots		
x_p		
y_1		H
\dots		
y_q		

on peut trouver deux paires de transformations simultanées

$$(x_1, \dots, x_p) \rightarrow (x_1^*, \dots, x_p^*) \text{ (respectivement } (\tilde{x}_1, \dots, \tilde{x}_p)) \\ (y_1, \dots, y_q) \rightarrow (y_1^*, \dots, y_q^*) \text{ (respectivement } (\tilde{y}_1, \dots, \tilde{y}_q))$$

telles que les matrices correspondantes des x_1^*, \dots, y_q^* (resp. des $\tilde{x}_1, \dots, \tilde{y}_q$) soient de forme

	$x_1^* \dots x_p^* \quad y_1^* \dots y_q^*$	$\tilde{x}_1 \dots \tilde{x}_p \quad \tilde{y}_1 \dots \tilde{y}_q$
x_1^*	α	β
\dots		
x_p^*		
y_1^*		
\dots		
y_q^*	γ	

où α et γ (resp. β) peuvent être fixés a priori, et où F^* (resp. $\tilde{\beta}$ et $\tilde{\gamma}$) ont des structures précisées dans l'appendice à cet article; la seule restriction est essentiellement que le rang des matrices α , γ et β (notion définie ci-dessous) ne soit pas supérieur au rang de la matrice B , H et F respectivement.

On peut évidemment atteindre au moyen de la même paire de transformations simultanées le double objectif α , γ et β si ces trois matrices sont "compatibles".

4. Dans la présentation usuelle de l'analyse des corrélations canoniques, on détermine séquentiellement les paires (ξ_1, η_1) , (ξ_2, η_2) , etc. de variables canoniques par des propriétés de corrélation nulle (entre ξ_i et ξ_k , entre η_i et η_k , entre ξ_i et η_k) et de corrélation maximum (entre ξ_k et η_k); c'est d'ailleurs cette corrélation s_k entre ξ_k et η_k que l'on appelle la k ème corrélation canonique.

La définition simultanée (et non séquentielle) de paires (ξ_k, η_k) , comme elle a été introduite au paragraphe 2, est connue des spécialistes depuis plusieurs années. Il est regrettable qu'elle n'ait pas acquis droit de cité dans les manuels de statistique, car elle me semble plus intuitive que la définition de type séquentiel.

L'ACC peut donc être considérée comme le cas spécial de la méthode générale du paragraphe 3, où les objectifs $\alpha \equiv I_p$, $\gamma \equiv I_q$ et $\beta \equiv \begin{array}{|c|c|} \hline A & O \\ \hline O & O \\ \hline \end{array}$ sont compatibles.

5. L'analyse factorielle des correspondances (AFC) sert à analyser des tables de contingence, des fonctions de probabilité bidimensionnelles ou, plus généralement, n'importe quelle matrice (rectangulaire) C de nombres non négatifs c_{ij} . On peut la considérer comme l'analyse des corrélations canoniques entre variables dichotomiques (cf., par exemple, Escoufier, 1971; ou Hill, 1974).

Si L (respectivement K) désigne la matrice diagonale dont les éléments diagonaux sont les totaux de lignes (resp. de colonnes) de la matrice C , alors la matrice

$\begin{array}{|c|c|} \hline L & C \\ \hline C' & K \\ \hline \end{array}$, en AFC, joue le rôle de la matrice $\begin{array}{|c|c|} \hline B & F \\ \hline F' & H \\ \hline \end{array}$ en ACC.

C'est donc bien un cas particulier de la méthode générale du paragraphe 3.

6. L'analyse en composantes principales (ACP) sert à simplifier la description d'un seul groupe de variables (x_1, \dots, x_p) . On peut cependant la présenter comme l'ACC.

Appelons B la matrice des variances et covariances de variables x_1, \dots, x_p ; c'est une matrice (symétrique et définie positive) quelconque.

Il est facile de montrer que si $p = q$ et si le groupe de variables (y_1, \dots, y_p) n'est qu'une "copie" du groupe (x_1, \dots, x_p) , alors l'ACP sert à remplacer la structure compli-

quée $P = \begin{bmatrix} B & B \\ B & B \end{bmatrix}$ par la structure simple $\Pi = \begin{bmatrix} V & W \\ W' & B' \end{bmatrix}$, où W est une matrice à lignes non-corrélées deux à deux et où la matrice V est diagonale. V contient les variances des nouvelles variables (x_1^*, \dots, x_p^*) , variables que l'on appelle les composantes principales du groupe (x_1, \dots, x_p) .

Cette définition de type simultané me semble plus intéressante que la définition usuelle, où x_1^*, \dots, x_p^* sont définies séquentiellement par des propriétés de corrélation nulle (entre x_i^* et x_k^*) et de variance maximum (pour chaque x_k^*).

En conclusion, l'ACP est le cas particulier de la méthode générale des transformations simultanées où $B \equiv F \equiv H$, et où l'objectif est $\mathcal{X} \equiv V$ et $\mathcal{Y} \equiv B$.

7. Dans le reste de cet article, nous allons développer un cas particulier qui nous semble encore plus important que les précédents, et qui n'a été étudié que récemment (Gabriel, 1971; Schellhorn, 1976).

Il s'agit du cas d'une matrice quelconque

$$A = \begin{array}{|c|c|c|} \hline a_{11} & \dots & a_{1m} \\ \hline \dots & \dots & \dots \\ \hline a_{nl} & \dots & a_{nm} \\ \hline \end{array} = \begin{array}{|c|} \hline l_1' \\ \hline \dots \\ \hline l_n' \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline c_1 & \dots & c_m \\ \hline \end{array}$$

à n lignes l_1', \dots, l_n' et m colonnes c_1, \dots, c_m .

Pour ne pas multiplier les interprétations possibles des formules mathématiques, nous nous occuperons du cas le plus courant où chaque colonne c_i représente n mesures d'une variable, et où chaque ligne l_j représentent une unité statistique où l'on a mesuré les m variables.

Dans la section intitulée "La décomposition singulière d'une matrice", nous rappellerons que A peut être représentée de façon unique sous la forme d'une somme de matrices de rang un aux propriétés intéressant l'analyse de données. Cette propriété - encore trop peu connue et trop peu utilisée - permet de construire une paire de transformations simultanées

$$A = \begin{bmatrix} l_1 & \dots & l_n \end{bmatrix}' \longrightarrow U = \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \quad ("effets des unités")$$

$$A' = \begin{bmatrix} c_1 & \dots & c_m \end{bmatrix}' \longrightarrow V' = \begin{bmatrix} v_1 & \dots & v_m \end{bmatrix}' \quad ("effets des variables")$$

telle que

$$P = \begin{array}{|c|c|} \hline AA' & A^{(2)} \\ \hline A^{(2)'} & A'A \\ \hline \end{array} \text{ soit transformée en } II = \begin{array}{|c|c|} \hline UU' & UV' \\ \hline VU' & VV' \\ \hline \end{array} \text{ où}$$

AA' = $n \times$ (matrice des "similarités brutes" entre unités);

$A'A$ = $n \times$ (matrice des variances et covariances empiriques)

lorsque les variables sont centrées; $nVV' = A'A$.

UV' = A (la matrice originale, dont chaque élément est donc représenté comme le produit "scalaire" d'un "effet unité" par un "effet variable");

UU' = $n \times$ (matrice de "projection des moindres carrés")
= $n \times$ (matrice des similarités entre unités).

Les (u_1, \dots, u_n) et (v_1, \dots, v_m) sont analysés dans la section intitulée "Variables et unités statistiques"; en outre, ils se prêtent bien à une représentation graphique très utile qui est présentée dans la dernière section de cet article.

LA DECOMPOSITION SINGULIERE D'UNE MATRICE

Toute matrice A possède une *unique* "décomposition singulière" (terminologie mathématique traditionnelle, assez malheureuse, sinon ... singulière; les psychométriciens l'appellent la décomposition d'Eckart-Young); nous allons voir en détail ce que cela signifie au moyen d'un exemple de taille minimum (4 lignes et 3 colonnes).

- Supposons que la mesure de 3 variables chez 4 "sujets" fournisse les 12 nombres suivants:

$$A = \begin{array}{|c|c|c|} \hline -1 & 13 & 1 \\ \hline -1 & 7 & 7 \\ \hline 3 & 3 & 9 \\ \hline 11 & 1 & 7 \\ \hline \end{array}$$

Les lignes $l_1' = [-1 | 13 | 1]$, ..., $l_4' = [11 ; 1 ; 7]$

représentent les 4 sujets (ou unités statistiques);

les colonnes $c_1 = \begin{bmatrix} -1 \\ -1 \\ 3 \\ 11 \end{bmatrix}$, $c_2 = \begin{bmatrix} 13 \\ 7 \\ 3 \\ 1 \end{bmatrix}$, $c_3 = \begin{bmatrix} 1 \\ 7 \\ 9 \\ 7 \end{bmatrix}$

représentent les 3 variables

2. A est égale à la somme des trois matrices suivantes:

$$\underline{A = A_1 + A_2 + A_3}$$

(par exemple, en première ligne et seconde colonne, $13 = 6 + 6 + 1$).

$$A_1 = \begin{bmatrix} 3 & 6 & 6 \\ 3 & 6 & 6 \\ 3 & 6 & 6 \\ 3 & 6 & 6 \end{bmatrix} \quad A_2 = \begin{bmatrix} -6 & 6 & -3 \\ -2 & 2 & -1 \\ 2 & -2 & 1 \\ 6 & -6 & 3 \end{bmatrix} \quad A_3 = \begin{bmatrix} 2 & 1 & -2 \\ -2 & -1 & 2 \\ -2 & -1 & 2 \\ 2 & 1 & -2 \end{bmatrix}$$

Les matrices A_1 , A_2 et A_3 sont construites comme suit:

$$A_1 = 18 [0,5;0,5;0,5;0,5] \cdot [0,\bar{3};0,\bar{6};0,\bar{6}] \\ = s_1 \quad g_1 \quad d_1'$$

$$A_2 = 13,4164 [-0,6708;-0,2236;0,2236;0,6708] \cdot [0,\bar{6};-0,\bar{6};0,\bar{3}] \\ = s_2 \quad g_2 \quad d_2'$$

$$A_3 = 6 [0,5;-0,5;-0,5;0,5] \cdot [0,\bar{6};0,\bar{3};-0,\bar{6}] \\ = s_3 \quad g_3 \quad d_3'$$

Les trois nombres $s_1 = 18$, $s_2 = 13,4164$ ($= 3\sqrt{20}$) et $s_3 = 6$ s'appellent les *valeurs singulières* de la matrice A; on les regroupe dans la matrice S suivante:

$$S = \begin{bmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \\ 0 & 0 & 0 \end{bmatrix}$$

Les valeurs singulières sont des nombres positifs ou nuls. On appelle matrice des *vecteurs singuliers "à gauche"* (de A) la matrice G suivante (il y a 4 vecteurs g, car il y a 4

lignes dans la matrice A) :

$$G = \begin{array}{|c|c|} \hline 0,5 & -0,6708 (= -3/\sqrt{20}) \\ \hline 0,5 & -0,2236 (= -1/\sqrt{20}) \\ \hline 0,5 & 0,2236 (= 1/\sqrt{20}) \\ \hline 0,5 & 0,6708 (= 3/\sqrt{20}) \\ \hline \end{array} \begin{array}{|c|c|} \hline 0,5 & -0,2236 (= -1/\sqrt{20}) \\ \hline -0,5 & 0,6708 (= 3/\sqrt{20}) \\ \hline -0,5 & -0,6708 (= -3/\sqrt{20}) \\ \hline 0,5 & 0,2236 (= 1/\sqrt{20}) \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline & g_1 & g_2 & g_3 & g_4 \\ \hline \end{array}$$

On appelle matrice des *vecteurs singuliers "à droite"* (de A) la matrice D suivante (il y a 3 vecteurs d, car il y a 3 colonnes dans la matrice A) :

$$D = \begin{array}{|c|c|c|} \hline 1/3 & 2/3 & 2/3 \\ \hline 2/3 & -2/3 & 1/3 \\ \hline 2/3 & 1/3 & -2/3 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline d_1 & d_2 & d_3 \\ \hline \end{array} .$$

Dans cet exemple, la matrice D est égale à sa transposée

$$D' = \begin{array}{|c|} \hline d_1' \\ \hline d_2' \\ \hline d_3' \\ \hline \end{array}$$

ce n'est en général pas le cas.

Les produits matriciels figurant ci-dessus ($g_1 d_1'$, $g_2 d_2'$, $g_3 d_3'$), et qui permettent de déterminer A_1^1 , A_2^2 et A_3^3 respectivement, se calculent comme on le montre pour A_3^3 (ce qui suffit) :

$$A_3 = 6 \begin{array}{|c|c|} \hline 0,5 & \\ \hline -0,5 & \\ \hline -0,5 & \\ \hline 0,5 & \\ \hline \end{array} \begin{array}{|c|} \hline 0,6; 0,3; -0,6 \\ \hline \end{array} = 6 \begin{array}{|c|c|c|} \hline (0,5) \cdot (0,6) & (0,5) \cdot (0,3) & (0,5) \cdot (-0,6) \\ \hline (-0,5) \cdot (0,6) & (-0,5) \cdot (0,3) & (-0,5) \cdot (-0,6) \\ \hline (-0,5) \cdot (0,6) & (-0,5) \cdot (0,3) & (-0,5) \cdot (-0,6) \\ \hline (0,5) \cdot (0,6) & (0,5) \cdot (0,3) & (0,5) \cdot (-0,6) \\ \hline \end{array}$$

$$= \begin{array}{|c|c|c|} \hline 6 \cdot (0,3) & 6 \cdot (0,1\bar{6}) & 6 \cdot (-0,3) \\ \hline 6 \cdot (-0,3) & 6 \cdot (-0,1\bar{6}) & 6 \cdot (-0,3) \\ \hline 6 \cdot (-0,3) & 6 \cdot (-0,1\bar{6}) & 6 \cdot (-0,3) \\ \hline 6 \cdot (0,3) & 6 \cdot (0,1\bar{6}) & 6 \cdot (-0,3) \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 2 & 1 & -2 \\ \hline -2 & -1 & 2 \\ \hline -2 & -1 & 2 \\ \hline 2 & 1 & \textcircled{-2} \\ \hline \end{array}$$

Une matrice construite comme l'est A_3 , c'est-à-dire au moyen de deux vecteurs g_3 et d_3 , est dite *matrice de rang = 1*; A_1 et A_2 sont donc aussi des matrices de rang = 1. Les matrices dont tous les éléments sont nuls sont dites *matrices de rang = 0*. Toutes les autres matrices sont de rang supérieur à 1; plus exactement, le rang d'une matrice est égal au nombre de ses valeurs singulières non nulles; par exemple, la matrice A est de rang = 3. Le rang d'une matrice M ne peut pas dépasser le plus petit des deux nombres suivants:

(nombre des lignes de M) et (nombre des colonnes de M).

Le rang est ainsi une mesure assez simple de la complexité d'une matrice; c'est exactement un "nombre de dimensions" ou "nombre de degrés de liberté". Donc les matrices de rang = 1 sont les matrices (non nulles) de complexité minimum.

3. On remarquera aussi que le mode de formation de A_3 (ou A_1 , ou A_2) détaillé ci-dessus correspond à une *analyse multiquantitative* de A_3 (alors que les calculs d'analyse de variance correspondent à une analyse additive):

- la 3ème valeur singulière $s_3 = 6$ équivaut à une mesure d'importance globale ("effet global");
- le 3ème vecteur singulier à gauche $g_3 = [0,5; -0,5; -0,5; 0,5]'$ comprend les 4 "effets de ligne";
- le 3ème vecteur singulier à droite $d_3 = [2/3; 1/3; -2/3]'$ comprend les 3 "effets de colonne".

C'est ainsi que le nombre -2, qui figure à l'intersection de la 4ème ligne et de la 3ème colonne de la matrice A_3 , est égal au produit:

$$(\text{effet global}=6) \times (\text{effet de la 4e ligne}=0,5) \times (\text{effet de la 3e colonne}=-0,6).$$

La *décomposition singulière* de la matrice A est cette représentation sous forme d'agrégat de trois analyses multiquantitatives

$$A_1 = s_1 \begin{bmatrix} g_1 \\ d_1 \end{bmatrix}', \quad A_2 = s_2 \begin{bmatrix} g_2 \\ d_2 \end{bmatrix}', \quad A_3 = s_3 \begin{bmatrix} g_3 \\ d_3 \end{bmatrix}'$$

dont chacune explore une nouvelle "direction" en augmentant de 1 le rang de la matrice somme-partielle:

$$\begin{array}{lll} \underline{A = A_1 + A_2 + A_3} & \text{rang } (A_1) & = 1 \\ & \text{rang } (A_1 + A_2) & = 2 \\ & \text{rang } (A_1 + A_2 + A_3) & = \text{rang } (A) = 3 \end{array}$$

Un lecteur familiarisé avec le produit matriciel remarquera que l'on peut aussi écrire $A = G S D'$, ce qui est une forme équivalente de la décomposition singulière de A.

4. La décomposition singulière ne peut être obtenue pratiquement que par un ordinateur (problème des valeurs propres et vecteurs propres de deux matrices symétriques; cf. l'appendice); mais ses propriétés remarquables (I à VI ci-dessous), en font un instrument indispensable en analyse des données multidimensionnelles.

I La décomposition singulière existe toujours; elle est essentiellement unique.

II A_1 est la matrice de rang = 1 à distance *minimum* de A (c'est donc la meilleure approximation B de rang = 1 au sens des moindres carrés, l'approximation dont la matrice résiduelle $A-B$ est la plus petite possible).

De même $A_1 + A_2$ est la matrice de rang = 2 à distance minimum de la matrice A:

$$A_1 + A_2 = \begin{array}{|c|c|c|} \hline & 3-6 & 6+6 \\ \hline & 3-2 & 6+2 \\ \hline & 3+2 & 6-2 \\ \hline & 3+6 & 6-6 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline & -3 & 12 \\ \hline & 1 & 8 \\ \hline & 5 & 4 \\ \hline & 9 & 0 \\ \hline \end{array} .$$

Plus généralement si $M = M_1 + M_2 + \dots + M_r$ est la décomposition singulière d'une matrice M , alors $M_1 + M_2 + \dots + M_s$ ($s < r$) est la meilleure approximation de rang = s de la matrice M.

III Lorsque $M = M_1 + \dots + M_r$ est une matrice "carrée" (autant de lignes que de colonnes), l'approximation $M_1 + \dots + M_s$ ($s < r$) vérifie en même temps une propriété de maximum parmi toutes les transformations simultanées de M; cette propriété, qui est énoncée exactement dans l'appendice, comprend comme cas spéciaux les propriétés de variance maximum des composantes principales et les propriétés de corrélation maximum des variables canoniques.

IV Les propriétés II et III ci-dessus sont valables pour autant que les valeurs singulières aient été rangées dans leur ordre décroissant: $s_1 \geq s_2 \geq s_3 \geq \dots \geq 0$ (ce qui a été fait dans l'exemple numérique). En outre:

$\frac{s_1^2}{s_1^2 + s_2^2 + s_3^2}$ (= 60% dans notre exemple) mesure la qualité de l'approximation de A par la matrice A_1 de rang = 1;

$\frac{s_1^2 + s_2^2}{s_1^2 + s_2^2 + s_3^2}$ (= 93% dans notre exemple) mesure la qualité de l'approximation de A par la matrice $A_1 + A_2$ de rang = 2;

(etc.)

- V Les vecteurs singuliers à gauche g_1, g_2, g_3, g_4 sont normés ($\|g_k\| = 1$) et perpendiculaires deux à deux ($g_i^T g_k = 0$). Voici deux vérifications, parmi les dix qui sont nécessaires:

$$\|g_3\| = \sqrt{(0,5)^2 + (-0,5)^2 + (-0,5)^2 + (-0,5)^2} = \sqrt{0,25 \cdot 4} = \sqrt{1} = 1 ;$$

$$g_2^T g_3 = (-3/\sqrt{20}) \cdot (0,5) + (-1/\sqrt{20}) \cdot (-0,5) + (1/\sqrt{20}) \cdot (-0,5) + (3/\sqrt{20}) \cdot (0,5) \\ = (3/\sqrt{20}) \cdot 0 + (1/\sqrt{20}) \cdot 0 = 0 .$$

Les vecteurs singuliers à droite d_1, d_2, d_3 sont aussi normés et perpendiculaires deux à deux.

Cette propriété de perpendicularité est précieuse en statistique, car cela signifie que si $X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3$ et Y_4 sont des variables de même variance (par exemple des variables réduites) et non corrélées deux à deux, alors les variables dérivées

$$(-3/\sqrt{20}) X_1 + (-1/\sqrt{20}) X_2 + (1/\sqrt{20}) X_3 + (3/\sqrt{20}) X_4 \quad \text{et}$$

$$(0,5) Y_1 + (-0,5) Y_2 + (-0,5) Y_3 + (0,5) Y_4$$

sont non corrélées.

- VI On peut définir facilement des fonctions de la matrice A, et c'est essentiellement la seule façon de le faire. Voici quelques exemples:

$$\log(A) = \log(s_1) \boxed{g_1 d_1'} + \log(s_2) \boxed{g_2 d_2'} + \log(s_3) \boxed{g_3 d_3'} ;$$

$$e^A = e^{s_1} \boxed{g_1 d_1'} + e^{s_2} \boxed{g_2 d_2'} + e^{s_3} \boxed{g_3 d_3'} ;$$

$$A^{(k)} = s_1^k \boxed{g_1 d_1'} + s_2^k \boxed{g_2 d_2'} + s_3^k \boxed{g_3 d_3'} \quad (k \geq 0) ;$$

$$A^{(-k)} = s_1^{-k} \boxed{d_1 g_1'} + s_2^{-k} \boxed{d_2 g_2'} + s_3^{-k} \boxed{d_3 g_3'} \quad (k \geq 0) ;$$

(puissances de l'inverse généralisé $A^+ = A^{(-1)}$ de Moore-Penrose) ;

$$A^{(-1/2)} = \frac{1}{\sqrt{s_1}} \begin{bmatrix} d_1 g_1' \\ d_1 g_2' \\ d_1 g_3' \end{bmatrix} + \frac{1}{\sqrt{s_2}} \begin{bmatrix} d_2 g_1' \\ d_2 g_2' \\ d_2 g_3' \end{bmatrix} + \frac{1}{\sqrt{s_3}} \begin{bmatrix} d_3 g_1' \\ d_3 g_2' \\ d_3 g_3' \end{bmatrix}$$

VARIABLES ET UNITES STATISTIQUES

Les unités statistiques et les variables correspondent respectivement aux lignes et aux colonnes de la matrice de données A.

Nous montrons d'abord que les vecteurs singuliers à droite (colonnes de D) sont des vecteurs dérivés des lignes de A et que les vecteurs singuliers à gauche (colonnes de G) sont des vecteurs dérivés des colonnes de A. C'est la première interprétation des matrices G et D, qui fournissent donc respectivement une meilleure représentation des variables et des unités statistiques (par des vecteurs perpendiculaires deux à deux).

Mais on peut aller plus loin, grâce à la décomposition singulière, et définir une paire de transformations simultanées des unités et variables en "effets des unités" et "effets des variables".

1. Les vecteurs singuliers à droite d sont dérivés des unités; chaque d' est une combinaison linéaire des lignes l' de la matrice A. Considérons, par exemple, le troisième vecteur singulier à droite d_3 :

$$\begin{aligned} d_3' &= \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix} = \frac{0,5}{6} \begin{bmatrix} -1 & 13 & 1 \end{bmatrix} \\ &\quad + \frac{-0,5}{6} \begin{bmatrix} -1 & 7 & 7 \end{bmatrix} \\ &\quad + \frac{-0,5}{6} \begin{bmatrix} 3 & 3 & 9 \end{bmatrix} \\ &\quad + \frac{0,5}{6} \begin{bmatrix} 11 & 1 & 7 \end{bmatrix} \end{aligned}$$

Symboliquement :

$\frac{g_{31}}{s_3}$	$\begin{bmatrix} 1_1' \\ 1_2' \\ 1_3' \\ 1_4' \end{bmatrix}$
$\frac{g_{32}}{s_3}$	$\begin{bmatrix} 1_1' \\ 1_2' \\ 1_3' \\ 1_4' \end{bmatrix}$
$\frac{g_{33}}{s_3}$	$\begin{bmatrix} 1_1' \\ 1_2' \\ 1_3' \\ 1_4' \end{bmatrix}$
$\frac{g_{34}}{s_3}$	$\begin{bmatrix} 1_1' \\ 1_2' \\ 1_3' \\ 1_4' \end{bmatrix}$

Ce qu'il y a d'intéressant dans ce résultat, c'est que les numérateurs g_{3k} des coefficients sont les composantes du troisième vecteur singulier à gauche g_3 , alors que les dénominateurs sont tous égaux à la troisième valeur singulière s_3 . Schématiquement, pour les trois vecteurs singuliers à droite:

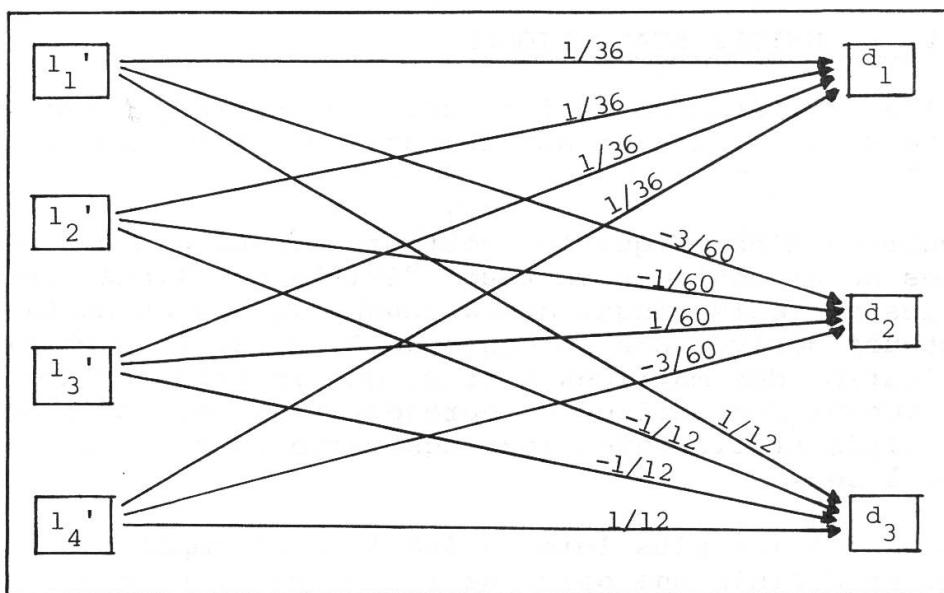


Figure 1

2. Les trois premiers vecteurs singuliers à gauche g sont dérivés des variables; chacun de ces g est une combinaison linéaire des colonnes c de la matrice A .

Considérons, par exemple, le second vecteur singulier à gauche g_2

$$g_2 = \begin{pmatrix} -3/\sqrt{20} \\ -1/\sqrt{20} \\ 1/\sqrt{20} \\ 3/\sqrt{20} \end{pmatrix} = \frac{2/\sqrt{3}}{3\sqrt{20}} \begin{pmatrix} -1 \\ -1 \\ 3 \\ 11 \end{pmatrix} + \frac{-2/\sqrt{3}}{3\sqrt{20}} \begin{pmatrix} 13 \\ 7 \\ 3 \\ 1 \end{pmatrix} + \frac{1/\sqrt{3}}{3\sqrt{20}} \begin{pmatrix} 1 \\ 7 \\ 9 \\ 7 \end{pmatrix}$$

$$\text{Symboliquement : } g_2 = \frac{d_{21}}{s_2} \begin{pmatrix} c_1 \end{pmatrix} + \frac{d_{22}}{s_2} \begin{pmatrix} c_2 \end{pmatrix} + \frac{d_{23}}{s_2} \begin{pmatrix} c_3 \end{pmatrix}.$$

On observe de nouveau que les numérateurs d_{2k} des coefficients sont les composantes du second vecteur singulier à droite d_2 , alors que les dénominateurs sont tous égaux à la seconde valeur singulière s_2 .

Schématiquement, pour les trois premiers vecteurs singuliers à gauche:

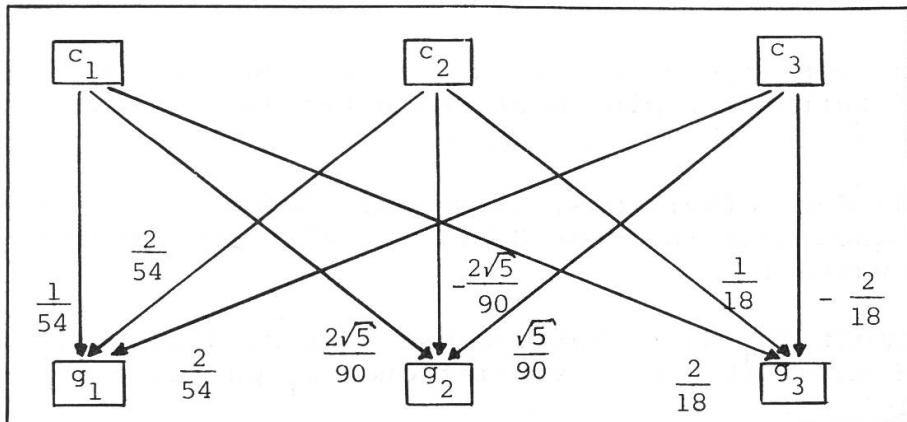


Figure 2

3. Désignons par n le nombre d'unités statistiques, c'est-à-dire le nombre de lignes de la matrice A (dans notre cas, $n = 4$).

$$\begin{aligned} \text{Les lignes } u'_1 &= [2/2 ; -6/\sqrt{20} ; 2/2 ; -2/\sqrt{20}] \\ u'_2 &= [2/2 ; -2/\sqrt{20} ; -2/2 ; 6/\sqrt{20}] \\ u'_3 &= [2/2 ; 2/\sqrt{20} ; -2/2 ; -6/\sqrt{20}] \\ u'_4 &= [2/2 ; 6/\sqrt{20} ; 2/2 ; 2/\sqrt{20}] \end{aligned}$$

de la matrice $\sqrt{n} G$, au contraire des colonnes g_k , représentent des effets des unités statistiques (ce sont les "effets de ligne" introduits plus haut, mais amplifiés par un facteur \sqrt{n}).

De même, les colonnes v'_1, v'_2, v'_3 de la matrice $(1/\sqrt{n}) SD'$ — c'est-à-dire, dans notre cas, de $\frac{1}{2} \times$ le produit matriciel de la matrice S par la transposée de la matrice D —, soit:

$$v_1 = \begin{bmatrix} s_1/6 \\ 2s_2/6 \\ 2s_3/6 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ \sqrt{20} \\ 2 \\ 0 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 2s_1/6 \\ -2s_2/6 \\ s_3/6 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ -\sqrt{20} \\ 1 \\ 0 \end{bmatrix}$$

$$v_3 = \begin{bmatrix} 2s_1/6 \\ s_2/6 \\ -2s_3/6 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ \sqrt{20}/2 \\ -2 \\ 0 \end{bmatrix}$$

représentent des *effets des variables* (ce sont les "effets de colonne" introduits plus haut, mais transformés par un facteur s_k/\sqrt{n}).

4. En effet, de ces définitions, il découle (comme l'ont remarqué quelques auteurs, dont Gabriel, 1971) les interprétations suivantes:

- a) Tout élément a_{ij} de la matrice est égal au produit (scalaire) d'un effet "unité statistique" u_i' par un effet "variable" v_j .

Par exemple $a_{1,2} = 13 = u_1' v_2$, donc

$$\begin{aligned} a_{1,2} &= (2/2)(2s_1/6) + (-6/\sqrt{20})(-2s_2/6) + (2/2)(s_3/6) + (-2/\sqrt{20})(0) \\ &= 6 + 6 + 1 + 0 \end{aligned}$$

Cette représentation est d'autant plus intéressante que, comme tout produit scalaire, $u_1' v_2$ se laisse interpréter par un triple produit:

$$\begin{aligned} u_1' v_2 &= (\text{longueur de } u_1) \times (\text{longueur de } v_2) \times \\ &\quad (\text{cosinus de l'angle entre les directions } u_1 \text{ et } v_2) \\ &= 2 \times 7,55 \times \cos(30^\circ 34' 37''). \end{aligned}$$

La "longueur" de $u_1 = \|u_1\| = \sqrt{(2/2)^2 + (-6/\sqrt{20})^2 + (2/2)^2 + (-2/\sqrt{20})^2} = \sqrt{4} = 2$ est une mesure de "l'importance" de l'unité 1 (cf. ci-dessous).

$v_2 = 7,55$ est une mesure de "l'importance" de la variable 2 (cf. ci-dessous).

$\cos(30^\circ 34' 37'') = 0,86$ est une mesure du degré "d'association" entre unité 1 et variable 2; cette association est maximum lorsque $\cos = \pm 1$, minimum lorsque $\cos = 0$.

- b) L'effet différentiel $u_1 - u_2$ (entre unités 1 et 2). La différence $a_{13} - a_{23}$ entre les unités 1 et 2, en ce qui concerne la variable 3, vaut aussi

$$\frac{a_{13} - a_{23}}{\|u_1 - u_2\| \cdot \|v_3\|} = \frac{(u_1 - u_2)' v_3}{\|u_1 - u_2\| \cdot \|v_3\|} \cdot \cos(\text{angle entre effet } u_1 - u_2 \text{ et effet } v_3).$$

- c) L'effet différentiel $v_1 - v_2$ (entre variables 1 et 2). La différence $a_{31} - a_{32}$ entre les variables 1 et 2, pour l'unité 3, vaut aussi

$$\frac{a_{31} - a_{32}}{\|u_3\| \cdot \|v_1 - v_2\|} = \frac{u_3' (v_1 - v_2)}{\|u_3\| \cdot \|v_1 - v_2\|} \cdot \cos(\text{angle entre effet } u_3 \text{ et effet } v_1 - v_2).$$

- d) L'effet d'interaction entre paire d'unités et paire de variables. En analyse de variance, la combinaison linéaire suivante est appelée l'effet d'interaction entre paire (u_1, u_2) et paire (v_1, v_2) :

$$\frac{a_{11} - a_{12} - a_{21} + a_{22}}{\|u_1 - u_2\| \cdot \|v_1 - v_2\|} = \frac{(u_1 - u_2)' (v_1 - v_2)}{\|u_1 - u_2\| \cdot \|v_1 - v_2\|} \cdot \cos(\text{angle entre les deux}).$$

- e) Si les variables sont centrées (c'est-à-dire si la somme de chaque colonne c_k de la matrice A est nulle), alors:

i) la variance empirique de la $k^{\text{ème}}$ variable est égale au carré $\|v_k\|^2$ de la longueur du vecteur v_k ;

ii) la covariance empirique entre les $i^{\text{ème}}$ et $k^{\text{ème}}$ variables est égale au produit scalaire $v_i' v_k$;

iii) le coefficient de corrélation empirique entre les variables i et k est égal à $(v_i' v_k) / (\|v_i\| \cdot \|v_k\|)$, soit le cosinus de l'angle entre les directions v_i et v_k .

iv) la distance euclidienne $\|v_i - v_k\|$ entre effets de variables est proportionnelle à la distance euclidienne $\|c_i - c_k\|$ entre colonnes de la matrice A:

$$\underline{\|v_i - v_k\| = \|c_i - c_k\| / \sqrt{n}} ; \text{ en outre,}$$

$\|v_i - v_k\|$ = écart-type de la variable "(variable i) - (variable k)".

5. Les interprétations du paragraphe précédent correspondent aux propriétés intrinsèques de nos trois variables; aux propriétés intrinsèques de nos quatre unités statistiques (cf. ci-dessous les points c, d et e) correspondent des propriétés "extrinsèques" des trois variables (essentiellement le fait que ces trois variables déterminent la matrice Q introduite ci-dessous). C'est ce que nous allons expliquer dans ce paragraphe avant de poursuivre (au paragraphe 6) la liste des interprétations des effets u_i et v_j .

Supposons que nos quatre unités statistiques nous fournissent des informations non seulement sur nos trois variables, mais aussi sur d'autres variables y, z, \dots ; il suffira de considérer le cas d'une unique variable y . Nous disposons donc d'un vecteur (colonne) y de quatre observations:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

Nous allons aussi changer l'optique de recherche adoptée jusqu'à présent, soit l'étude des relations entre les trois variables; nous voulons maintenant étudier l'impact de nos trois variables sur la variable y — on pourrait aussi dire que nous voulons "prédirer" y au moyen de nos trois variables.

Les variables du début jouent maintenant le rôle de *trois variables exogènes* (ou variables "prédictrices", "indépendantes", "prédéterminées"), alors que y est une variable dite "endogène" (ou variable "critère", "dépendante").

La méthode des moindres carrés permet de trouver la combinaison \hat{y} de trois variables exogènes qui est la plus proche du vecteur d'observation y ; elle montre que cette meilleure approximation s'interprète géométriquement comme une projection (perpendiculaire, et non oblique):

$$Q : y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \longrightarrow \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{bmatrix}$$

et, ce qui est plus important, que chaque composante \hat{y}_k du vecteur "ajusté" \hat{y} s'obtient par combinaison linéaire des composantes y_i du vecteur original y :

$$\hat{y}_k = \omega_k' y = \omega_{k1}(y_1) + \omega_{k2}(y_2) + \omega_{k3}(y_3) + \omega_{k4}(y_4).$$

Dans notre cas, le projecteur des moindres carrés est la matrice Q suivante:

$$Q = \begin{pmatrix} \omega_1' \\ \omega_2' \\ \omega_3' \\ \omega_4' \end{pmatrix} = \frac{1}{20} \begin{pmatrix} 19 & 3 & -3 & 1 \\ 3 & 11 & 9 & -3 \\ -3 & 9 & 11 & 3 \\ 1 & -3 & 3 & 19 \end{pmatrix}$$

Ainsi, la meilleure approximation (ou "valeur ajustée")

$$\text{de } y_1 \text{ est } \hat{y}_1 = (19y_1 + 3y_2 - 3y_3 + y_4) / 20,$$

$$\text{de } y_2 \text{ est } \hat{y}_2 = (3y_1 + 11y_2 + 9y_3 - 3y_4) / 20,$$

$$\text{de } y_3 \text{ est } \hat{y}_3 = (-3y_1 + 9y_2 + 11y_3 + 3y_4) / 20,$$

$$\text{de } y_4 \text{ est } \hat{y}_4 = (y_1 - 3y_2 + 3y_3 + 19y_4) / 20.$$

Les propriétés matricielles de Q figurent en appendice; les propriétés plus directement utilisables dans notre analyse de données sont les suivantes:

- a) Le carré de la longueur de chaque vecteur-ligne ω_i' (ou vecteur-colonne, puisque Q est toujours symétrique) est égal à l'élément q_{ii} sur la diagonale de la matrice.

Exemple de la seconde ligne:

$$\|\omega_2\|^2 = (3/20)^2 + (11/20)^2 + (9/20)^2 + (-3/20)^2 = 11/20.$$

- b) Le produit scalaire des vecteurs ω_i' et ω_k est égal à l'élément q_{ik} en ligne i et colonne k .

Exemple des lignes 3 et 4:

$$\begin{aligned} \omega_3' \omega_4 &= (-3/20)(1/20) + (9/20)(-3/20) + (11/20)(3/20) \\ &\quad + (3/20)(19/20) \\ &= 3/20. \end{aligned}$$

- c) Le rang de la matrice A (= 3 dans notre exemple) est égal à la somme des éléments diagonaux de la matrice Q ($= \frac{19}{20} + \frac{11}{20} + \frac{11}{20} + \frac{19}{20}$ dans notre exemple); ces

éléments diagonaux q_{ii} représentent la contribution individuelle de chaque unité statistique i à la complexité totale de notre matrice de données A .

Ainsi, dans notre exemple, les deux "gros contribuables" sont les unités 1 et 4.

- d) Chacun des éléments diagonaux q_{kk} de Q est aussi une mesure de la sensibilité de l'ajustement lors de modification (erreur, changement, etc.) de y_k .

Exemple de la deuxième composante:

Si y_2 est remplacé par $y_2 + c$, alors tous les \hat{y}_i sont remplacés par des \hat{y}_i^* (que l'on calculera au moyen des formules ci-dessus pour \hat{y}_i , en substituant $y_2 + c$ à y_2); le nouveau vecteur ajusté $\hat{y}^* = [y_1^*; y_2^*; y_3^*; y_4^*]'$ est alors à la distance $\|\hat{y} - \hat{y}^*\|$ de l'ancien vecteur ajusté \hat{y} .

Il est facile de montrer que la modification relative vaut

$$\frac{\|\hat{y} - \hat{y}^*\|}{c} = \sqrt{q_{22}} \quad (= \sqrt{\frac{11}{20}} = 0,74 \text{ dans notre cas}).$$

- e) Chacun des éléments non-diagonaux q_{ij} de Q est une mesure de la similarité entre unités i et j ; on obtient une mesure standardisée, variant entre -1 et +1, en utilisant

$$\frac{q_{ij}}{\sqrt{q_{ii} q_{jj}}} \quad (\text{qui est égal au cosinus de l'angle entre } \omega_i \text{ et } \omega_j; \text{ cf. l'appendice})$$

Voici la matrice des similarités standardisées pour notre exemple:

	unité 1	unité 2	unité 3	unité 4
--	---------	---------	---------	---------

unité 1	1	+0,21	-0,21	+0,05
unité 2	+0,21	1	+0,82	-0,21
unité 3	-0,21	+0,82	1	+0,21
unité 4	+0,05	-0,21	+0,21	1

6. Voici la suite des interprétations que l'on peut donner aux effets u_i et v_k :

- a) Le carré $\|u_i\|^2$ de la longueur du vecteur u_i (à trois composantes) est égal à $n \cdot q_{ii}$.

Exemple pour la seconde unité:

$$\|u_2\|^2 = (2/2)^2 + (-2/\sqrt{20})^2 + (-2/2)^2 = 44/20 = 4 \cdot \omega_2^2$$

- b) Le produit scalaire $u_i' u_k$ (vecteur à trois composantes) est égal à $n \cdot q_{ik}$.

Exemple pour les unités 3 et 4:

$$u_3' u_4 = (2/2)(2/2) + (2/\sqrt{20})(6/\sqrt{20}) + (-2/2)(2/2) = 12/20 \\ = 4 \cdot \omega_3' \omega_4 .$$

- c) Le cosinus de l'angle entre les directions u_i et u_k est égal au coefficient de similarité standardisé entre unité i et unité k.

Exemple pour les unités 3 et 4:

$$\frac{u_3' u_4}{\|u_3\| \cdot \|u_4\|} = \frac{\frac{12}{20}}{\sqrt{\frac{44}{20} \cdot \frac{76}{20}}} = + 0,21 = \frac{q_{34}}{\sqrt{q_{33} q_{44}}}$$

- d) La distance euclidienne $\|v_i - v_k\|$ entre effets de variables (à trois composantes) est égale à ce que l'on appelle la *distance de Mahalanobis*

$$\underline{(l_i^{-1})' (\frac{1}{n} A'A)^{-1} (l_i^{-1})}$$

entre les lignes i et k de la matrice A; c'est une distance "pondérée" (alors que la distance euclidienne)

$$\|l_i^{-1}\| = (l_i^{-1})' (l_i^{-1})$$

n'est pas pondérée), familière aux statisticiens; la matrice $(\frac{1}{n} A'A)^{-1}$ des coefficients de pondération n'est autre, lorsque les variables sont centrées, que l'inverse de la matrice des variances et covariances empiriques de nos trois variables.

- e) Le carré $\|u_i - v_k\|^2$ de la distance euclidienne entre effet unité u_i et effet variable v_k (tous deux, des vecteurs à trois composantes) dépend à la fois de la contribution de l'unité i à la complexité totale (rang) et de la contribution de la variable k à la dispersion totale (variance).

Plus précisément, lorsque les variables sont centrées, on a la relation:

$$\underline{\|u_i - v_k\|^2 = n \cdot q_{ii} + variance (variable k) - 2 a_{ik} .}$$

LA REPRESENTATION GRAPHIQUE

Il est évident que si la matrice A est de rang > 2, seule une approximation de rang = 2 peut être représentée exactement sur le plan du dessin.

- Puisque les grandeurs qui se laissent le mieux interpréter sont des longueurs de vecteurs (variances et nombres q_{kk}), des angles (coefficients de corrélation, coefficients de similarité) et des distances entre extrémités de vecteurs ($\|u_i - u_k\|$, $\|v_i - v_k\|$, $\|u_i - v_k\|$), le plus naturel serait de représenter les u_i et les v_k par des vecteurs du plan. C'est ce que l'on ferait si les lignes de la matrice correspondaient à un second groupe de variables (cas de l'ACC et de l'AFC) plutôt qu'à des unités statistiques.

Dans notre cas, pour des raisons de lisibilité du graphique, nous adopterons le principe de représentation suivant:

Pour les variables: les vecteurs $\hat{v}_1, \hat{v}_2, \dots$, formés des deux premières composantes de v_1, v_2, \dots

Pour les unités : les points-extrémités des vecteurs $\hat{u}_1, \hat{u}_2, \dots$ formés des deux premières composantes de u_1, u_2, \dots

- Le lecteur pourra s'exercer à construire la représentation graphique de notre exemple.

Ainsi il représentera la variable (colonne de la matrice A)

c_1	c_2	c_3
-------	-------	-------

par un vecteur allant de l'origine (0;0) vers le point

$$(3; \sqrt{20}) = (6; -\sqrt{20}) = (6; \sqrt{20}/2) = \\ (+3,00; +4,47) \quad (+6,00; -4,47) \quad (+6,00; +2,24) .$$

De même, il représentera l'unité statistique (ligne de la matrice A)

1	2	3	4
---	---	---	---

par le point

$$(1; -6/\sqrt{20}) = (1; -2/\sqrt{20}) = (1; 2/\sqrt{20}) = (1; 6/\sqrt{20}) = \\ (+1,00; -1,34) \quad (+1,00; -0,45) \quad (+1,00; +0,45) \quad (+1,00; +1,34)$$

Les matrices exactes des variances et covariances empiriques d'une part, des corrélations empiriques d'autre part, sont les suivantes:

	c_1	c_2	c_3
c_1	32,44	-1,13	22,88
c_2	-1,13	54,75	21,75
c_3	22,88	21,75	42,75

	c_1	c_2	c_3
c_1	1	-0,03	0,61
c_2	-0,03	1	0,45
c_3	0,61	0,45	1

Bien que le graphique ainsi construit représente rigoureusement $A_1 + A_2$, et non $A_1 + A_2 + A_3 = A$, et bien que les variables ne soient pas centrées, on peut y lire cependant que:

- a) variance (c_1) < variance (c_3) < variance (c_2), en comparant les longueurs de \hat{v}_1 , \hat{v}_2 et \hat{v}_3 ;
 - b) correl (c_1, c_2) ≈ 0 (car \hat{v}_1 et \hat{v}_2 sont presque perpendiculaires), correl (c_1, c_3) $>$ correl (c_2, c_3) (car l'angle entre \hat{v}_1 et \hat{v}_3 est inférieur à celui entre \hat{v}_2 et \hat{v}_3);
 - c) c_1 et c_3 forment un sous-groupe en ce qui concerne les distances euclidiennes deux à deux entre les c_k (car la distance entre les extrémités de \hat{v}_1 et \hat{v}_3 est sensiblement plus petite que celle entre les extrémités de \hat{v}_2 et \hat{v}_3);
 - d) les unités 1 et 4 sont les plus "gros contribuables" à la complexité de A et à la sensibilité du projecteur Q (car \hat{u}_1 et \hat{u}_4 sont plus éloignés de l'origine que \hat{u}_2 et \hat{u}_3);
 - e) si l'on avait essayé de construire une échelle à partir des similarités standardisées entre unités, on aurait obtenu $u_1 \overline{u_2} \overline{u_3} \overline{u_4}$, échelle qui se lit bien sur le graphique;
 - f) les valeurs absolues $|a_{ik}|$ des éléments de A correspondent aux positions relatives des points \hat{u}_i et des vecteurs \hat{v}_k (plus exactement, aux angles entre directions \hat{u}_i et \hat{v}_k).
3. Toutes les relations exactes, énumérées dans la section "Variables et unités statistiques", sont valables seulement approximativement dans la représentation graphique.

Lorsqu'une matrice $A = A_1 + \dots + A_r$ est remplacée par $A_1 + A_2$, la qualité de l'ajustement vaut

$$\frac{s_1^2 + s_2^2}{s_1^2 + s_2^2 + \dots + s_r^2}$$

(= 93 % dans notre exemple)

pour les effets différentiels et d'interaction entre paires d'unités et de variables (les a_{ih} et $u_i'v_j$);

$$\boxed{\frac{s_1^4 + s_2^4}{s_1^4 + \dots + s_r^4}}$$

(= 99,1 % dans notre exemple)

pour les variances, covariances et coefficients de corrélation de variables centrées (les $v_i'v_k$ et $\|v_i\|^2$);

$$\boxed{\frac{s_1^o + s_2^o}{s_1^o + \dots + s_r^o}} = \frac{2}{r}$$

(= $\frac{2}{3} = 66,7$ % dans notre exemple)

pour les similitudes entre unités et contributions à la complexité (les q_{ik} , $u_i'u_k$ et $\|u_i\|^2$).

Donc, les corrélations sont les mieux représentées, puis viennent les effets différentiels et d'interaction; enfin les similitudes et contributions à la complexité sont souvent assez faussées dans la représentation graphique (ce qui n'est heureusement pas le cas dans notre exemple où $r = 2+1$).

Il n'est pas difficile de construire une paire de transformations simultanées aux propriétés inverses, où les similitudes sont les mieux représentées et les corrélations les moins bien représentées.

4. Nous terminons par un exemple plus complexe; afin que le lecteur puisse comparer la technique particulière exposée ci-dessus à l'ACP et l'ACC, nous avons repris volontairement un exemple de 18 variables et 18 unités statistiques analysés par Bertier et Bouroche (1975) selon l'ACP et (partiellement) selon l'ACC.

Notre graphique 1 correspond à la figure 1 (page 120) de ces auteurs. Au lieu de leur figure 2 (page 121), correspondant aux valeurs singulières s_1 et s_3 , nous avons préféré dessiner le graphique 2, correspondant aux valeurs singulières s_2 et s_3 , c'est-à-dire la meilleure approximation de la matrice résiduelle $A-A_1$ ($= A_2 + A_3 + \dots + A_{17} + A_{18}$); la direction d'analyse qui correspond à A_1 est interprétée par Bertier et Bouroche (page 121) comme étant le revenu par habitant; la matrice résiduelle analysée dans le graphique 2 est donc la matrice originale "corrigée" des effets du revenu par habitant (autre terminologie: matrice "lissée"). Faute de place, nous ne commenterons que brièvement le seul graphique 1.

$$\begin{aligned}
 a) \quad & (s_1^4 + s_2^4) / (s_1^4 + \dots + s_{18}^4) = 83,6\%; \\
 & (s_1^2 + s_2^2) / (s_1^2 + \dots + s_{18}^2) = 56,6\%; \\
 & (s_1^\circ + s_2^\circ) / (s_1^\circ + \dots + s_{18}^\circ) = 11,1\%.
 \end{aligned}$$

Ces trois rapports nous indiquent que la qualité de l'ajustement est acceptable pour les corrélations, passable pour les effets différentiels et effets d'interaction, et médiocre pour les similitudes et contributions à la complexité. On étudiera donc ces dernières sur la base des valeurs numériques exactes, et non sur la seule base du graphique; ce dernier, pourtant, nous montre utilement la disposition des unités statistiques (distances et angles) dans le plan de meilleure approximation (PMA), ce qui est un aperçu synthétique non négligeable.

- b) Les calculs ont été effectués à partir de variables réduites (variances = 1 et covariances = corrélations). On ne peut donc pas distinguer les variables par leurs variances; on notera cependant que dix variables sont bien représentées dans le PMA (vecteurs de longueur voisine de 1), alors que les variables TES, LOG, ELE, CAL, AIN, TAP, DEN et FBC le sont mal (de ces huit, LOG, CAL et FBC sont bien représentées dans le graphique 2).
- c) Dans le PMA, on distingue nettement trois groupes de variables (FBCF est mal placé par Bertier et Bouroche):
 - I TES, REC, EDU, LOG, ELE;
 - II CAL, PNB, AIN, TLV, ASS, PIA;
 - III TAP, EXP, IMP, RES, POP, DEN, FBC.

Les variables d'un même groupe sont plus fortement corrélées entre elles qu'avec les variables des autres groupes.

- d) Dans le PMA, certains groupements d'unités statistiques apparaissent nettement:
 - A) FI, DA, NO, SU (pays scandinaves);
 - B) PO, ES, GR (pays méditerranéens n'appartenant pas au Marché commun).

Une entité comme le Marché commun (AL, GB, FR, BL, IT), par contre, apparaît en ordre très dispersé dans le PMA.

- e) Les unités statistiques éloignées du centre (US, PO, GR, SU, etc.) contribuent plus fortement que les autres aux deux dimensions "revenu par habitant" et "importance des dépenses publiques" du PMA. On notera que les échelles utilisées pour dessiner les unités et les variables sont différentes.

5. Notons les différences principales avec les graphiques d'ACP et d'analyse factorielle.

- a) On peut en principe représenter n'importe quelle matrice de nombres réels a_{ij} ; la représentation sera utile dans la mesure où les propriétés de la section "Variables et unités statistiques" seront pertinentes (interprétation par produits scalaires et par projections perpendiculaires).
- b) Les axes de coordonnées servent uniquement à construire le graphique; une rotation d'axes n'a pas de sens, "c'était avant qu'il fallait agir", en choisissant une autre paire de transformations simultanées.
- c) Les rôles joués par les variables et par les unités statistiques apparaissent de façon plus symétrique (et mieux que dans les "Q-analysis" et "R-analysis" des psychométriciens):

<u>unités</u>	<u>variables</u>
similarités	corrélations
complexité	dispersion
effets différentiels et d'interaction	

- d) On a tendance à interpréter les graphiques d'ACP et d'AFC en termes de distances entre points; en fait, c'est avant tout en termes d'angles et de projections perpendiculaires qu'il faut lire le message graphique. En ce qui concerne l'AFC, on peut montrer (Schellhorn, 1976) que la notion de barycentre est liée à celle de projection perpendiculaire.

APPENDICE

1. Les valeurs propres λ (respectivement μ) et vecteurs propres d (resp. g) de $A'A$ (resp. AA') sont définis par les relations

$$\underline{A'A \ d = \lambda d} \quad (\text{resp. } \underline{AA' \ g = \mu g}).$$

On peut montrer que les λ et μ sont des nombres réels non-négatifs et que, si on les énumère en ordre décroissant, on a

$$\lambda_1 = \mu_1 \geq \lambda_2 = \mu_2 \geq \dots \geq \lambda_r = \mu_r > \lambda_{r+1} = \mu_{r+1} = \dots = 0;$$

r est égal au rang de la matrice A ; les racines carrées positives $s_i = \sqrt{\lambda_i} = \sqrt{\mu_i}$ sont les valeurs singulières de A .

Les vecteurs propres normés d_i (resp. g_i) de $A'A$ (resp. AA') sont les vecteurs singuliers à droite (resp. à gauche) de A ; ils vérifient les relations $Ad_i = s_i g_i$ et $A'g_k = s_k d_k$.

Soit $G = \begin{bmatrix} g_1 & \dots & g_r \end{bmatrix}$, $D = \begin{bmatrix} d_1 & \dots & d_r \end{bmatrix}$,

$$S = \begin{bmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & s_r \end{bmatrix}$$

On vérifie les relations suivantes:

$$G'G = I_r = D'D$$

$$AD = GS$$

$$A'G = DS$$

$$A = G S D' = s_1 g_1 d_1' + s_2 g_2 d_2' + \dots + s_r g_r d_r'$$

$$A'A = D S^2 D' = s_1^2 d_1 d_1' + \dots + s_r^2 d_r d_r'$$

$$AA' = G S^2 G' = s_1^2 g_1 g_1' + \dots + s_r^2 g_r g_r'$$

$$A^{(k)} = G S^k D' \quad A^\dagger = D S^{-1} G'$$

Le projecteur des moindres carrés $Q = A(A'A)^\dagger A' = GG' = g_1 g_1' + \dots + g_r g_r'$ vérifie les propriétés $Q = Q' = Q^2 = Q^\dagger Q = QQ'$; ses valeurs singulières non-nulles valent + 1. De même pour $A'(AA')^\dagger A = DD'$.

Rangs: rang (A) = rang ($A'A$) = rang (Q) = trace (Q) = $q_{11} + \dots + q_{nn}$.

Les lecteurs un peu mathématiciens pourront consulter l'article de Hawkins et Ben-Israel (1973).

2. Les effets des unités sont les colonnes de la matrice $U' = \sqrt{n} G'$; les effets des variables sont les colonnes de la matrice $V' = (1/\sqrt{n}) SD'$.

La paire de transformations simultanées est la suivante:

$$\begin{aligned} GSD' &= A \longrightarrow A \left[n^{\frac{1}{2}} D S^{-1} \right] = U \\ DSG' &= A' \longrightarrow A' \left[n^{-\frac{1}{2}} G \right] = V ; \end{aligned}$$

$$\begin{array}{|c|c|} \hline AA' & A^{(2)} \\ \hline A^{(2)'} & A'A \\ \hline \end{array} = \begin{array}{|c|c|} \hline GS^2 G' & GS^2 D' \\ \hline DS^2 G' & DS^2 D' \\ \hline \end{array} \text{ est transformée en}$$

$$\begin{array}{|c|c|} \hline UU' & UV' \\ \hline VU' & VV' \\ \hline \end{array} = \begin{array}{|c|c|} \hline nGG' & GSD' \\ \hline DSG' & n^{-\frac{1}{2}} DS^2 D' \\ \hline \end{array}$$

3. La méthode des transformations simultanées est basée sur la remarque suivante:

Si X et Y sont les bases de deux sous-espaces vectoriels de dimension finie, et si l'on transforme X en XS et Y en YT , alors la matrice de tous les produits intérieurs

B	F
F'	H

devient

$S'BS$	$S'FT$
$T'F'S$	THT

.

L'ACC est le cas particulier suivant:

Soit $F^* = B^{-\frac{1}{2}} F H^{-\frac{1}{2}}$ la matrice F standardisée (cela signifie que si F est une matrice de covariances, alors F^* consiste en pseudo-corrélations); soit $F^* = GSD'$ sa décomposition singulière. Alors

$$\begin{bmatrix} x_1 & \dots & x_p \end{bmatrix} = X \longrightarrow X \begin{bmatrix} B^{-\frac{1}{2}} G \end{bmatrix} = \begin{bmatrix} \xi_1 & \dots & \xi_r \end{bmatrix}$$

$$\begin{bmatrix} y_1 & \dots & y_q \end{bmatrix} = Y \longrightarrow Y \begin{bmatrix} H^{-\frac{1}{2}} D \end{bmatrix} = \begin{bmatrix} \eta_1 & \dots & \eta_r \end{bmatrix}$$

est la paire de transformations simultanées cherchée (les ξ_j et η_k avec $j > r$ et $k > r$ peuvent être négligés ici).

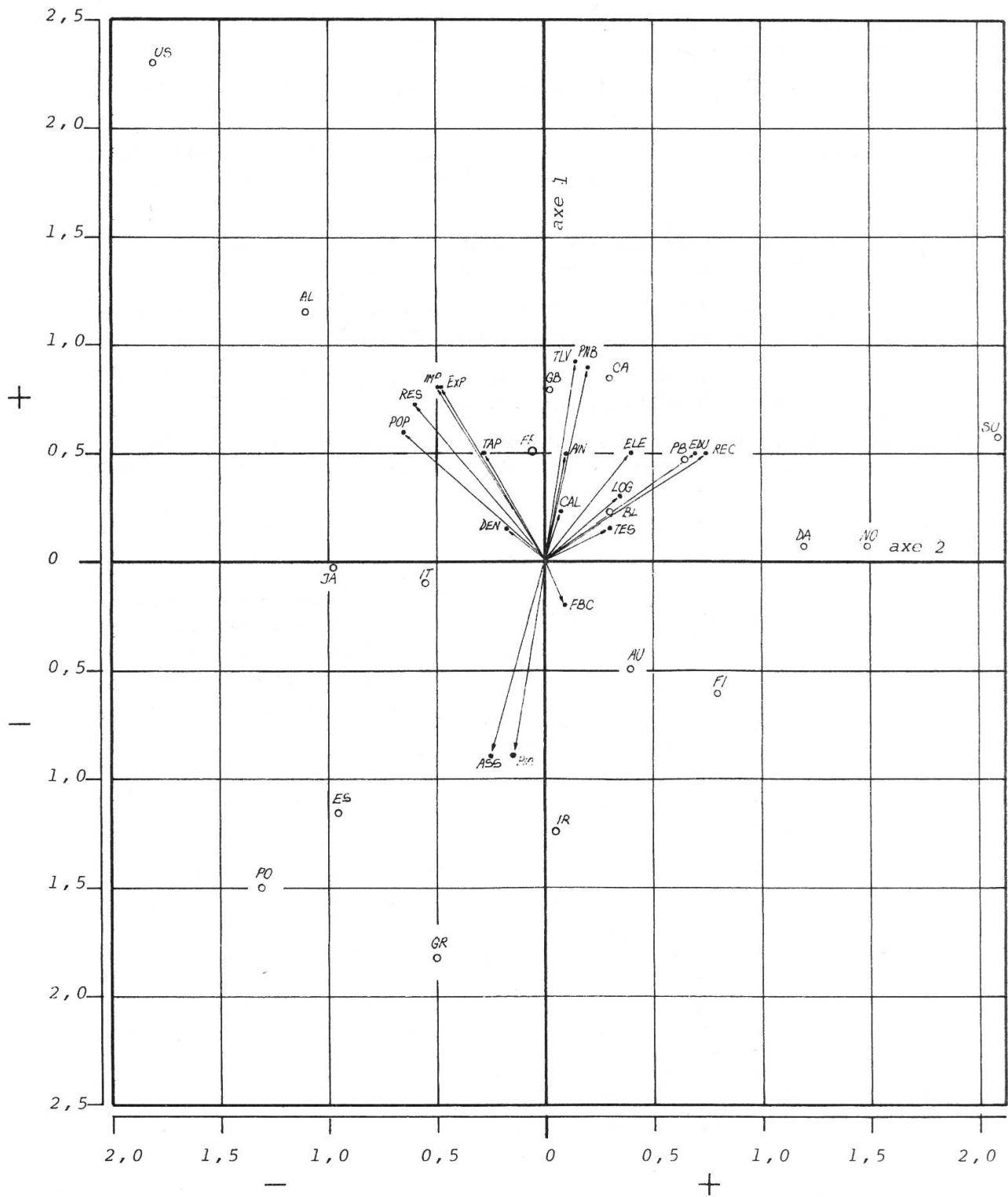
4. Les propriétés d'optimum découlent des deux propriétés suivantes, où $\sigma_1(M) \geq \sigma_2(M) \geq \sigma_3(M) \geq \dots$ désignent les valeurs singulières de la matrice M , rangées dans l'ordre décroissant.

- a) Pour toutes les représentations d'une matrice A comme somme de deux matrices perpendiculaires ($A = B+C$ avec $B'C = \boxed{0}$) où $\text{rang}(A) = r > k = \text{rang}(B)$, on vérifie que:
 $\underline{\sigma_i(A-B)} \geq \sigma_i(A - (A_1 + \dots + A_k)) = \sigma_i(A_{k+1} + \dots + A_r) = s_{i+k}$
 pour $1 \leq i \leq k$. Ainsi le résidu d'ordre k ($A_{k+1} + \dots + A_r$) minimise simultanément toutes les valeurs singulières dans le cas d'une transformation $A \rightarrow A-B$ du type ci-dessus.
- b) Pour toutes les transformations $A \rightarrow S'AT$ d'une matrice $A = GSD'$ de rang r au moyen de matrices S' et T telles que $\text{rang}(S'T) = k \leq r$, $\sigma_i(S') \leq 1$, $\sigma_i(T) \leq 1$, on vérifie que $\underline{\sigma_i(S'AT)} \leq \sigma_i(G_k^T A D_k) = \sigma_i(\bar{A}_1 + \dots + \bar{A}_k) = s_i$
 pour $1 \leq i \leq k$. Ainsi la meilleure approximation d'ordre k ($\bar{A}_1 + \dots + \bar{A}_k$) maximise simultanément toutes les valeurs singulières dans le cas d'une transformation $A \rightarrow S'AT$ du type ci-dessus.

BIBLIOGRAPHIE

- Bertier P. et Bouroche J.-M. (1975): Analyse de données multidimensionnelles. Paris, P.U.F.
- Escoufier Y. (1971): Les liaisons entre groupes d'aléas. Rev. Stat. Appl., vol. 19, p. 5-17.
- Gabriel K.R. (1971): The biplot graphic display of matrices with application to principal component analysis. Biometrika, vol. 58, pp. 453-467.
- Gower J.C. (1971): A general coefficient of similarity and some of its properties. Biometrics, vol. 27, pp. 857-874.
- Hawkins J.B. et Ben-Israel A. (1973): On generalized matrix functions. Linear and multilinear algebra, vol. 1, pp. 163-171.
- Hill M.O. (1974): Correspondence analysis: a neglected multivariate method. Applied statistics, vol. 23, pp. 340-354.
- Schellhorn J.P. (1976): Analyse d'impact par transformations simultanées. Cahiers du Département d'économétrie, Université de Genève.

Jean-Pierre Schellhorn
Département d'Econométrie
Université de Genève
12, rue de Candolle
1211 Genève 4

Graphique 1

Graphique 2