REVUE DE STATISTIQUE APPLIQUÉE

J. PAGÈS

Analyse factorielle multiple et analyse procustéenne

Revue de statistique appliquée, tome 53, n° 4 (2005), p. 61-86

<http://www.numdam.org/item?id=RSA_2005__53_4_61_0>

Rev. Statistique Appliquée, 2005, **LIII** (4), 61-86

ANALYSE FACTORIELLE MULTIPLE ET ANALYSE PROCUSTÉENNE

J. PAGÈS

Laboratoire de mathématiques appliquées,
Agrocampus-Rennes, 65 rue de Saint-Brieuc, CS 84215
35042 Rennes cedex France
E-mail : jerome.pages@agrocampus-rennes.fr

RÉSUMÉ

L'analyse procustéenne est la méthode de référence pour superposer plusieurs représentations de nuages de points homologues. De son côté, l'analyse factorielle multiple fournit aussi une telle représentation. Cet article compare les principes d'obtention de ces deux représentations. Un petit exemple construit spécialement pour mettre en évidence les différences entre les deux méthodes illustre cette comparaison. Un exemple issu de données réelles illustre la parenté entre les résultats des deux méthodes que l'on observe souvent en pratique.

Mots-clés : *Analyse factorielle multiple, analyse procustéenne, analyse sensorielle*

ABSTRACT

Procustes analysis is the reference method to superimpose several representations of the same set of points. Multiple factor analysis provides also such a representation. This paper compares the principles to get those two representations. A small example, built especially to highlight the differences between the two methods, illustrates this comparison. An example resulting from real data illustrates the similarity between the results of the two methods that can be often observed in practice.

Keywords : *Multiple factor analysis, procustes analysis, sensory analysis*

Le principe de l'Analyse Procustéenne (AP) remonte à 1952 (Green). À ce moment le problème est posé de la façon suivante : disposant de deux nuages de points homologues situés dans le même espace, comment faire tourner l'un des nuages pour le faire coïncider le mieux possible avec l'autre? Depuis, l'AP a donné lieu à de nombreux développements dont on trouve un historique dans Dijksterhuis & Gower (1991/2); en particulier, l'Analyse Procustéenne Généralisée (APG; Gower, 1975) s'applique à un ensemble de plus de deux nuages.

Cette méthode est très utilisée par les Anglo-saxons dans le domaine de l'évaluation sensorielle pour traiter les résultats d'épreuves descriptives dans lesquelles J juges évaluent I produits à l'aide de plusieurs critères (*cf.* par exemple Piggott, 1986). Dans ce type d'épreuve, les juges utilisent habituellement les mêmes critères. Mais, on rencontre aussi une méthodologie dans laquelle chaque juge utilise ses propres

critères d'évaluation (Free Choice Profiling; Williams & Arnold, 1984). Les variables n'étant plus homologues d'un juge à l'autre, les méthodes usuelles de traitement ne s'appliquent pas. En revanche, on dispose de J nuages de points homologues et l'APG s'applique. Naturellement cette méthode s'applique aussi lorsque les critères sont identiques d'un juge à l'autre.

Ce type de données relève aussi de l'Analyse Factorielle Multiple (AFM; Escofier & Pagès, 1998) et il est intéressant de comparer les deux approches.

L'APG étant peu répandue en France, nous commençons cette comparaison par quelques rappels sur cette méthode. Notre présentation s'appuie essentiellement sur les travaux de Dijksterhuis et Gower (Dijksterhuis, 1995; Dijksterhuis & Gower, 1991/2).

1. Analyse Procustéenne

1.1. Données, notations

On dispose de J nuages (notés N_I^j) de I points homologues ($i^j \in N_I^j$). Ces nuages évoluent dans des espaces de même dimension K_c. Les coordonnées des points de N_I^j sont rassemblées dans la matrice X_j de dimensions (I, K_c).

Si initialement les N_I^j évoluent dans des espaces de dimensions différentes (K_j), on se ramène au cas précédent en choisissant $K_c = \max(K_j,\ j = 1, J)$: si $K_j < K_c$, alors on considère que N_I^j a une inertie nulle dans K_c-K_j directions, ce que l'on obtient concrètement en ajoutant des colonnes de 0 à la matrice X_j initiale.

En analyse procustéenne les individus sont toujours affectés du poids 1. Nous adoptons ces poids dans cet article.

1.2. Objectifs

Les nuages N_I^j sont placés dans un même espace R^{Kc}. Chaque nuage est centré et il n'y a pas lieu d'opérer de translation. On transforme alors les N_I^j de façon à faire coïncider le mieux possible les points homologues. Dans la version originale, seules les transformations orthogonales, qui ne modifient pas les distances entre les points d'un même nuage c'est-à-dire les rotations et les symétries, sont autorisées. On peut aussi autoriser les homothéties mais, sauf mention explicite du contraire, nous ne les envisageons pas.

À l'issue des transformations, le nuage N_I^j possède de nouvelles coordonnées que l'on rassemble dans la matrice Y_j. La quantité que l'on minimise s'écrit :

$$\sum_{j>l} trace(Y_j - Y_l)'(Y_j - Y_l)$$

Formellement, Y_j se déduisant de X_j par une transformation orthogonale, on peut écrire $Y_j = X_j\ T_j$ avec $T_j\, T_j' = I_d$ (en notant I_d la matrice identité de taille convenable). Le modèle procustéen peut alors s'écrire :

$$q_j X_j T_j = Z + E_j$$

avec Z, matrice de taille (I, K_c) contenant les coordonnées de la configuration dite moyenne (en référence à la façon dont elle est calculée), E_j une matrice de résidus, et q_j un scalaire présent dans le modèle lorsque l'on autorise les homothéties.

Remarque à propos du nuage moyen. – Dès lors que les N_I^j sont placés dans le même espace, on peut construire un nuage moyen N_I, dont chaque point i est l'isobarycentre de ses points homologues dans les N_I^j. Ce nuage est analogue, jusqu'à un certain point, au nuage moyen noté N_I^* en l'AFM (*cf.* Escofier & Pagès, 1998 p. 154). Pour faciliter la comparaison entre les deux méthodes, nous appelons N_I le nuage moyen quelle que soit l'analyse.

1.3. Méthodes et variantes

1.3.1. Selon le nombre de nuages

Premier cas : $J = 2$. C'est le cas de la méthode originale. Il existe une solution analytique (*cf.* Arnold & Williams 1986 et Saporta 1990 p.195) que nous rappelons brièvement.

Soit X_1 et X_2 les tableaux contenant les données initiales; on cherche à transformer X_1 pour «l'ajuster» à X_2.

Soit la décomposition en valeurs singulières :

$$X_1' X_2 = U \wedge V'$$

avec U et V deux matrices orthogonales et $\wedge$ une matrice diagonale ne comportant que des termes positifs ou nuls.

Alors l'ajustement du tableau X_1 est donné par :

$$Y_1 = X_1 U V'$$

La dissymétrie de la solution est seulement apparente; du point de vue de la position relative des points, il revient au même d'ajuster X_1 à X_2 ou X_2 à X_1.

Second cas : $J > 2$. C'est le cas de l'APG. On ne connaît pas de solution analytique. On procède par un algorithme itératif qui, à chaque pas, ajuste successivement chaque nuage N_I^j au nuage moyen (au premier pas le premier nuage sert de nuage moyen), le nuage moyen étant lui-même recalculé après les rotations des N_I^j. Plus précisément, le principe de l'algorithme original de Gower (1975) peut être décrit ainsi :

(1) initialisation du nuage moyen Z (par la première configuration);

(2) ajustement de chacun des J nuages N_I^j au nuage moyen; mise à jour des N_I^j par le résultat de ces ajustements; ces ajustements sont réalisés successivement;

(3) mise à jour du nuage moyen Z à partir des J nuages N_I^j («ajustés»);

(4) mise à jour du critère d'ajustement pour l'ensemble des N_I^j;

(5) reprendre en (2) tant que l'amélioration du critère est supérieure à un seuil fixé.

Il est souhaité que, à la fin de l'algorithme, chaque N_I^j soit ajusté au nuage moyen N_I qui contient les centres de gravité des points homologues.

Cet algorithme converge, et la propriété souhaitée ci-dessus est vérifiée en pratique, mais sa convergence vers un optimum global du critère est incertaine. Plusieurs travaux ont eu pour objet l'amélioration de cet algorithme.

1.3.2. Selon le nombre de dimensions

Premier cas : $K_c = 2$ ou 3. La solution peut être examinée directement et globalement par une représentation graphique.

Second cas : $K_c > 3$. La solution ne peut être examinée qu'en projection sur des sous-espaces. Dans la variante la plus classique (Gower, 1975), à l'issue de l'AP (ou de l'APG), on projette les N_I^j sur les axes factoriels de N_I. Il a aussi été proposé d'utiliser les axes factoriels de l'ensemble N_I^J des N_I^j (on note N_I^J le nuage union des N_I^j).

Compte tenu du fait qu'en pratique on se limite à l'examen d'un sous-espace de R^{Kc}, Peay (1988) a proposé de ne chercher à faire coïncider les N_I^j que sur un nombre réduit de dimensions fixé à l'avance. Cette variante améliore bien évidemment le critère restreint au sous-espace examiné. Mais les solutions obtenues pour les différents nombres de dimensions ne s'emboîtent pas.

1.3.3. Influence, sur les objectifs, du nombre de dimensions

Lorsque $K_c > 3$, on n'étudie pas globalement l'homologie entre les espaces dans lesquels évoluent les N_I^j mais sous-espace par sous-espace. En pratique, on étudie même souvent les représentations dimension par dimension. Ce point de vue rapproche l'APG de l'analyse canonique généralisée, à savoir la recherche d'une suite de directions communes à plusieurs nuages de points homologues. Par la suite nous faisons souvent référence à ce point de vue.

2. Comparaison entre les deux méthodes

Rappelons que l'AFM repose sur une ACP du tableau X de taille (I, K) juxtaposant en ligne les tableaux X_j. Dans cette ACP, les variables du groupe j sont pondérées par $1/\lambda_1^j$ (en notant λ_1^j la première valeur propre de l'ACP séparée du groupe j).

2.1. Représentation des N_I^j

À chaque tableau X_j correspond un nuage N_I^j.

En AFM, les N_I^j sont placés dans l'espace R^K, somme directe des R^{Kj}. Ainsi les N_I^j ne sont pas vraiment dans le même espace; le caractère simultané de cette représentation est artificiel; il se justifie en tant que cadre d'interprétation de la méthode.

En APG, les N_I^j sont tous placés dans le même espace R^{Kc}. Cette représentation correspond à une homologie globale des espaces R^{Kc} *a priori* distincts qui contiennent chacun un nuage N_I^j. Attention : au départ il s'agit d'une homologie globale et non d'une homologie dimension initiale par dimension initiale ce qui est le cas lorsque les variables sont les mêmes d'un groupe à l'autre.

Au départ, c'est-à-dire avant les transformations, cette représentation superposée des N_I^j est artificielle, à l'instar de celle de l'AFM, et se justifie en tant que cadre d'interprétation de la méthode.

L'objet de l'APG est, à partir de cette homologie globale, d'identifier les dimensions homologues des N_I^j. Ces dimensions homologues sont celles qui induisent la même structure sur les individus. On retrouve ici la notion de facteur commun de l'AFM.

Presque toutes les différences entre les deux méthodes dérivent de la différence entre les deux modes de représentation des N_I^j.

2.2. Nuage moyen

Dans les deux analyses, le nuage moyen N_I contient les points i, centres de gravité des ensembles $\{i^j; j = 1, J\}$ (d'où l'appellation nuage moyen). Mais ces deux nuages moyens N_I, étant construits dans des espaces différents, ne possèdent pas la même signification d'une méthode à l'autre.

2.2.1. En AFM

- les i^j associés à un même i appartiennent à des sous-espaces orthogonaux;
- leurs coordonnées se juxtaposent et ne s'additionnent pas deux à deux;
- le carré de la distance entre deux points moyens i et l s'écrit :

$$d^2(i,l) = \frac{1}{J^2} \sum_j d^2(i^j, l^j);$$

- ainsi deux points i et l sont d'autant plus éloignés qu'ils le sont dans chacun des groupes et ce, quelles que soient les directions de cet éloignement dans chacun des groupes;
- dans R^K l'inertie totale de N_I est égale à l'inertie totale de N_I^J (union des N_I^j) divisée par J^2.

2.2.2. En APG

- les i^j appartiennent au même espace R^{Kc};
- le calcul des coordonnées de i résulte d'une «véritable» moyenne entre les coordonnées homologues des $\{i^j | j = 1, J\}$;
- la distance entre les deux points moyens i et l dépend, comme en AFM, de la distance entre les individus i et l dans chaque groupe mais aussi du fait que, le

long des directions homologues, les écarts entre les individus i et l sont dans le même sens ou non (*cf.* Figure 1);

- il en résulte que, à inertie de N_I^J constante, l'inertie de N_I est d'autant plus grande que les écarts entre les points sont identiques dans les directions homologues c'est-à-dire que les N_I^j se ressemblent. D'où la dénomination de nuage consensus attribué au nuage moyen en APG.

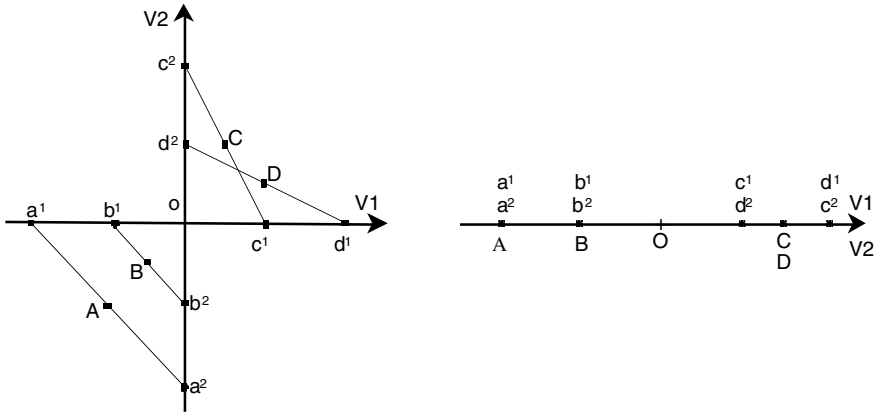


FIGURE 1

Jeu 1 de données choisies : 4 individus (A,B,C,D) décrits par deux groupes de chacun une seule variable $(\{V1\}, \{V2\})$. *Représentations des nuages* N_I *et* N_I^j, *dans l'espace* R^K *de l'AFM (à gauche) et sur l'axe unique de l'analyse procustéenne (à droite). A et B diffèrent pour chaque groupe : en AFM les points moyens diffèrent. Il en est de même pour C et D. Sur l'axe de l'AP, les différences entre A et B vont dans le même sens : les points moyens diffèrent. Sur ce même axe, les différences entre C et D vont en sens contraire : les points moyens sont confondus.*

2.3. Objectif, critère, algorithme

2.3.1. Problématique; modèle

D'un point de vue très général, on recherche dans les deux cas des facteurs communs aux N_I^j. Ces facteurs communs seront visualisés à l'aide d'une représentation superposée des N_I^j.

En AFM :

- le nombre de facteurs communs n'est pas spécifié;
- les facteurs communs peuvent être communs à tous les groupes ou à certains d'entre eux;
- l'objet de l'analyse est de détecter leur existence et de les mettre en évidence.

En APG :

- on suppose qu'il existe une homologie complète entre les espaces dans lesquels évoluent initialement les N_I^j; en pratique on met en évidence des bases homologues de ces sous-espaces; finalement, en termes d'analyse canonique, cela revient à supposer qu'il existe K_c facteurs communs aux J groupes;
- l'objet de l'analyse est d'identifier ces facteurs communs.

2.3.2. Critère

Dans les deux méthodes on peut l'exprimer à partir de la représentation superposée. On considère la partition de N_I^J en I classes contenant chacune les points associés à un même individu. Relativement à cette partition, l'inertie intra-classes de N_I^J s'écrit :

$$\sum_i \sum_j d^2(i^j, i) = \frac{1}{2I} \sum_i \sum_{j,j} d^2(i^j, i^l)$$

En AP ou en APG, on cherche à minimiser cette quantité. Lorsque les homothéties ne sont pas autorisées, il revient au même de maximiser l'inertie inter-classes associée qui n'est autre que l'inertie de N_I. Pour l'APG, on cherche à chaque étape la rotation d'un nuage N_I^j qui satisfait ce critère pour l'ensemble des dimensions.

Remarque. – Lorsque les homothéties sont autorisées, une solution triviale consiste à choisir le coefficient 0 pour chaque nuage. Pour l'éviter, on adopte la contrainte de ne pas modifier l'inertie totale de N_I^J (union des N_I^j).

En AFM, on cherche axe par axe à maximiser l'inertie inter-classes projetée. Malgré cette parenté, les deux critères ne correspondent que partiellement car, en procédant axe par axe, l'inertie totale n'est plus fixe ce qui annule l'équivalence entre minimiser l'inertie intra et maximiser l'inertie inter (remarquons au passage que cette équivalence est aussi perdue dans la méthode de Peay).

Ainsi entre les deux méthodes :

- les représentations géométriques diffèrent; les nuages moyens n'ont pas exactement la même signification;
- les quantités à maximiser diffèrent malgré une certaine parenté;
- les types de transformation des N_I^j diffèrent (rotation ou projection).

2.4. Propriétés des représentations des N_I^j

2.4.1. En APG

- les transformations des N_I^j sont orthogonales (composées éventuellement avec des homothéties);
- la forme des N_I^j est parfaitement respectée : c'est là une contrainte très forte, spécificité de l'APG;

- si K_c est supérieur à 3, on ne peut examiner la représentation superposée des N_I^j qu'à l'aide de projections par exemple sur les axes principaux de N_I ; ces projections sont effectuées après l'ajustement.

2.4.2. En AFM

- la projection a lieu simultanément à l'ajustement;
- la projection de N_I^j se fait sur des axes qui n'appartiennent pas à R^{K_j} ou, selon un autre point de vue, sur des axes non orthogonaux de R^{K_j} ; il en résulte une déformation des N_I^j, même si le nuage N_I est parfaitement représenté. Des éléments concernant ces déformations sont donnés en annexe.

2.5. Premier bilan

L'AFM est une analyse factorielle particulière et une analyse multicanonique (au sens de Carroll, 1968) particulière. Elle n'est pas une analyse procustéenne si l'on considère comme caractéristique de cette dernière la non déformation des N_I^j.

Il n'en reste pas moins que les problématiques des analyses procustéennes et de l'AFM sont apparentées :

- toutes deux peuvent s'articuler autour de la notion de facteur commun;
- toutes deux incluent une représentation superposée des N_I^j et d'un nuage moyen.

2.6. Harmonisation de l'inertie des N_I^j

En AFM l'harmonisation de l'inertie des N_I^j est effectuée avant l'analyse :

- à l'intérieur des groupes par la réduction, sur option, des colonnes;
- entre les groupes par la surpondération des variables qui équivaut à une homothétie des N_I^j.

En APG, deux harmonisations, non exclusives, sont possibles.

Avant l'analyse

Ici la problématique est la même qu'avant une AFM (ou toute autre analyse de ce type de tableau multiple). Dans la pratique anglo-saxonne l'usage est le suivant : dans tous les cas l'inertie de N_I^J est ramenée à 100; sur option, l'inertie de chaque N_I^j est ramenée à $100/J$.

Pendant l'analyse

La transformation de chaque N_I^j peut inclure ou non une homothétie. Cette homothétie est différente selon les N_I^j mais concerne l'ensemble des dimensions de R^{Kc} (il a été proposé par R. Lafosse (1985) une variante dans laquelle le rapport d'homothétie peut varier selon les axes mais il ne s'agit pas d'une analyse

procustéenne au sens où nous l'entendons puisque les N_I^j ne sont plus identiques (à une homothétie près)). L'introduction de telles homothéties :

- améliore par principe le critère;
- revient en fait à faire jouer des rôles différents aux N_I^j. À la limite, ceci permet d'éliminer un nuage de l'analyse (coefficient nul).

2.7. Relations entre les facteurs homologues

Rappel d'une propriété de l'AFM

Nous reprenons les notations de l'AFM :

F_s : coordonnées de N_I le long de l'axe de rang s;

F_s^j : coordonnées de N_I^j le long de l'axe de rang s.

En AFM, ces facteurs bénéficient de la propriété suivante :

$\forall\, s, j : r(F_s^j, F_s) \geqslant 0$ une variable canonique n'est jamais liée négativement à la variable générale de même rang.

Cette propriété constitue le moins que l'on puisse demander à un facteur commun. En revanche, deux variables canoniques de même rang peuvent être liées négativement.

Cas de l'APG

La relation entre facteurs homologues ne semble pas avoir été abordée en ces termes à propos de l'APG. Nous montrons ci-dessous que, de ce point de vue, l'APG vérifie les mêmes propriétés que l'AFM.

Propriété : en APG, $\forall\, s, j : r(F_s^j, F_s) \geqslant 0$

À l'issue de l'APG, chaque N_I^j est ajusté à N_I. Il suffit donc de vérifier que dans l'analyse procustéenne de deux groupes, les facteurs homologues ne sont jamais corrélés négativement. Soit X_1 et X_2 les deux tableaux de données initiaux. On ajuste X_1 à X_2 ce qui conduit à représenter le groupe 1 par (*cf.* § 1.3.1.) :

$$Y_1 = X_1 U V'$$

Les variables sont centrées. Au coefficient I près, les covariances entre facteurs homologues de Y_1 et de X_2 sont les termes diagonaux de :

$$X_2' Y_1 = X_2' X_1 U V' = V \wedge U' U V' = V \wedge V'$$

Les termes de $\wedge$ étant positifs ou nuls, la matrice $X_2' Y_1$ est semi-définie positive et ses termes diagonaux sont positifs ou nuls. Le nuage moyen étant $(Y_1 + X_2)/2$, il faut aussi considérer les matrices $X_2'(Y_1 + X_2)/2$ et $Y_1'(Y_1 + X_2)/2$: la propriété ci-dessus assure la positivité de leurs termes diagonaux.

Ce résultat reste vrai si l'on fait subir aux nuages ajustés la même transformation orthogonale, ce qui se passe lorsque l'on projette sur les principales directions d'inertie du nuage moyen.

Exemple où les variables canoniques sont liées négativement

En revanche, en APG comme en AFM, on peut avoir des facteurs homologues corrélés négativement, ce que nous illustrons à l'aide d'un exemple (tableau 1).

TABLEAU 1
Trois individus (A,B,C) décrits par 3 groupes contenant chacun une variable

	$V1$	$V2$	$V3$
A	5	1	-3
B	-2	-2	-2
C	-3	1	5
Données			

	$V1$	$V2$	$V3$
$V1$	1		
$V2$	.40	1	
$V3$	$-.68$	.40	1
Corrélations			

Dans ce cas particulier ou chaque groupe ne comporte qu'une seule variable, l'APG considère les trois variables comme homologues (éventuellement en considérant leurs opposées) et la matrice des corrélations entre facteurs homologues est confondue (éventuellement, pour tenir compte des symétries, en changeant tous les signes d'une ou plusieurs lignes et des colonnes correspondantes) avec la matrice des corrélations entre variables initiales. Lorsqu'une variable est corrélée positivement avec deux variables elles-mêmes liées négativement entre elles (ce qui est le cas ici), il existe des facteurs homologues corrélés négativement.

Remarque. – Si, sur les données de l'exemple précédent on réalise une APG avec homothéties, alors le groupe 2 se voit affecter un coefficient 0 ce qui illustre au passage comment l'introduction d'homothéties permet d'exclure un groupe.

Mais l'introduction d'homothéties ne règle de façon satisfaisante le problème des corrélations négatives entre dimensions homologues que dans le cas unidimensionnel puisque l'homothétie s'applique uniformément à toutes les dimensions d'un groupe.

2.8. Résultats

Remarque préliminaire. – Lorsque le modèle procustéen est exactement vérifié (*i.e.* les N_I^j se déduisent l'un de l'autre par rotation ou symétrie), les deux méthodes fournissent le «bon résultat», à savoir une représentation superposée des N_I^j dans laquelle les points homologues sont confondus et les formes des N_I^j parfaitement respectées. Pour l'APG, c'est évident puisque le nuage moyen est identique à chaque N_I^j après rotation. Pour le cas de l'AFM, supposons 3 groupes tels que $X = (Z, ZA, ZB)$ avec $A'A = I_d$ et $B'B = I_d$; les composantes principales de l'AFM sont vecteurs propres de $X' = 3Z'Z$: ainsi le nuage moyen et chaque nuage partiel ont les mêmes composantes principales.

Dans la mise en évidence de facteurs communs, le cadre de l'APG, comparé à celui de l'AFM, est très contraignant puisqu'il suppose que :

- il existe K_c facteurs communs orthogonaux;
- les facteurs sont communs à tous les groupes.

Ces contraintes pèsent sur l'ensemble des résultats (en particulier sur les premiers facteurs) puisque l'on recherche un optimum global. Ainsi, lorsqu'il existe un facteur commun à certains groupes seulement.

- En AFM les N_I^j qui ne possèdent pas ce facteur sont orthogonaux à ce facteur; ils n'ont aucune influence sur lui.
- En APG, toute direction est nécessairement commune à tous les groupes et le facteur commun sera superposé avec des directions de certains groupes avec lesquelles il n'a rien à voir. Ainsi, dans ce cas, la mise en évidence même du facteur commun peut être perturbée : la configuration moyenne ne correspond pas à celle du facteur commun mais est déformée par les représentations des groupes qui n'ont rien à voir avec lui mais qui lui sont quand même superposées.

2.9. Aides à l'interprétation

Nous présentons successivement les principales aides de l'APG en précisant leur signification et le cas échéant leurs équivalents en AFM.

En APG la représentation superposée fournit un cadre dans lequel l'inertie totale de N_I^J peut être décomposée de multiples façons et induire un système exhaustif d'indicateurs. L'inertie totale (en pratique rendue égale à 100 même si, dans l'article original de Gower (1975) cette inertie est fixée à J) est d'abord décomposée en inertie inter (inertie du consensus) et inertie intra (inertie résiduelle). Ces trois inerties sont ensuite elles-mêmes décomposées de trois façons (un exemple de ces décompositions est donné § 3.3.3.).

2.9.1. Décomposition par dimension

- L'inertie inter indique l'importance relative des dimensions; on retrouve les valeurs propres de l'ACP de N_I; cet indicateur est le même que dans l'AFM.
- L'inertie intra indique le degré de «consensualité» de la dimension; cet indicateur est équivalent à celui de l'AFM, dans laquelle on rapporte l'inertie inter à l'inertie totale.

2.9.2. Décomposition par groupe

- L'inertie intra mesure la ressemblance entre N_I^j et N_I. En AFM, on calcule des coefficients de corrélation canoniques et les mesures Lg qui permettent d'évaluer la liaison entre N_I^j et N_I axe par axe (Rappel : la mesure Lg $(z, \{v_k, k = 1, K\})$ entre une variable z et un ensemble de variables v_k est l'inertie projetée des variables v_k sur z; *cf.* Escofier & Pagès, 1998 p. 161). La différence entre les méthodes porte ici sur deux points : la nature de l'indicateur et le fait de le calculer par axe ou globalement; ces différences s'inscrivent bien dans les optiques différentes des méthodes mais rien n'interdit d'introduire dans l'une les indices de l'autre, à l'exception des sommes d'inertie intra sur plusieurs axes qui n'ont pas de sens en AFM.

Remarque. – On peut se demander pourquoi l'inertie intra ne peut être cumulée axe par axe puisque ces axes sont orthogonaux dans R^K (même s'ils ne le sont pas dans R^{K_j}) : l'analyse réalisée étant une analyse «inter», dans la base des axes factoriels la matrice de variance interclasses est diagonale alors que la matrice de variance intraclasses n'a aucune raison de l'être.

- La décomposition de l'inertie inter n'a pas de sens ici.

2.9.3. Décomposition par individu

- L'inertie inter est la contribution des individus au nuage moyen; cet indicateur existe aussi en AFM (attention le nuage moyen n'a pas exactement le même sens) mais il est surtout utilisé axe par axe dans l'optique de l'analyse factorielle.
- L'inertie intra indique si l'individu est globalement l'objet d'un consensus ou non; en AFM, cet indicateur est calculé axe par axe mais ne peut pas être cumulé sur plusieurs directions; en APG il peut être calculé globalement et décomposé axe par axe.

2.9.4. En résumé

- les deux méthodes possèdent des systèmes d'indicateurs qui permettent de balayer exhaustivement les thèmes qui surgissent lors de l'examen d'une représentation superposée;
- en AFM il n'est toutefois pas possible de cumuler les inerties projetées des N_I^j sur plusieurs axes;
- en APG, le caractère forcé de la superposition, qui pèse sur le consensus, pèse par voie de conséquence sur les indicateurs ce qui sera illustré par un exemple (§ 3).

2.10. Représentation des variables

Dans les deux méthodes, on calcule les coefficients de corrélation entre les variables initiales et les dimensions du nuage moyen. Ces coefficients sont représentés graphiquement comme en ACP.

Remarquons au passage que, en APG, cette représentation ne correspond à une projection du nuage N_K de l'AFM que si les nuages N_I^j sont projetés sur les axes principaux de N_I et non sur ceux de N_I^J (ce qui est un inconvénient pour cette option déjà évoquée au § 1.2.2.).

Ici la différence entre les deux méthodes tient au fait qu'en AFM les variables initiales jouent un rôle actif direct (*via* les liaisons intra et inter groupes) dans la représentation du nuage moyen alors qu'en APG elles interviennent de façon indirecte. Il en résulte que :

- la représentation des variables possède sa propre optimalité en AFM ce qui n'est pas le cas en APG;
- la relation de transition qui exprime la coordonnée d'un individu moyen en fonction des coordonnées des variables n'existe pas en APG;

- *a fortiori* la relation qui en AFM exprime la coordonnée d'un individu partiel (i^j) en fonction des coordonnées des variables n'existe pas en APG.

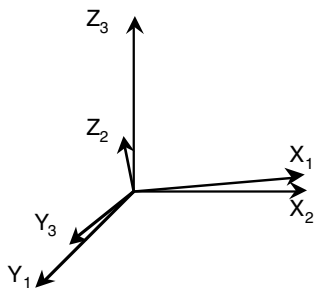
3. Étude d'un jeu de données choisies

3.1. Données

4 individus (A,B,C,D) sont décrits par 3 groupes comportant chacun 2 variables. Les données sont rassemblées dans le tableau 2 et illustrées figures 2 et 3.

TABLEAU 2
Jeu de données choisies

	groupe 1		groupe 2		groupe 3	
	$X1$	$Y1$	$X2$	$Z2$	$Y3$	$Z3$
A	6	6	6	-2	3	-6
B	6	-6	6	2	-3	6
C	-6	6	-6	2	3	6
D	-6	-6	-6	-2	-3	-6



Les variables, étant centrées, se situent dans un sous-espace de dimension 3 *ce qui les rend représentables.*

Ici, les coefficients de corrélation valent soit 0 *soit* 1.

Les variables parfaitement corrélées (ex. X_1 *et* X_2 *) sont légèrement écartées afin d'être représentées distinctement.*

FIGURE 2
Jeu de données choisies : représentation des 6 variables dans l'espace des variables R^4

Les 6 variables sont centrées. Elles sont construites à partir de 3 variables X, Y et Z, non corrélées deux à deux et de variance 1, qui ont été multipliées par 2, 3 ou 6. Les variables ne sont pas réduites pour obtenir des directions d'inégales inerties. En procédant ainsi, chaque groupe présente un facteur commun avec chacun des deux autres; ces facteurs communs ne sont pas forcément associés à la même inertie d'un groupe à l'autre.

L'inertie maximum est la même dans chaque groupe ce qui élimine l'influence de la pondération de l'AFM : les deux analyses opèrent sur les mêmes données.

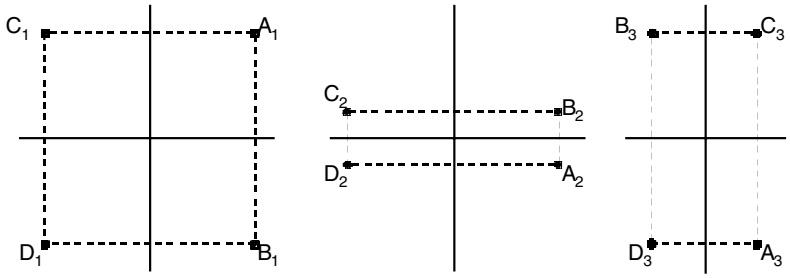


FIGURE 3
Jeu de données choisies : représentation des quatre individus pour chacun des trois groupes

Ces données, ainsi que les suivantes, ont été traitées par les logiciels SPAD (2002) pour l'AFM et OP & P (1992) pour l'APG.

3.2. Résultats de l'AFM

3.2.1. Inerties projetées du nuage moyen

Elles sont : 2, 1.25, 10/9, 0, 0, 0 (en pourcentage : .459, .287, .255, 0, 0, 0). Trois dimensions d'importances comparables sont nécessaires (et suffisantes) pour représenter les données.

3.2.2. Mesures de liaison entre facteurs et groupes

Deux mesures sont utilisées et rassemblées dans le tableau 3, dans lequel on trouve à l'intersection du groupe j et du facteur s :

- le coefficient de corrélation (dit canonique) entre F_s et F_s^j ;
- la mesure de liaison Lg entre F_s et le groupe j.

TABLEAU 3
Liaisons entre variables générales et variables canoniques

	$F1$	$F2$	$F3$
Groupe 1	1	1	0
Groupe 2	1	0	1
Groupe 3	0	1	1
$r(F_s, F_s^j)$			

$F1$	$F2$	$F3$
1	1	0
1	0	1/9
0	1/4	1
$\mathtt{Lg}(F_s, j)$		

Les coefficients de corrélation canoniques indiquent que $F1$ est commun aux groupes 1 et 2, $F2$ aux groupes 1 et 3, $F3$ aux groupes 2 et 3. Les mesures Lg précisent, par exemple, que $F3$ correspond à la principale direction d'inertie du groupe 3 et à une direction d'inertie peu importante du groupe 2. On retrouve ici exactement la structure qui a servi de base pour construire les données.

3.2.3. Inerties des individus

Les 4 individus ont :

- la même contribution à l'inertie pour chaque facteur;
- la même inertie intra pour chaque facteur.

On retrouve ici la symétrie des individus bien visible sur les données.

3.2.4. Représentation superposée (Figure 4)

Cette représentation précise la nature des facteurs communs. Ainsi, le premier facteur oppose A et B d'une part à C et D d'autre part, opposition qui existe dans les groupes 1 et 2 et non dans le groupe 3.

Bien que N_I soit parfaitement représenté, les nuages N_I^j sont déformés. Cette déformation présente deux aspects décrits en annexe. Ainsi, par rapport aux N_I^j initiaux, les représentations des N_I^j dans l'AFM :

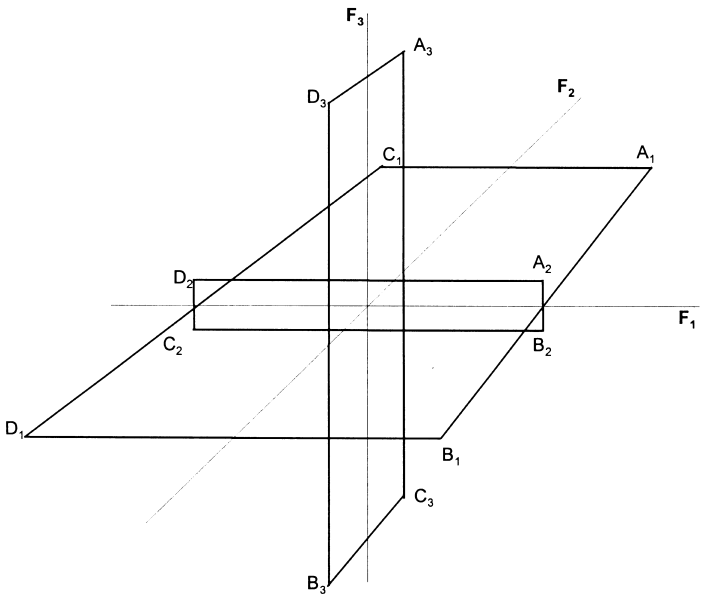


FIGURE 4 (début)

	Coordonnées		
	axe 1	axe 2	axe 3
A_1	-2.12	2.68	0.00
B_1	-2.12	-2.68	0.00
C_1	2.12	2.68	0.00
D_1	2.12	-2.68	0.00
A_2	-2.12	0.00	0.32
B_2	-2.12	0.00	-0.32
C_2	2.12	0.00	-0.32
D_2	2.12	0.00	0.32
A_3	0.00	0.67	2.85
B_3	0.00	-0.67	-2.85
C_3	0.00	0.67	-2.85
D_3	0.00	-0.67	2.85

Valeurs exactes et arrondies				
$\frac{1}{\sqrt{10}}$	$\frac{3}{2\sqrt{5}}$	$\frac{3}{\sqrt{2}}$	$\frac{6}{\sqrt{5}}$	$\frac{9}{\sqrt{10}}$
0.32	0.67	2.12	2.68	2.85

FIGURE 4 (fin)
Jeu de données choisies : représentation superposée issue de l'AFM. Les coordonnées des N_I^j sont multipliées par le nombre de groupes actifs J (ici égal à 3) de façon à ce que le nuage global N_I soit au centre de gravité des N_I^j.

- subissent une homothétie de rapport $1/\sqrt{\lambda_s}$ le long de chaque facteur, associé à la valeur propre λ_s, du nuage N_I. Ainsi, le nuage N_I^1 a initialement la même inertie dans toutes les directions mais en projection est plus allongé le long de F2. Pour ce premier groupe de cet exemple, le calcul numérique est aisé car les deux valeurs propres non nulles de son ACP séparée sont égales. Ainsi, entre les axes 1 et 2 de l'AFM, le carré du rapport des coordonnées des individus partiels – coordonnées identiques en valeur absolue pour tous les points et pour chacun des axes 1 et 2 – est l'inverse de celui des valeurs propres de l'AFM. Soit, numériquement, puisque $\lambda_1 = 2$ et $\lambda_2 = 1.25$:

$$\left[\frac{F_2(i)}{F_1(i)}\right]^2 = \left[\frac{6}{\sqrt{5}}\,\frac{\sqrt{2}}{3}\right]^2 = \frac{\lambda_1}{\lambda_2} = \frac{2}{1.25}$$

- subissent une homothétie de rapport $\sqrt{\lambda_s^j}$ le long de leurs propres facteurs (λ_s^j étant la s^e valeur propre dans l'analyse de N_I^j). Ils sont en quelque sorte caricaturés. Ainsi, pour $j = 3$, le rectangle formé par N_I^j est plus étiré en projection qu'initialement (l'écart entre les deuxième et troisième valeurs propres de N_I est faible et joue peu ici). Pour ce troisième groupe de cet exemple, un calcul simple est encore possible car les facteurs de l'AFM coïncident, au rang près, avec ceux de son ACP séparée : l'axe 2 (resp. 3) de l'AFM coïncide avec l'axe 2 (resp. 1) de l'ACP séparée du groupe 3. Ainsi, entre les axes 2 et 3, le carré du rapport des coordonnées des individus s'exprime simplement en fonction des valeurs propres de l'AFM et de celles de l'ACP du groupe 3. Soit, numériquement, puisque $\lambda_1^3 = 1$, $\lambda_2^3 = 1/4$, $\lambda_2 = 1.25$ et $\lambda_3 = 10/9$:

$$\left[\frac{F_3(i)}{F_2(i)}\right]^2 = \left[\frac{9}{\sqrt{10}}\ \frac{2\sqrt{5}}{3}\right]^2 = \frac{\lambda_2}{\lambda_3}\ \frac{\lambda_1^3}{\lambda_2^3} = \left[\frac{5}{4}\ \frac{9}{10}\right] 4$$

3.3. Résultats de l'APG

3.3.1. Inerties projetées du nuage moyen

Par construction ce nuage est contenu dans un plan. Les pourcentages d'inertie projetée sont 63,5 % et 36,5 %.

3.3.2. Représentation des individus (Figure 5)

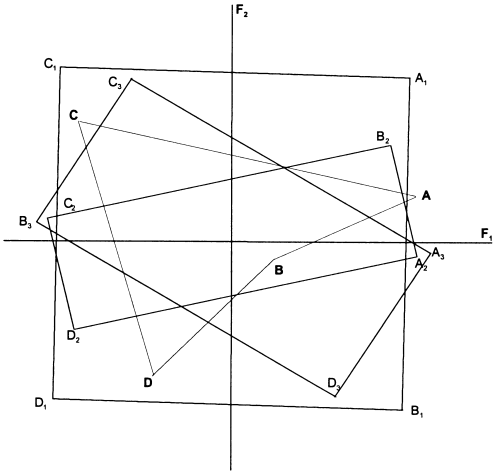


FIGURE 5
Jeu de données choisies : représentation superposée issue de l'APG

Chaque nuage N_I^j est parfaitement représenté. Bien évidemment, les points homologues ne se superposent pas exactement puisque les données ne respectent pas

le modèle procustéen. À la différence des N_I^j, le nuage moyen ne possède aucune symétrie : les individus y jouent des rôles différents. Ceci n'est évidemment pas conforme aux données. Cela étant, on retrouve les grandes lignes du premier plan de l'AFM, à savoir un premier axe qui oppose globalement $\{\mathrm{A}, \mathrm{B}\}$ à $\{\mathrm{C}, \mathrm{D}\}$ et un second qui oppose globalement $\{\mathrm{A}, \mathrm{C}\}$ à $\{\mathrm{B}, \mathrm{D}\}$.

3.3.3. Indicateurs d'adéquation au modèle procustéen

La décomposition de l'inertie totale en inertie inter et inertie intra est examinée globalement ou à son tour décomposée de différentes façons (tableau 4).

TABLEAU 4
Décompositions de l'inertie totale selon l'APG

décomposition globale	I. inter 58.3	I. intra 41.7	totale 100
par dimension			
1	37.0 (58 %)	26.2	63.3 (100 %)
2	21.3 (56 %)	15.5	37.7 (100 %)
par groupe			
1		13.3 (29 %)	45.9 (100 %)
2		9.6 (38 %)	25.5 (100 %)
3		18.8(65 %)	28.7 (100 %)
par individu			
A	20.6	4.4	25
B	1.4	23.6	25
C	21.4	3.6	25
D	15.0	10.1	25

L'inertie totale est fixée à 100.

Décomposition globale : elle met en évidence une inertie inter :

- sensiblement inférieure à 100 : le modèle procustéen est loin d'être vérifié;
- sensiblement supérieure à 100/3 : les trois nuages ont des éléments de structure en commun.

Décomposition par dimension : les éléments communs précités sont mis en évidence de la même façon par les deux dimensions.

Décomposition par groupe : le nuage moyen – qui s'interprète comme la structure commune mise en évidence – s'apparente surtout aux groupes 1 et 2 et beaucoup moins au groupe 3 (plus forte inertie intra pour ce groupe). Cette décomposition suggère que les groupes 1 et 2 se ressemblent plus entre eux qu'ils ne ressemblent au groupe 3. Ceci est visible en AFM (Tableau 2) et dans les données (figure 3 : la direction commune entre N_I^1 et N_I^2 est une direction d'inertie maximum).

Décomposition par individu : les différentes représentations des individus A et C sont très regroupées autour de leur centre de gravité; inversement celles de l'individu B sont très écartées. Ceci suggère que l'individu B joue un rôle particulier dans les données ce qui est contradictoire avec les inerties identiques des individus. De fait, si l'on permute dans la configuration de l'APG, d'une part les individus A et D et d'autre part les individus B et C, on obtient une solution aussi bonne que la précédente mais qui suggère cette fois que c'est l'individu C qui est très particulier.

3.4. Conclusion

Dans cet exemple de très faible dimension, le caractère contraignant du modèle pèse lourdement dans l'APG de données qui ne vérifient pas le modèle. Les spécificités des groupes gênent une mise en évidence claire des structures communes.

4. Étude d'un jeu de données choisi

4.1. Données

Nous reprenons les données de dégustation de vins dont le traitement par AFM a déjà été publié (Escofier & Pagès 1998). Ces données ont été recueillies par R. MORLAT et C. ASSELIN (INRA Angers) dans le cadre de leur essai terroir.

Les individus consistent en 21 vins du Val de Loire. Pour chaque individu on dispose de 27 variables quantitatives dont les valeurs sont les moyennes des notes attribuées par un jury de 36 dégustateurs. Les variables sont subdivisées en groupes qui correspondent aux phases de la dégustation :

1 Olfaction au repos :	**5** variables (ex. : intensité de l'odeur)
2 Vision :	**3** variables (ex. : intensité de la couleur)
3 Olfaction après agitation :	**10** variables (ex. : persistance aromatique)
4 Gustation :	**9** variables (ex. : astringence)

4.2. Résultats de l'AFM sur données centrées réduites

De l'examen des coefficients de corrélation canoniques de l'AFM (Tableau 5) il résulte que :

TABLEAU 5

Coefficients de corrélation canoniques de l'AFM entre les variables générales (en colonnes) et les variables canoniques associées (en lignes)

	$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$
$Gr1$	.888	.956	.887	.482	.419	.274	.422
$Gr2$	.926	.221	.158	.221	.176	.082	.212
$Gr3$	.969	.894	.896	.566	.659	.487	.464
$Gr4$	.950	.868	.299	.254	.517	.565	.424

- le premier facteur est commun aux 4 groupes;
- le deuxième est commun aux groupes 1,3 et 4;
- le troisième facteur est commun aux groupes 1 et 3.

L'interprétation de ces facteurs (analysés par ailleurs; Escofier & Pagès 1998 p. 139) est la suivante :

$F1$: puissance et harmonie;

$F2$: cas particulier de 2 vins présentant un «défaut» perceptible au nez et en bouche;

$F3$: opposition entre notes florales et notes fruitées.

4.3. Les analyses à comparer

Face à ce type de données, notre pratique consiste à réaliser une AFM du tableau centré réduit (on peut aussi choisir de ne pas réduire). La pratique anglo-saxonne consiste à réaliser une APG (avec ou sans homothéties) du tableau non réduit. Nous réalisons donc ces trois analyses. En pratique, les deux APG conduisant presque exactement aux mêmes résultats, nous ne considérons plus que celle sans homothéties.

Les différences entre ces analyses peuvent provenir d'une part des méthodes elles-mêmes et d'autre part de la transformation des données. Aussi réalisons-nous en complément :

- une AFM sur le tableau non réduit;
- une APG sur le tableau réduit;
- une APG sur le tableau réduit et pondéré comme en AFM (la variance de chaque variable du groupe j est égale à $1/\lambda_1^j$ afin de rendre égales les inerties axiales maximum de chaque groupe; rappel : λ_1^j est la première valeur propre de l'ACP séparée du groupe j).

Nous utilisons les abréviations suivantes :

	sigle	méthode	données
1	$afm1$	AFM	réduites
2	$afm2$	AFM	non réduites
3	$apg1$	APG sans homothéties	non réduites
4	$apg2$	APG sans homothéties	réduites
5	$apg3$	APG sans homothéties	réduites et pondérées façon AFM

4.4. Comparaison entre les variables générales

Le tableau 6 rassemble les coefficients de corrélation entre les variables générales (de même rang) des différentes analyses.

TABLEAU 6
Coefficients de corrélation entre les représentations moyennes de 5 analyses

	Facteur 1			
	$afm1$	$afm2$	$apg1$	$apg2$
$afm1$ (red.)				
$afm2$ (n. red.)	1.00			
$apg1$ (n. red.)	.99	1.00		
$apg2$ (red.)	1.00	1.00	.99	
$apg3$ (red.+afm)	1.00	1.00	.99	1.00

	Facteur 2			
	$afm1$	$afm2$	$apg1$	$apg2$
	.99			
	.98	.99		
	1.00	.99	.99	
	1.00	.99	.99	1.00

	Facteur 3			
	$afm1$	$afm2$	$apg1$	$apg2$
$afm1$ (red.)				
$afm2$ (n. red.)	.80			
$apg1$ (n. red.)	.89	.68		
$apg2$ (red.)	.97	.74	.95	
$apg3$ (red.+afm)	.97	.72	.93	1.00

	Facteur 4			
	$afm1$	$afm2$	$apg1$	$apg2$
	.73			
	.41	.57		
	.42	.58	.97	
	.44	.61	.96	.99

Les deux premières variables générales sont pratiquement identiques entre toutes les analyses. La troisième variable générale est globalement stable d'une analyse à l'autre. À ce niveau la différence entre AFM et APG est très faible lorsque l'on réduit et notable lorsque l'on ne réduit pas. La quatrième variable générale :

- est stable d'une APG à l'autre;
- relativement stable entre les deux AFM;
- variable entre AFM et APG.

Pour rendre compte de ces différences, il est raisonnable d'invoquer les différences entre les modèles sous-jacents aux deux types d'analyses. Mais ces différences sont, dans ce cas, sans répercussions pratiques puisqu'elles se réfèrent à des directions de très faible inertie. La stabilité entre les 3 APG attire l'attention; toutefois cette quatrième dimension n'a pu être interprétée à l'aide des corrélations avec les variables initiales.

4.5. Comparaison entre les pourcentages d'inertie

TABLEAU 7
Pourcentages d'inertie de la représentation moyenne dans les 5 analyses

	$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$	cumul
$afm1$ (red.)	49.38	19.49	8.78	5.31	3.86	2.89	2.51	92.21
$afm2$ (n. red.)	54.09	21.19	7.13	4.27	3.46	2.10	1.72	93.96
$apg1$ (n. red.)	78.31	13.94	2.87	2.14	.94	.66	.51	99.37
$apg2$ (red.)	65.43	19.87	6.67	3.50	1.77	1.25	.98	99.31
$apg3$ (red.+afm)	65.61	20.05	7.23	3.18	1.67	.97	.72	99.43

La comparaison entre les pourcentages d'inertie (Tableau 7) des différentes analyses fait apparaître deux phénomènes :

- à méthode égale, le premier pourcentage d'inertie est plus grand lorsque l'on ne réduit pas les données; ce résultat tient au fait que, dans ces données, les variables qui composent le premier axe sont aussi celles qui ont les plus grandes variances;
- à transformation des données égales, l'APG fournit un premier pourcentage d'inertie plus grand que celui de l'AFM. Ceci tient aux différents modes de représentation entre les deux méthodes.

Rappelons que, dans les deux méthodes, la configuration moyenne contient les points moyens des individus homologues dans les représentations partielles. Il en résulte que :

- si un facteur est commun à tous les groupes, la structure commune est bien traduite dans la configuration moyenne;
- si un facteur est commun à quelques groupes seulement, la structure commune est contractée dans la configuration moyenne puisque cette dernière prend en compte les représentations des groupes qui ne présentent pas cette structure commune.

Ceci explique que sur ces données l'APG conduit à un premier pourcentage d'inertie très important : il est le seul à correspondre à un facteur commun aux quatre groupes. On retrouve ici la déformation du nuage consensus en APG pour un facteur qui n'est commun qu'à certains groupes. Ce phénomène joue moins dans l'AFM, méthode qui amplifie la dispersion des représentations des N_I^j sur les axes associés aux faibles valeurs propres de l'analyse globale (*cf.* § 3.2.4.). Cette amplification, qui peut être vue par ailleurs comme un artefact, apparaît ici comme un moyen de préserver la représentation de N_I tout en bénéficiant d'une représentation superposée avec i situé à l'isobarycentre des i^j.

4.6. Comparaison globale des analyses

Nous représentons chaque analyse par les 7 premiers facteurs du nuage moyen. Pour comparer globalement ces ensembles de 7 facteurs, on calcule les coefficients RV (Escoufier, 1971) entre les représentations des vins qu'ils conduisent (*cf.* tableau 8).

TABLEAU 8
Coefficients RV entre les représentations moyennes de 5 analyses

	$afm1$	$afm2$	$apg1$	$apg2$	$apg3$
$afm1$	1	.991	.952	.986	.988
$afm2$		1	.966	.982	.985
$apg1$			1	.979	.975
$apg2$				1	.998
$apg3$					1

Ces coefficients mettent en évidence l'étroite parenté entre les cinq analyses; les différences d'inertie entre les dimensions constituent pratiquement les seules différences entre les résultats de ces analyses.

5. Conclusion

Fondamentalement AFM et APG sont des méthodes différentes qui n'ont pas les mêmes objectifs. En particulier l'APG construit une représentation des données dans un cadre très contraignant (chaque dimension est commune à tous les groupes). Il semble donc en première analyse que l'APG doit être réservée à des applications très particulières.

Concrètement, l'APG fournit une représentation exacte des N_I^j. Le nuage moyen n'est qu'un intermédiaire pour obtenir cette représentation superposée. Sa représentation peut exprimer imparfaitement la structure commune mise en évidence par la représentation superposée. De son coté l'AFM est axée sur la représentation du nuage moyen et en donne une représentation qui bénéficie pleinement de la dualité de l'analyse factorielle. En revanche la représentation des N_I^j subit des distorsions.

En pratique la recherche de dimensions communes à des groupes de variables s'effectue souvent sur des données comportant, outre quelques dimensions communes, un assez grand nombre de dimensions à la fois non communes et d'inertie peu importante. Dans de telles situations la contrainte de l'APG n'est pas gênante puisque les dimensions non communes qu'elle superpose concernent des dimensions d'inertie peu importante qui ne sont pas examinées. Ceci explique la convergence des résultats que l'on peut observer.

Annexe : deux propriétés de la représentation superposée des N_I^j en AFM

Notations complémentaires

Soit x_{ik} le terme général du tableau X de taille (I, K) juxtaposant en ligne les tableaux X_j. Soit m_k le poids affecté à la variable k, M la matrice diagonale de taille (K, K) contenant l'ensemble des m_k et M_j la matrice diagonale de taille (K_j, K_j) contenant les m_k des variables du seul groupe j. On suppose pour simplifier, mais sans nuire à la généralité, que la matrice des poids des individus est la matrice identité.

Soit $W_j = X_j M_j X_j'$ la matrice des produits scalaires (entre individus) associée au j^e tableau.

L'AFM repose sur une ACP du tableau X; on note u_s le vecteur unitaire du s^e axe d'inertie du nuage moyen ($u_s \in R^K$), F_s la composante principale associée, z_s la composante principale normée associée (F_s et $z_s \in R^I$), λ_s la valeur propre associée, et S le nombre de valeurs propres non nulles.

Soit Z la matrice de taille (I, I) dont les colonnes sont les vecteurs propres normés de XMX'; les S premières colonnes de Z sont les z_s. On a : $Z'Z = ZZ' =$ identité.

La coordonnée de l'individu partiel i^j le long de l'axe de rang s de l'AFM est noté $F_s(i^j) = F_s^j(i)$. F_s^j est le vecteur de ces I coordonnées. En outre, on note $G_s(k)$ la coordonnée de la variable k le long de l'axe s (dans R^I).

Enfin, λ_s^j représente la $\mathrm{s^e}$ valeur propre de l'ACP séparée du groupe j.

Reconstitution des N_I^j

La projection de N_I^j sur u_s , notée F_s^j , se déduit de z_s par (Escofier & Pagès, 1998, p. 164) :

$$F_s^j = \frac{1}{\lambda_s} W_j F_s = \frac{1}{\sqrt{\lambda_s}} W_j z_s$$

Soit $\{\sqrt{\lambda_s} F_s^j; s = 1, S\}$ l'ensemble des coordonnées des projections du nuage N_I^j (multipliées par $\sqrt{\lambda_s}$). L'ACP de la matrice de taille (I, S) ayant pour colonnes $\{\sqrt{\lambda_s} F_s^j; s = 1, S\}$ revient à faire celle du tableau $W_j Z$ (de taille (I, I) mais dont les $(I - S)$ dernières colonnes ne comportent que des zéros). Ce qui conduit à diagonaliser la matrice :

$$W_j Z Z' W_j = W_j W_j$$

dont les vecteurs propres sont les mêmes que ceux de W_j et les valeurs propres les carrés de celles de W_j.

Ainsi, l'ACP de $\{\sqrt{\lambda_s} F_s^j; s = 1, K\}$ conduit aux mêmes facteurs que celle de X_j, les valeurs propres de cette ACP étant les carrés de celles de X_j. En ce sens, l'ensemble des F_s^j permet une reconstitution de X_j. Cette reconstitution n'est en général pas parfaite même si le nuage N_I est parfaitement représenté dans l'AFM puisque, pour cette reconstitution :

[1] les facteurs F_s^j ont été au préalable multipliés par $\sqrt{\lambda_s}$;

[2] les valeurs propres de la reconstitution obtenue sont les carrés de celles de la représentation exacte.

Conséquences quant à l'interprétation de la représentation superposée

Le long de l'axe de rang s du nuage moyen, l'inertie de la projection de N_I^j est sujette à deux types d'artefacts. Elle a tendance à être amplifiée lorsque la direction de projection correspond :

- à une direction d'inertie faible du nuage N_I. (du fait de [1]); cette déformation est identique pour tous les N_I^j;
- à une direction d'inertie importante de N_I^j (du fait de [2]); cette déformation varie selon les N_I^j.

Relation de transition partielle

En contrepartie de cette déformation, la représentation superposée des N_I^j bénéficie d'une relation de transition, dite partielle, exprimant la coordonnée $F_s^j(i)$ d'un individu partiel i^j en fonction des coordonnées $G_s(k)$ des variables du groupe j. De la relation usuelle concernant les points moyens :

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} x_{ik} m_k G_s(k)$$

on déduit la relation concernant les points partiels :

$$F_s^j(i) = F_s(i^j) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K_j} x_{ik} m_k G_s(k)$$

Ainsi, la position des i^j s'interprète en relation avec celle des variables selon la règle usuelle : un individu apparaît du côté des variables pour lesquelles il a de fortes valeurs et à l'opposé de celles pour lesquelles il a de faibles valeurs.

Références

ARNOLD G. M. and WILLIAMS A. A. (1986), The use of generalized procustes techniques in sensory analysis. In *Statistical procedures in food research*. Pigott J. R. Eds. Elsevier applied sciences. 233-253

CARROLL J. D. (1968), A generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th annual convention of the American Psychological Association*. p 227-228.

DIJKSTERHUIS G. B (1995), *Multivariate data analysis in sensory and consumer science*. Thesis. Rijksuni Versiteit Leiden.

DIJKSTERHUIS G. B. and GOWER J. C. (1991/2), The interpretation of generalized Procustes analysis and allied methods. *Food Quality and Preference*, 3 p. 67-87.

ESCOFIER B. et PAGES J. (1998), *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*. 3e édition, Dunod. Paris.

GOWER J.-C. (1975), Generalized Procrustes Analysis. *Psychometrika*, 40, p. 33-51

GREEN B. F. (1952), The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17, p. 429-440.

LAFOSSE R. (1985), *Analyses procustéennes de deux tableaux*. Thèse de 3e cycle. Toulouse.

LANGRON S.P. (1983), The application of Procustes statistics to sensory profiling. In. *Sensory quality in Foods & Beverages. Definition measurement & control*. Williams & Atkin (Eds), Ellis Horwood Ltd, Chichester, p.89-95.

OP & P (1992), Procustes PC version 2.2. *A personal computer program for generalized procustes analysis*. Utrecht; OP & P Software Development.

PEAY E.R. (1988), Multidimensional rotation and scaling of configurations to optimal agreement. *Psychometrika*, 53, p. 199-208.

PIGGOTT J. R. (1986), *Statistical procedures in food research*. Elsevier

SAPORTA G. (1990), *Probabilités, analyse des données et statistique*. Technip, Paris.

SPAD. (2002), Système portable pour l'analyse des données. Logiciel diffusé par SPAD Groupe Test & GO Paris.

WILLIAMS A. and ARNOLD G.M. (1984), A new approach to the sensory analysis of foods and beverages. *In Progress in flavour research 1984. Proceedings of the* 4^{th} *Weurman flavour research symposium*. J. Adda (Ed.) Elsevier Amsterdam p. 35-50.