



**HAL**  
open science

## Deux méthodes d'analyse factorielle du lien entre deux tableaux de variables partitionnés

G Kissita, P Cazes, M Hanafi, R Lafosse

► **To cite this version:**

G Kissita, P Cazes, M Hanafi, R Lafosse. Deux méthodes d'analyse factorielle du lien entre deux tableaux de variables partitionnés. *Revue de Statistique Appliquée*, 2004, 52 (3), pp.73-92. hal-04474462

**HAL Id: hal-04474462**

<https://oniris.hal.science/hal-04474462>

Submitted on 23 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# REVUE DE STATISTIQUE APPLIQUÉE

G. KISSITA

P. CAZES

M. HANAFI

R. LAFOSSE

## **Deux méthodes d'analyse factorielle du lien entre deux tableaux de variables partitionnés**

*Revue de statistique appliquée*, tome 52, n° 3 (2004), p. 73-92

[http://www.numdam.org/item?id=RSA\\_2004\\_\\_52\\_3\\_73\\_0](http://www.numdam.org/item?id=RSA_2004__52_3_73_0)

© Société française de statistique, 2004, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

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

## DEUX MÉTHODES D'ANALYSE FACTORIELLE DU LIEN ENTRE DEUX TABLEAUX DE VARIABLES PARTITIONNÉS

G. KISSITA\*, P. CAZES\*\*, M. HANAFI\*\*\*, R. LAFOSSE\*\*\*\*

\* ISG, Université Marien Ngouabi, Brazzaville, gakissita@yahoo.fr

\*\* Lise-Ceremade, Université Paris 9-Dauphine, cazes@ceremade.dauphine.fr

\*\*\* ENITIAA-INRA, Nantes, hanafi@enitiaa-nantes.fr

\*\*\*\* LSP, Université Paul Sabatier, Toulouse 3, lafosse@cict.fr

### RÉSUMÉ

Deux ensembles  $X$  et  $Y$  de variables mesurées sur les mêmes individus sont respectivement formés de sous-ensembles de variables  $\{X_i\}$ ,  $i = 1, \dots, k_x$  et  $\{Y_j\}$ ,  $j = 1, \dots, k_y$ . Une première analyse, nommée *concorG*, revient à remplacer chacun des tableaux  $X_i$  et  $Y_j$  par un nombre réduit de variables latentes capables de résumer le lien de  $X$  avec  $Y$ . Le lien d'un tableau  $X_i$  avec un tableau  $Y_j$  constituant un lien partiel, les variables latentes de la deuxième analyse nommée *concorGM* permettent de comparer entre eux tous les liens partiels possibles. Par un choix de  $k_x + k_y$  métriques euclidiennes, ces deux analyses conduisent à des analyses canoniques, des analyses explicatives (ACPVI généralisée), ou des analyses procrustéennes et se prêtent à diverses possibilités d'applications.

**Mots-clés :** *Analyse factorielle, Corrélation entre multi-tableaux.*

### ABSTRACT

Two data sets  $X$  and  $Y$  measured on the same set of individuals are respectively formed with subsets of variables  $\{X_i\}$ ,  $i = 1, \dots, k_x$  and  $\{Y_j\}$ ,  $j = 1, \dots, k_y$ . A first analysis, named *concorG*, amounts to replace each of the subsets  $X_i$  and  $Y_j$  by few latent variables which can sum up the relationship between  $X$  and  $Y$ . The aim of the second analysis named *concorGM* is to investigate the  $k_x k_y$  relationships, between each  $X_i$  and each  $Y_j$ . By choosing  $k_x + k_y$  euclidean metrics, these two analyses lead to canonical, or redundancy, or procrustes analyses, and so to various applications.

**Keywords :** *Factor Analysis, Correlation between multi-sets.*

### 1. Introduction

Le nombre de circonstances où les données sont recueillies sous la forme de multi-tableaux va croissant, notamment pour les données sensorielles, écologiques, chimométriques, temporelles et d'imagerie. Dans ce papier les données correspondent à des sous-ensembles de variables qui sont appariés parce que toutes les variables considérées sont mesurées sur un même ensemble fixé d'individus. Ce type de données se prête à des analyses bien connues comme les analyses discriminantes, les analyses

des correspondances, les analyses de redondance ou ACPVI, les régressions PLS, ou les analyses procruste. Le type de données, des choix différents de métriques, et des façons différentes de choisir les orthogonalités entre solutions successives, font que les deux analyses proposées dans ce papier peuvent correspondre chacune à tous ces types d'analyses.

Cazes P. & coll. (1977) proposent une étude du lien entre deux séries de variables qualitatives en considérant des critères de covariances. Vivien M. (2002) considère l'étude d'un lien entre deux groupes d'ensembles de variables quantitatives, à partir d'un critère de covariance également. Dans cette dernière analyse nommée *acimoG*, les tableaux sont modifiés avant de recalculer une nouvelle solution, contrairement à ce qui est proposé par Cazes. Chacune de ces deux analyses correspond à une étude symétrique du lien. En variant la façon de modifier les tableaux, Vivien définit des régressions (*acimoG-PLS*), ne correspondant plus à une analyse symétrique du lien.

Les deux analyses du lien de plusieurs ensembles avec plusieurs autres ensembles proposées dans ce papier sont basées sur des critères de carrés de covariances. La nature symétrique ou non symétrique du lien dépend d'un choix de métriques fait au départ par l'utilisateur, en fonction de la nature de ses données et de ses préoccupations. Les façons de modifier les tableaux sont induites par ce choix.

Ces deux analyses sont des propositions d'extension d'analyses définies par Lafosse et Hanafi (analyse *concor*, 1997), Esposito et Scepi (1999), Balbi et Esposito (1999) ou Hanafi et Lafosse (développement de *concor*, 2001), où l'association de plusieurs ensembles de variables quantitatives n'est considérée qu'avec un seul autre ensemble. L'*acimo-PLS* de Vivien et Sabatier (2001), où un ensemble explique plusieurs autres ensembles, est fondée au premier pas sur la première solution de l'analyse *concor*, mais s'en éloigne notamment par des conditions d'orthogonalité différentes aux pas suivants.

La première des deux analyses proposée ici est une analyse du lien entre deux tableaux  $X$  et  $Y$ , nommée *concorG* (*concor* généralisée), permettant de détailler les contributions partielles des sous-tableaux formant chacun des deux tableaux  $X$  et  $Y$ .

La deuxième analyse, nommée *concorGM* (*concor* généralisée et multiple) permet de comparer l'intensité du lien partiel d'un sous-tableau de  $X$  avec un sous-tableau de  $Y$ , aux intensités de tous les autres liens partiels possibles entre sous-tableaux.

C'est l'analyse *concorGM* qui est considérée pour traiter une application de type analyse discriminante, puis une application de type analyse procruste. Les résultats sont confrontés aux résultats fondés sur le critère de l'*acimoG* et sur celui de *concorG*. Les programmes ont été réalisés en langage R et en langage *scilab*.

## 2. Analyse *concorG*

L'analyse *concorG* est une généralisation de l'analyse *Concor* introduite par Lafosse et Hanafi (1997). Dans *Concor* on analyse le lien entre deux ensembles  $X$  et  $Y$  de variables centrées mesurées sur les mêmes individus, sachant que l'un des deux tableaux est partitionné en sous-ensembles de variables. Il s'agit alors d'évaluer la contribution de chaque sous-tableau de la partition. Ainsi le tableau de données  $Y$

étant de dimension  $n \times q$ , le tableau de données  $X$  de dimension  $n \times p$  est formé des sous-tableaux  $X_i$ ,  $i = 1, \dots, k_x$  de dimensions respectives  $n \times p_i$ , avec  $\sum p_i = p$ .

Par des changements de métriques, plusieurs types de lien peuvent alors être considérés (Hanafi et Lafosse, 2001). L'analyse ConcorG se ramenant (du moins au premier pas) à une analyse interbatterie de Tucker (1958) écrite avec des métriques quelconques (Lafosse, 1997), comprend donc comme cas particuliers les analyses classiques, telle que l'analyse canonique de deux tableaux, ou telle que l'ACPVI, et constitue une généralisation de ces techniques. Pour être général, on note  $N$  la métrique euclidienne sur l'espace d'individus  $R^q$ , et  $M_i$  les métriques euclidiennes respectives sur les espaces d'individus  $R^{p_i}$ . La métrique  $M$  considérée sur  $R^p$  est la métrique des blocs diagonaux  $M_i$ , les autres blocs étant nuls.

Certaines matrices de l'exposé sont alors considérées comme matrices d'applications linéaires entre espaces métriques, et par suite une décomposition en valeurs singulières d'une matrice  $A$  considérée comme application linéaire de  $(R^q, N)$  dans  $(R^p, M)$  et notée

$$A = U\Delta V',$$

correspond à une écriture où les colonnes de  $U$  sont les vecteurs singuliers à gauche  $M$ -orthonormés et où celles de  $V$  sont les vecteurs singuliers à droite  $N$ -orthonormés, les valeurs singulières se trouvant sur la diagonale de la matrice diagonale  $\Delta$  (Greenacre, 1984). Quand il s'agira uniquement de la décomposition en valeurs singulières usuelle ( $M$  et  $N$  sont alors les métriques identité), cela sera indiqué dans l'exposé.

La métrique diagonale des poids affectés aux individus définie sur  $R^n$  est notée  $D$ .

Pour éviter de trop alourdir les écritures, on convient ici et par la suite d'utiliser des mêmes notations pour désigner les vecteurs à calculer apparaissant dans l'expression d'un critère à optimiser, et les vecteurs solutions correspondants.

La première solution de l'analyse Concor consiste à définir  $k_x + 1$  combinaisons linéaires des variables des tableaux respectifs  $X_i$  et  $Y$ , notées  $X_i M_i u_{i1}$  et  $Y N v_1$  ( $u_{i1} \in R^{p_i}$  et  $v_1 \in R^q$ ). Sous contraintes de normes égales à 1 des  $k_x + 1$  vecteurs  $u_{11}, u_{21}, \dots, u_{k_x 1}$  et  $v_1$ , ces  $k_x + 1$  composantes sont solutions du critère à maximiser

$$f(u_{11}, u_{21}, \dots, u_{k_x 1}, v_1) = \sum_{i=1}^{k_x} \text{cov}^2(X_i M_i u_{i1}, Y N v_1). \quad (1)$$

On note maintenant encore  $(u_{11}, u_{21}, \dots, u_{k_x 1}, v_1)$  ce premier  $(k_x + 1)$ -uple solution de Concor. Le deuxième  $(k_x + 1)$ -uple solution  $(u_{12}, u_{22}, \dots, u_{k_x 2}, v_2)$  est obtenu comme le premier, mais après avoir remplacé chacun des tableaux  $X_i$  par le tableau déflationné  $X_i - X_i M_i u_{i1} u'_{i1}$ . Cela revient alors à maximiser le critère (1) sous les mêmes contraintes de normes avec en plus les contraintes d'orthogonalité

$$u_{i2} \perp u_{i1}, \quad \forall i = 1, \dots, k_x.$$

On procède ainsi par déflations successives, pour obtenir les solutions suivantes, toutes orthogonales aux précédentes selon  $k_x$  orthogonalités respectives. Si on souhaite que

l'analyse du lien concerne tous les tableaux simultanément, le nombre de solutions calculées sera pris au plus égal à  $\min(\text{rang}(X_i))$ . Dans d'autres circonstances ce nombre pourra être plus élevé.

Le couple  $(u_1, v_1)$  étant le premier couple de vecteurs singuliers de la matrice  $X'DY$ , considérée comme celle d'une application linéaire de  $(R^q, N)$  dans  $(R^p, M)$ , l'obtention de la première solution s'effectue à partir de ce premier couple, les vecteurs  $u_{i1}$  étant les sous-blocs respectifs normés à 1 du vecteur  $u_1$  et à la solution on a

$$f(u_{11}, u_{21}, \dots, u_{k_x1}, v_1) = \text{cov}^2(XMu_1, YNv_1).$$

Une démonstration est donnée dans l'annexe 6.1.

Maintenant on suppose que le tableau  $Y$  est lui aussi partitionné en  $k_y$  sous-ensembles des variables  $Y_j$ ,  $j = 1, \dots, k_y$ , avec les métriques respectives  $N_j$  et on considère que la métrique  $N$  est la métrique bloc-diagonale des blocs diagonaux  $N_j$ . Voulant maximiser sous  $k_y + 1$  contraintes de norme ( $v_{j1} \in R^{q_j}$  et  $u_1 \in R^p$ ) le critère

$$g(v_{11}, v_{21}, \dots, v_{k_y1}, u_1) = \sum_{j=1}^{k_y} \text{cov}^2(Y_j N_j v_{j1}, XM u_1),$$

on obtient aussi la solution depuis le même premier couple  $(u_1, v_1)$  précédent, avec  $v_{j1}$  le  $j$ -ème sous-bloc normé à 1 de  $v_1$ . Finalement à la solution on peut écrire

$$\begin{aligned} \sum_{j=1}^{k_y} \text{cov}^2(Y_j N_j v_{j1}, XM u_1) &= \text{cov}^2(XM u_1, YN v_1) \\ &= \sum_{i=1}^{k_x} \text{cov}^2(X_i M_i u_{i1}, YN v_1) = \lambda_1^2, \end{aligned}$$

où  $\lambda_1$  est la première valeur singulière (avec les métriques associées) de la matrice  $X'DY$ .

Le calcul de la première solution de l'analyse concorG (Concor généralisée) consiste à réaliser ce double calcul pour obtenir les  $k_x + k_y$  composantes partielles  $X_i M_i u_{i1}$  et  $Y_j N_j v_{j1}$ , et les deux composantes globales  $XM u_1$  et  $YN v_1$ .

La propriété suivante montre que les composantes partielles de la première solution de concorG sont des représentants unidimensionnés de leurs tableaux respectifs pouvant caractériser le lien de  $X$  avec  $Y$ .

### Propriété 2.1

Soit  $Z_x$  le tableau ayant pour colonnes les  $k_x$  composantes partielles  $X_i M_i u_{i1}$ , et  $Z_y$  le tableau ayant pour colonnes les  $k_y$  composantes partielles  $Y_j N_j v_{j1}$ . Alors,  $(a_1, b_1)$  étant le premier couple singulier de la décomposition en valeurs singulières usuelle de  $Z_x^t D Z_y$ , on a

$$(Z_x a_1, Z_y b_1) = (XM u_1, YN v_1).$$

Une démonstration se trouve dans l'annexe 6.2.

La deuxième solution de *concorG* se calcule comme la première, mais après avoir déflationné chaque tableau  $Y_j$  et chaque tableau  $X_i$ , ce qui permet d'obtenir  $k_x + k_y + 2$  vecteurs,  $u_{i2}$ ,  $v_{j2}$ ,  $u_2$  et  $v_2$ , orthogonaux aux vecteurs respectifs de la première solution. On définit ainsi 2 nouveaux ensembles de composantes partielles  $\{X_i M_i u_{i2}\}$  et  $\{Y_j N_j v_{j2}\}$  associés aux deux nouvelles composantes globales  $X M u_2$  et  $Y N v_2$ . Les solutions successives sont calculées de la même manière, chaque fois après avoir à nouveau déflationné les  $k_x + k_y$  tableaux des deux partitions.

Parce que chacun des deux systèmes de vecteurs  $\{u_s\}$  et  $\{v_s\}$  est un système orthogonal, et parce que les composantes globales associées aux  $s = 1, \dots, S$  solutions successives vérifient

$$\text{cov}(X M u_s, Y N v_{s'}) = 0, \text{ pour } s \neq s',$$

les solutions successives correspondent à un découpage effectif du lien entre  $X$  et  $Y$ . La mise en correspondance de parts d'inertie distinctes issues de chacun des deux tableaux ( $\text{cov}(X M u_s, Y N v_s) \neq 0$ ), constitue ainsi une véritable analyse de ce lien. Cela peut être condensé en énonçant la propriété suivante, soulignant bien que l'ensemble des solutions de l'analyse *concorG* constitue une analyse du lien entre une part de  $X$  et une part de  $Y$  (parts extraites avec les projecteurs orthogonaux  $U U' M$  et  $V V' N$ ).

### *Propriété 2.2*

Soit  $U$  la matrice ayant pour colonnes les  $S$  vecteurs  $M$ -orthonormés  $u_s$ , et  $V$  la matrice ayant pour colonnes les  $S$  vecteurs  $N$ -orthonormés  $v_s$ . Alors les couples  $(u_s, v_s)$  sont tous les couples singuliers de la matrice  $(X M U U')' D (Y N V V')$ .

En réalité, comme cela a été dit, chacune des solutions de l'analyse *concorG* s'obtient depuis le calcul d'un premier couple singulier. Ce n'est plus le cas pour l'analyse suivante, nommée *concorGM* (*concor généralisée multiple*), qui nécessite un algorithme particulier.

## **3. Analyse *concorGM***

### **3.1. Le critère**

Toutes les notations précédentes sont reprises, les matrices  $X$  et  $Y$  étant toutes deux partitionnées. Les vecteurs solutions sont différents pour des analyses différentes, bien que notés de la même façon pour faciliter la correspondance entre les termes analogues.

Généralisant *Concor* (*Concor* est retrouvée quand l'un des deux tableaux n'est pas partitionné), l'analyse *concorGM* est fondée sur le critère  $k$  suivant, à maximiser

sous les  $k_x + k_y$  contraintes de norme, celles considérées pour le critère *concorG*,

$$k(u_{11}, \dots, u_{k_x 1}, v_{11}, \dots, v_{k_y 1}) = \sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \text{cov}^2(X_i M_i u_{i1}, Y_j N_j v_{j1}). \quad (2)$$

Quand  $X = Y$  et quand la partition est commune, ce critère devient le critère  $\Psi_3$  optimisé par Hanafi (1997, pp. 74), ici posé en dimension 1.

Le critère de l'acimoG (Vivien 2002) serait le critère  $k$  sous les mêmes contraintes, si ce n'était que les covariances ne sont pas élevées au carré en acimoG. L'auteur s'en sert pour réaliser des régressions de type PLS et le calcul des solutions successives ne correspond pas aux mêmes modifications de tableaux qu'en analyse *concorGM*. Quand  $X = Y$  et quand la partition est commune, le critère de l'acimoG devient le critère MAXBET (Ten Berge, 1986) posé en dimension 1. Finalement, l'analyse *concorGM* est bien différente de l'acimoG, mais les deux critères respectifs sont vraiment de même nature. C'est pourquoi dans le traitement de données du paragraphe 4 les valeurs des critères seront comparées.

### 3.2. Première solution

L'optimisation du critère (2) revient à poser celle du Lagrangien

$$L = \sum_{i=1}^{k_x} \sum_{j=1}^{k_y} (u'_{i1} M_i X'_i D Y_j N_j v_{j1})^2 + \sum_{i=1}^{k_x} \lambda_i (1 - u'_{i1} M_i u_{i1}) + \sum_{j=1}^{k_y} \mu_j (1 - v'_{j1} N_j v_{j1}).$$

L'annulation des  $k_x + k_y$  dérivées partielles de  $L$  par rapport aux vecteurs  $u_{i1}$  et  $v_{j1}$ , conduit aux  $k_x + k_y$  égalités

$$\sum_{j=1}^{k_y} \text{cov}(X_i M_i u_{i1}, Y_j N_j v_{j1}) X'_i D Y_j N_j v_{j1} = \lambda_i u_{i1}, \quad i = 1, \dots, k_x.$$

$$\sum_{i=1}^{k_x} \text{cov}(X_i M_i u_{i1}, Y_j N_j v_{j1}) Y'_j D X_i M_i u_{i1} = \mu_j v_{j1}, \quad j = 1, \dots, k_y.$$

Ces égalités, et la propriété 2.1, conduisent à proposer l'algorithme ci-après (la preuve de sa convergence est établie dans l'annexe 6.3).

### 3.3 Algorithme de calcul de la première solution

Choisir  $k_x$  vecteurs  $u_i^{(0)}$  respectivement  $M_i$ -normés, par exemple ceux donnés par la première solution de *concorG*. En prenant alors  $k = 1, 2, \dots$ , répéter la séquence :



\* Pour chaque  $j$ , calculer le premier vecteur singulier à gauche  $v_j^{(k)}$ ,  $N_j$ -normé (celui à droite étant  $I$ -normé), de la matrice  $q_j \times k_x$

$$Y'_j D \left[ X_1 M_1 u_1^{(k-1)} \ X_2 M_2 u_2^{(k-1)} \ \dots \ X_i M_i u_i^{(k-1)} \ \dots \ X_{k_x} M_{k_x} u_{k_x}^{(k-1)} \right]$$

\* Pour chaque  $i$ , calculer le premier vecteur singulier à gauche  $u_i^{(k)}$ ,  $M_i$ -normé (celui à droite étant  $I$ -normé), de la matrice  $p_i \times k_y$

$$X'_i D \left[ Y_1 N_1 v_1^{(k)} \ Y_2 N_2 v_2^{(k)} \ \dots \ Y_j N_j v_j^{(k)} \ \dots \ Y_{k_y} N_{k_y} v_{k_y}^{(k)} \right]$$

\* Calculer la valeur du critère  $k$ .

\* Arrêter l'algorithme quand l'accroissement obtenu pour cette valeur n'est plus sensible.

Nous définissons les solutions successives de *concorGM* comme la première, mais après déflations des tableaux  $X_i$  et des tableaux  $Y_j$ , comme cela est fait pour l'analyse *concorG*.

L'analyse *concorGM* est vraiment une analyse de lien entre les deux partitions car la valeur du critère dépend de la façon de partitionner les tableaux dès la première solution, même quand les métriques sont les métriques identité (ce qui n'est pas le cas pour le critère de *concorG*). Nous en déduisons un intérêt accru pour *concorGM*. Cependant les solutions calculées de *concorG* sont globales, alors que l'algorithme de *concorGM* peut converger vers des solutions locales. C'est pourquoi on a proposé d'initialiser cet algorithme en partant de la première solution de *concorG*.

## 4. Exemple d'applications

Les données sont issues d'une expérimentation conduite par le Syndicat des vins de Cahors (Laffargue et coll., 1983-1993).

Les 30 lignes de chaque tableau désignent 5x6 vins différents, la différence provenant de l'un des 5 porte-greffes possibles ou de l'un des 6 greffons possibles. On a effectué au moment de la récolte (en 1983, 1985, 1988, 1989 et 1990), 8 analyses physico-chimiques de ces 30 vins. Les 8 analyses physico-chimiques sont : **anthocyane** en mg/l, indice de réfraction du **rouge**, indice de réfraction du **jaune**, indice de **folin**, **degré** en alcool, extrait **sec**, **acidité totale** et **acidité volatile**. Chaque année est finalement représentée par un tableau 30x8 de colonnes centrées et réduites.

### 4.1. Une analyse de type discriminante

On veut savoir si les variables physico-chimiques peuvent discriminer les vins selon leur appartenance aux 6 catégories de greffons, ou bien plutôt selon leur appartenance aux 5 catégories de porte-greffes. Les 2 hypothèses de sous-groupes sont considérées simultanément car on veut désigner laquelle des deux s'avère dominante par rapport à l'autre, en les confrontant. De plus, pour ne pas

privilégier une année plutôt qu'une autre, les 5 années participent simultanément à l'étude.

Finalement les données sont ici constituées de 2 tableaux d'indicatrices  $Y_1$ ,  $30 \times 5$ , et  $Y_2$ ,  $30 \times 6$ , et de 5 tableaux d'explicatives  $X_i$ ,  $i = 1 \dots 5$ ,  $30 \times 8$ . Les 30 vins sont les 30 individus communs aux 7 tableaux.

L'analyse *concorGM*, l'analyse *concorG* et l'analyse fondée sur le critère de l'*acimoG*, sont réalisées en munissant les 7 espaces d'individus des métriques de Mahalanobis respectives, de sorte que le critère devient un critère somme de carrés de corrélations pour *concorGM*. Les 10 carrés de corrélations d'une solution relatifs aux 10 couples  $(Y_j, X_i)$  sont alors interprétés en termes de pouvoirs discriminants.

TABLE 1  
*Valeurs du critère concorGM pour les 4 solutions et les 3 analyses*

sol. n°	1	2	3	4
concorGM	4.19	4.03	2.53	1.95
concorG	4.13	3.86	2.31	1.75
acimoG	3.98	3.66	2.23	1.73

TABLE 2  
*Valeurs du critère acimoG pour les 4 solutions et les 3 analyses*

sol. n°	1	2	3	4
concorGM	5.73	5.20	3.72	2.24
concorG	5.80	5.29	4.44	3.12
acimoG	6.17	5.71	4.55	3.61

Quatre solutions successives sont calculées. Quatre est ici la valeur du rang de  $Y_1$  alors que le rang de  $Y_2$  est égal à 5, ceux des  $X_i$  valant 8. Donc 4 est le nombre maximal de solutions à calculer pour permettre de confronter des 2 hypothèses. Pour chaque solution de chaque analyse, la valeur du critère *concorGM* considéré dans la table 1 est la somme de 10 carrés de corrélation. Les 2 premières solutions de *concorGM* correspondent alors à 65 % de la dépendance analysée.

La table 2 établit la comparaison entre ces analyses du point de vue de l'*acimoG*, le critère somme des 10 corrélations étant celui alors calculé (avec 59 % de la dépendance extraite par les 2 premiers axes, selon l'*acimoG*).

La table 3 contient les parts relatives du critère *concorGM* de chacune des 2 premières solutions, parts prises par chacun des 10 couples  $(Y_j, X_i)$ .

Dans l'ensemble, les corrélations sont plus élevées avec  $Y_1$  qu'avec  $Y_2$ . Pour l'année 1988, la première solution indiquerait plutôt le contraire, mais la

TABLE 3

*L'intérieur de la table contient les carrés des corrélations entre les composantes, pour les 2 premières solutions des 3 analyses*

	concorGM		concorG		acimoG		
	$Y_1v_{11}$	$Y_2v_{21}$	$Y_1v_{11}$	$Y_2v_{21}$	$Y_1v_{11}$	$Y_2v_{21}$	
$X_1u_{11}$	<b>.75</b>	.14	<b>.76</b>	.12	<b>.52</b>	.34	1983
$X_2u_{21}$	<b>.73</b>	.07	<b>.71</b>	.10	<b>.52</b>	.31	1985
$X_3u_{31}$	.34	.46	<b>.54</b>	.20	.38	.38	1988
$X_4u_{41}$	<b>.83</b>	.03	<b>.76</b>	.09	<b>.49</b>	.29	1989
$X_5u_{51}$	<b>.83</b>	.01	<b>.81</b>	.03	<b>.65</b>	.09	1990
	$Y_1v_{12}$	$Y_2v_{22}$	$Y_1v_{12}$	$Y_2v_{22}$	$Y_1v_{12}$	$Y_2v_{22}$	
$X_1u_{12}$	.19	.62	<b>.52</b>	.12	.36	.37	1983
$X_2u_{22}$	<b>.86</b>	.00	<b>.84</b>	.03	<b>.56</b>	.18	1985
$X_3u_{32}$	<b>.77</b>	.00	<b>.74</b>	.02	<b>.53</b>	.10	1988
$X_4u_{42}$	<b>.69</b>	.01	<b>.71</b>	.02	<b>.69</b>	.05	1989
$X_5u_{52}$	<b>.81</b>	.08	<b>.81</b>	.06	<b>.62</b>	.21	1990

deuxième solution montre que ce n'est pas vraiment le cas. Finalement l'analyse révèle nettement que les variables physico-chimiques considérées permettent la discrimination des porte-greffes bien plus que des greffons.

Pour faciliter la comparaison, les déflations respectives réalisées pour le calcul de la deuxième solution de chacune des 3 analyses sont du type proposé en analyse concorGM ou concorG. En ce sens, on s'est écarté de la véritable définition de l'acimoG. On peut noter que la différenciation entre porte-greffe et greffon est moins nette depuis les composantes de l'acimoG, comme on peut s'y attendre. En effet, pour le critère de l'analyse concorGM l'élévation au carré des corrélations produit la recherche d'une solution accentuant les écarts entre corrélations faibles et corrélations fortes. On remarque que les corrélations obtenues depuis l'analyse concorG se situent dans une position intermédiaire entre les valeurs de concorGM et de acimoG.

La discrimination trouvée en analyse concorGM est aussi révélée par les graphiques correspondants de la figure 1, en permettant de détailler la participation de chacun des vins. Pour la figure 1a (tableaux  $Y_1$  et  $Y_2$ ), les vins d'un même porte-greffe (resp. greffon) ayant le même codage dans  $Y_1$  (resp.  $Y_2$ ) sont projetés au même endroit désigné par le numéro du porte-greffe (resp. du greffon), sur les repères ( $v_{j1}, v_{j2}$ ) respectivement obtenus pour  $j = 1$  et 2. Sur la figure 1b (projections des 30 vins sur les repères ( $u_{i1}, u_{i2}$ )), on a codé chaque vin par son numéro de porte-greffe, tandis que sur la figure 1c, identique à 1b, chaque vin est codé par son numéro de greffon. Tous les  $7 \times 2$  axes de 1b ou 1c ont la même échelle.

Ainsi voit-on se correspondre assez bien les positions des porte-greffes représentés sur le graphique gauche de la figure 1a, avec celles des porte-greffes représentés sur chaque graphique de la figure 1b. Cela est très net pour le porte-greffe 5 quelle que soit l'année et donc ce porte-greffe a joué un rôle important dans la discrimination mise en évidence. Par contre la correspondance entre positions respectives s'établit beaucoup moins bien quand on veut associer les greffons du graphique droit de la figure 1a avec ceux de chacun des graphiques de la figure 1c.

La figure 1d contient les graphiques des variables de chaque tableau  $X_i$  projetées sur les repères respectifs des composantes partielles, d'aux des repères de la figure 1b ou 1c. Ces repères sont orthonormés car les métriques de Mahalanobis sont celles utilisées. Quand on passe d'une année à une autre année, on ne voit pas les positions des variables se correspondre. Bien que la discrimination par les analyses physico-chimiques se réalise assez bien pour les porte-greffes, ces graphiques ne révèlent pas l'existence de deux premiers facteurs qui seraient communs à toutes les années. Finalement on peut conclure que les facteurs discriminants des porte-greffes sont plutôt à évaluer séparément pour chaque année.

#### 4.2. Une étude d'analogie entre vins vieux et vins jeunes

En fait les mêmes analyses physico-chimiques des 30 vins précédents ont été répétées 5 ans après la récolte. Pour chacune des années 1983, 1985, 1988, 1989 et 1990, on dispose donc d'un nouveau tableau 30x8 de variables centrées réduites. Ces tableaux sont maintenant notés  $Y_j$ ,  $j = 1, \dots, 5$ . On se propose alors de comparer les analyses des 5 vins jeunes ( $X_i$ ,  $i = 1, \dots, 5$ ) avec celles des 5 vins vieux, voulant savoir quelles associations deux à deux peuvent prédominer. En considérant simultanément toutes les associations possibles, si certaines s'avèrent sensiblement plus fortes que d'autres, on pourra juger celles-ci significatives (plus facilement que par des analyses séparées). L'analyse conorGM est cette fois réalisée avec les métriques identité et le critère est un critère somme de 25 carrés de covariances. On considère ici que la structure associée à un tableau correspond à l'ensemble des corrélations entre variables de ce tableau. Un lien fort établi entre deux sous-tableaux va correspondre avec les métriques identité à l'existence d'un lien fort entre deux structures et révélera ainsi deux structures internes plutôt analogues.

Voulant prédire les vins vieux à partir des vins jeunes, on aurait adopté les métriques de Mahalanobis relatives aux tableaux des vins jeunes et les métriques identité pour les tableaux des vins vieux; mais ce type d'analyse n'a pas été réalisé dans ce papier.

Huit solutions successives peuvent être calculées, ce nombre correspondant au nombre de colonnes de chacun des 10 tableaux (et à leur rang). Les deux premières solutions de l'analyse conorGM correspondent ainsi à 92% du cumul des 8 valeurs du critère somme des 25 carrés de covariances (table 4), avec une première solution presque 4 fois plus importante que la deuxième.

Outre la valeur du critère conorGM qui est donnée dans la table 4 pour les 3 méthodes conorGM, conorG et acimoG, on donne les valeurs du critère acimoG dans la table 5. On notera que ces critères ne sont directement comparables qu'au

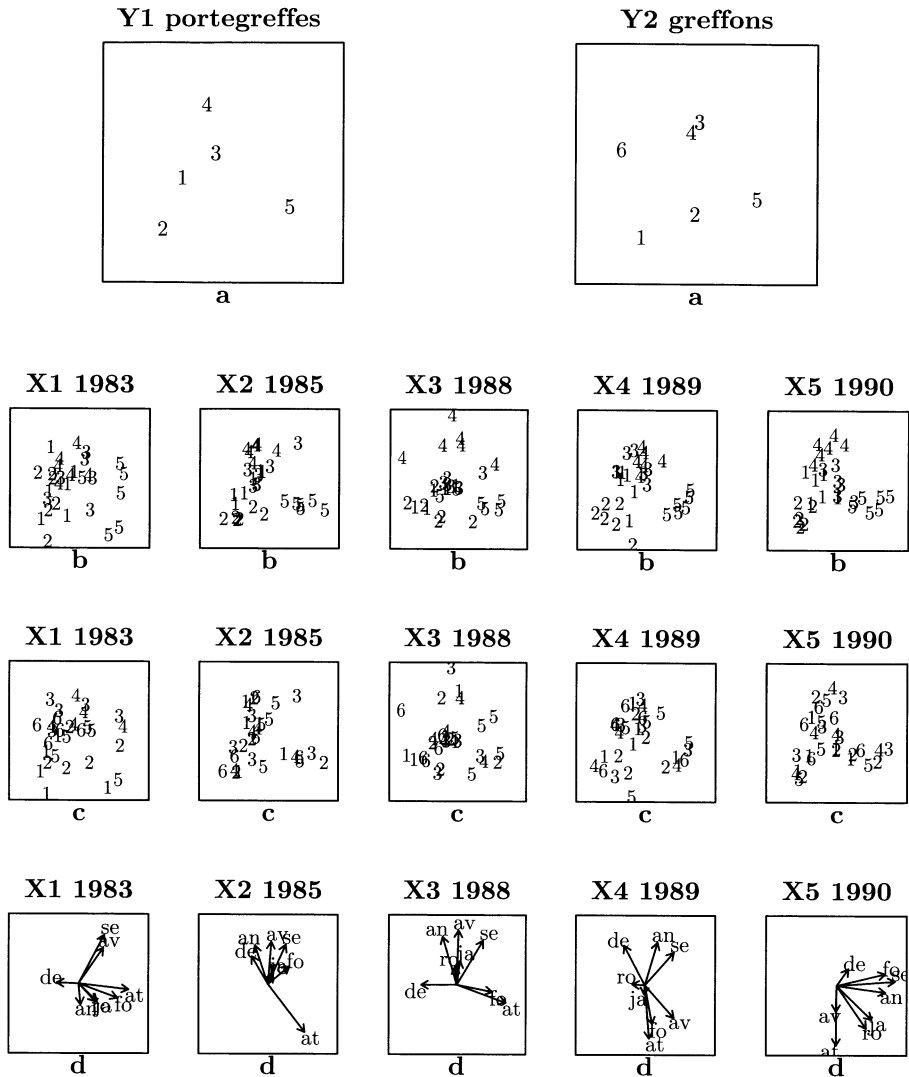


FIGURE 1

Représentation des 30 vins dans les repères  $(v_{j1}, v_{j2})$ ,  $j = 1, 2$  (ligne a) et dans les repères  $(u_{i1}, u_{i2})$ ,  $1 \leq i \leq 5$ , (lignes b et c). Les 30 vins sont symbolisés par leur porte-greffe (ligne a,  $Y_1$ , et ligne b), ou par leur greffon (ligne a,  $Y_2$ , et ligne c), la représentation des lignes b et c étant identique. En dernière ligne, les variables sont représentées dans les repères  $(X_i M_i u_{i1}, X_i M_i u_{i2})$ .

premier pas, du fait qu'aux pas suivants les tableaux déflationnés diffèrent suivant la méthode utilisée.

TABLE 4  
Valeurs du critère *concorGM* pour les 8 solutions des 3 analyses

solution n°	1	2	3	4	5	6	7	8
concorGM	80.8	20.9	7.16	1.4	.5	.3	.0	.0
concorG	76.0	26.5	10.0	1.9	.8	.3	.1	.0
acimoG	75.2	26.6	9.6	1.7	.9	.2	.0	.0

TABLE 5  
Valeurs du critère *acimoG* pour les 8 solutions des 3 analyses

solution n°	1	2	3	4	5	6	7	8
concorGM	36.5	19.3	9.3	5.0	2.5	1.3	.5	.2
concorG	39.1	23.0	14.0	6.1	3.5	2.0	1.0	.3
acimoG	39.2	22.9	14.0	5.9	3.6	2.1	.8	.2

TABLE 6

Carrés des corrélations entre les 2 composantes des 25 couples  $(X_i u_{is}, Y_j v_{js})$ . Les 5 premières lignes sont relatives à la première solution, les 5 lignes suivantes à la deuxième solution, orthogonale à la première de la même façon pour les 3 analyses. Les caractères gras mesurent l'association « vin jeune-vin vieux » d'une même année

concorGM					concorG					acimoG				
<b>.87</b>	.13	.00	.00	.06	<b>.80</b>	.20	.03	.02	.12	<b>.80</b>	.23	.04	.02	.13
.28	<b>.82</b>	.20	.16	.12	.32	<b>.81</b>	.28	.24	.21	.32	<b>.79</b>	.31	.26	.21
.00	.12	<b>.85</b>	.48	.09	.01	.17	<b>.82</b>	.49	.22	.01	.20	<b>.82</b>	.48	.23
.03	.16	.43	<b>.77</b>	.08	.10	.20	.44	<b>.76</b>	.18	.10	.22	.43	<b>.75</b>	.17
.07	.09	.16	.02	<b>.86</b>	.22	.21	.54	.19	<b>.76</b>	.18	.24	.52	.16	<b>.80</b>
<b>.74</b>	.04	.64	.29	.17	<b>.74</b>	.10	.61	.26	.24	<b>.76</b>	.06	.55	.20	.22
.13	<b>.73</b>	.08	.01	.02	.17	<b>.74</b>	.16	.03	.15	.15	<b>.76</b>	.12	.01	.11
.51	.00	<b>.82</b>	.25	.09	.47	.03	<b>.81</b>	.29	.11	.47	.01	<b>.81</b>	.30	.09
.37	.00	.56	<b>.46</b>	.29	.24	.07	.31	<b>.46</b>	.11	.19	.05	.27	<b>.49</b>	.06
.37	.03	.29	.29	<b>.69</b>	.29	.19	.35	.24	<b>.73</b>	.29	.13	.32	.19	<b>.73</b>

Les 5 premières lignes de la table 6 contiennent les valeurs  $\rho^2(X_i u_{i1}, Y_j v_{j1})$ , pour  $i = 1, \dots, 5$  et  $j = 1, \dots, 5$ , associées à la première solution. Les 5 dernières lignes contiennent les valeurs  $\rho^2(X_i u_{i2}, Y_j v_{j2})$  associées à la deuxième solution. Les

deuxièmes solutions des 3 analyses diffèrent du fait du critère et non pas du fait du type de déflations (on s'est donc encore ici écarté de la vraie définition de l'acimoG). Quelle que soit l'analyse et pour chaque solution, on observe des corrélations en général bien plus fortes quand il s'agit des couples  $(X_i u_{is}, Y_i v_{is})$ ,  $i = 1, \dots, 5$ , qui associent les mêmes années (en gras sur les diagonales). La faiblesse relative de presque tous les termes extra-diagonaux est évidente, notamment pour la première solution de l'analyse *concorGM*. Le fait de trouver ainsi des associations dominantes pour les vins de la même année implique que ces associations sont significatives. Par suite, envisager de prédire les analyses d'un vin vieux à partir des analyses du même vin faites au moment de la récolte semble concevable.

Pour repérer un vin, ou un porte-greffe, ou encore un greffon, responsable de cette association dominante, on pourrait juxtaposer les vins jeunes de l'année  $i$  projetés dans le repère  $(u_{i1}, u_{i2})$  avec les vins vieux de l'année  $i$  projetés dans le repère  $(v_{i1}, v_{i2})$ . Un même vin qui occupe la même position excentrée sur deux graphes juxtaposés contribue fortement à la dominance précédente, et encore plus fortement si cela se reproduit pour lui chaque année.

Notons qu'une autre analyse, celle effectuée avec les métriques de Mahalanobis, ne conduit pas à ces associations dominantes, et donc l'analogie mise en évidence entre vins de la même année n'apparaît qu'avec la prise en compte des structures internes aux sous-tableaux (avec les métriques identité).

## 5. Conclusion

Pour chaque solution de *concorG* ou de *concorGM*, des parts de variances et/ou des parts d'inertie sont mises en relation selon un critère d'association. Les solutions successives considèrent de nouvelles parts et le découpage ainsi induit définit la notion d'analyse. Un découpage relatif à chacun des sous-tableaux convient à l'étude du lien entre tous ces sous-tableaux. L'analyse *concorGM* étant effectivement une analyse entre les sous-tableaux, ce découpage est naturel. Par contre, la propriété 2.2 indique clairement que *concorG* analyse plus le lien entre  $X$  et  $Y$  que les liens entre les sous-tableaux, et de ce point de vue l'analyse *concorG* s'avère moins pertinente que *concorGM*.

Cependant, les traitements de données effectués indiquent que les premières solutions de *concorG* peuvent être proches des solutions de *concorGM*, dans la mesure où la valeur du critère *concorGM* calculée avec les solutions de *concorG* est élevée. Pour des raisons de rapidité de mise en oeuvre, ces solutions pourraient donc être recherchées à la place des solutions de *concorGM*. On a d'ailleurs ainsi pensé à initialiser l'algorithme de *concorGM* à partir du calcul de la première solution de *concorG*.

En outre, par construction et pour chaque solution de rang  $s$ , en analyse *concorG* il existe un lien entre les signes des vecteurs  $u_{is}$  et les signes des vecteurs  $v_{js}$  et pour cette raison les composantes de l'analyse *concorG* (et aussi celles de l'acimoG) pourraient être préférées à celles de l'analyse *concorGM*, notamment quand il s'agit de révéler les individus qui sont dans des positions analogues sur plusieurs graphiques juxtaposés.

Cependant, puisque ces signes sont totalement arbitraires en analyse *concorGM*, il est possible de les choisir. Pour une solution  $s$ , on peut par exemple choisir les signes de proche en proche, en considérant d'abord les vecteurs associés aux covariances  $cov(X_i M_i u_{is}, Y_j N_j v_{js})$  les plus élevées en valeur absolue, de sorte que ces covariances deviennent positives (en tout cas chaque fois que cela est possible, compte tenu des choix de signes déjà faits). Ainsi peut-on obtenir les graphiques de la figure 1. Il a été en effet possible de choisir les signes des vecteurs (et donc les orientations des axes) tels que toutes les covariances – en fait des corrélations – deviennent positives (par ailleurs on a constaté que les covariances calculées avec les composantes des deux premières solutions de *concorG* étaient aussi toutes positives, tout comme celles calculées en *acimoG*).

En prenant pour métriques  $M_i$  celles de Mahalanobis pour les tableaux respectifs  $X_i$ , et les métriques identité pour les tableaux  $Y_j$ , le critère *concorGM* devient un critère somme de variances expliquées et l'analyse devient une analyse de type régression. Il s'agit alors, parmi toutes les associations «tableau expliqué  $Y_j$  par tableau explicatif  $X_i$ », de désigner celles menant aux variances expliquées les plus élevées. Chaque système  $\{X_i u_{is}\}$ ,  $s = 1, 2, \dots$ , est alors formé de composantes explicatives de corrélations nulles deux à deux.

Cependant, dans certains cas (notamment quand le nombre des variables des tableaux  $X_i$  est très élevé) la mise en oeuvre de cette analyse devient incertaine. Pour éviter le calcul des métriques de Mahalanobis, on peut alors envisager une modification de l'analyse, recopiant une démarche proposée en régression *acimoG-PLS* par Vivien. Le critère d'association considéré pour calculer chaque solution est alors le critère *concorGM* de type procruste (toutes les métriques  $M_i$  et  $N_j$  sont les métriques identité), et les modifications successives des tableaux  $X_i$  sont différentes de celles proposées en analyse *concorGM*, car choisies pour obtenir des systèmes  $\{X_i u_{is}\}$ ,  $s = 1, 2, \dots$ , formés de composantes explicatives de corrélations nulles deux à deux. On obtient alors une régression «*concorGM-PLS*», qui peut se situer comme étant un compromis entre deux types d'analyse *concorGM*, le type procruste et le type régression décrit dans le paragraphe précédent. La différence entre *concorGM-PLS* et l'*acimoG-PLS* correspondante provient seulement du critère utilisé.

## 6. Annexe

Sans perte de généralité, nous faisons les trois démonstrations ci-après en supposant que les métriques euclidiennes sur les espaces d'individus sont les métriques identité.

### 6.1. Critère *concorG*

Il s'agit de montrer que, sous  $k_x + 1$  contraintes de norme, la maximisation du critère

$$f_1(u_1, u_2, \dots, u_{k_x}, v) = \sum_{i=1}^{k_x} cov^2(X_i u_i, Y v),$$



et, sous 2 contraintes de norme, celle du critère

$$f_2(a, b) = \text{cov}^2(Xa, Yb),$$

sont 2 optimisations relatives à un même problème avec à la solution,  $f_1 = f_2$  avec  $b = v$ , les  $k_x$  vecteurs  $u_i$  étant les  $k_x$  sous-blocs normés à 1 du vecteur  $a$ .

Le Lagrangien associé à la maximisation de  $f_2$  s'écrivant

$$L_2 = (a'X'DYb)^2 - \lambda(a'a - 1) - \mu(b'b - 1),$$

en annulant les dérivées partielles de  $L_2$  par rapport à  $a$  et  $b$ , et en posant

$$r = \text{cov}(Xa, Yb), \quad (3)$$

on obtient

$$r X'DYb = \lambda a \quad (4)$$

$$r Y'DXa = \mu b, \quad (5)$$

d'où l'on déduit immédiatement que

$$\lambda = \mu = r^2 \quad (6)$$

$$X'DYY'DXa = \lambda a \quad (7)$$

$$Y'DXX'DYb = \lambda b. \quad (8)$$

Le couple  $(a, b)$  est ainsi celui des vecteurs propres associés à la plus grande valeur propre des matrices  $X'DYY'DX$  et  $Y'DXX'DY$ .

Puisque  $\lambda = r^2$ ,  $a_i$  désignant le  $i$ -ème sous-vecteur de  $a$ , la restriction de (4) au bloc  $X_i$  s'écrit

$$X_i'DYb = r a_i. \quad (9)$$

Dans le cas de l'optimisation de  $f_1$ , le Lagrangien s'écrit

$$L_1 = \sum_{i=1}^{k_x} (u_i'X_i'DYv)^2 - \sum_{i=1}^{k_x} \lambda_i(u_i'u_i - 1) - \theta(v'v - 1).$$

Égalant à zéro les dérivées de  $L_1$  par rapport à  $u_i$  et  $v$ , et posant

$$r_i = \text{cov}(X_i u_i, Yv), \quad (10)$$

on a

$$r_i X_i'DYv = \lambda_i u_i \quad (11)$$

$$\sum_{i=1}^{k_x} r_i Y'DX_i u_i = \theta v. \quad (12)$$

On en déduit alors que

$$\lambda_i = r_i^2, \quad (13)$$

$$\theta = \sum r_i^2 = \sum \lambda_i. \quad (14)$$

Compte tenu de (13), l'égalité (11) s'écrit encore

$$X_i' DY v = r_i u_i, \quad (15)$$

ce qui, reporté dans (12), donne

$$\sum_{i=1}^{k_x} Y' D X_i X_i' DY v = \theta v. \quad (16)$$

Cette équation étant identique à (8) on en déduit

$$\theta = \lambda \quad (17)$$

$$v = b. \quad (18)$$

On déduit alors de (15) et de (9)

$$r_i u_i = r a_i. \quad (19)$$

Compte tenu de (13), (14) et (17), l'optimum de  $f_1$  est bien égal à celui de  $f_2$ , et on déduit de (19) que  $u_i$  est le  $i$ -ème sous-vecteur (normalisé) du vecteur  $a$ .

### Remarques

1) Pour maximiser le critère  $f_3$ , sous contraintes de norme des  $k_y + 1$  vecteurs  $u, v_1, \dots, v_j, \dots, v_{k_y}$

$$f_3(u, v_1, \dots, v_j, \dots, v_{k_y}) = \sum_{j=1}^{k_y} cov^2(Xu, Y_j v_j),$$

on suit le raisonnement ayant conduit à l'optimisation de  $f_1$ , après interversion des rôles de  $X$  et de  $Y$ . En posant

$$s_j = cov(Xu, Y_j v_j), \quad (20)$$

on obtient

$$u = a \quad (21)$$

$$s^2 = \sum_{j=1}^{k_y} s_j^2 = \lambda = r^2 \quad (22)$$

$$s_j v_j = r b_j. \quad (23)$$

$b_j$  étant le  $j$ -ème sous-vecteur de  $b$ .

2) Le couple  $(Xa, Yb)$  est le premier couple de Tucker (1958) des tableaux  $X$  et  $Y$ , car d'après (7) et (8) il vérifie les équations

$$XX'DYY'D(Xa) = \lambda Xa \quad (24)$$

$$YY'DXX'D(Yb) = \lambda Yb. \quad (25)$$

### 6.2. Démonstration de la propriété 2.1

Le couple  $(Xa, Yb)$  étant le premier couple de composantes de Tucker (1958) des tableaux  $X$  et  $Y$ , il faut montrer que  $c'$  est aussi le premier couple de composantes de Tucker des tableaux  $Z_x$  et  $Z_y$ . Il doit donc vérifier les équations (24) et (25), où  $X$  et  $Y$  sont à remplacer par  $Z_x$  et  $Z_y$  respectivement.

D'après les définitions de  $Z_x$  et  $Z_y$ , on a

$$Z_x Z'_x D Z_y Z'_y D(Xa) = \sum_{i=1}^{k_x} X_i u_i u'_i X'_i D \sum_{j=1}^{k_y} Y_j v_j v'_j Y'_j D Xa.$$

Or d'après (5), (6) et (23), on a

$$Y'_j D Xa = r b_j = s_j v_j,$$

d'où, puisque  $v'_j v_j = 1$

$$Z_x Z'_x D Z_y Z'_y D(Xa) = \sum_{i=1}^{k_x} X_i u_i u'_i X'_i D \sum_{j=1}^{k_y} s_j Y_j v_j.$$

Or

$$\sum_{j=1}^{k_y} s_j Y_j v_j = r \sum_{j=1}^{k_y} Y_j b_j = r Yb.$$

Donc

$$Z_x Z'_x D Z_y Z'_y D(Xa) = r \sum_{i=1}^{k_x} X_i u_i u'_i X'_i D Yb.$$

Comme d'après (9) et (19)

$$X'_i D Yb = r a_i = r_i u_i,$$

puisque  $u_i' u_i = 1$ , et compte tenu de (6) et de (19), on a

$$Z_x Z_x' D Z_y Z_y' D (Xa) = r \sum_{i=1}^{k_x} r_i X_i u_i = r^2 \sum_{i=1}^{k_x} X_i a_i = r^2 Xa = \lambda Xa,$$

ce qui montre que  $Xa$  est bien une composante de Tucker des tableaux  $Z_x$  et  $Z_y$ .

De la même façon on démontrerait que  $Yb$  est l'autre composante de Tucker de ces tableaux.

### 6.3. Convergence de l'algorithme *concorGM*

Afin de montrer la convergence de l'algorithme décrit au paragraphe 3.3, on montre que la valeur du critère *ConcorGM* augmente à chaque étape de l'algorithme. En effet, sachant les contraintes de norme à respecter, cette condition est suffisante (Hanafi, 1997).

La démonstration repose en effet sur le fait qu'en posant

$$f(u_1, u_2, \dots, u_{k_x}, v_j) = \sum_{i=1}^{k_x} \langle X_i u_i, Y_j v_j \rangle^2,$$

le critère *concorGM*

$$k(u_1, \dots, u_{k_x}, v_1, \dots, v_{k_y}) = \sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \text{cov}^2(X_i u_i, Y_j v_j) = \sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \langle X_i u_i, Y_j v_j \rangle^2$$

peut s'écrire

$$k(u_1, \dots, u_{k_x}, v_1, \dots, v_{k_y}) = \sum_{j=1}^{k_y} f(u_1, u_2, \dots, u_{k_x}, v_j). \quad (26)$$

Soient  $\{u_i^{(0)}\}$  et  $\{v_j^{(0)}\}$ , les  $k_x + k_y$  vecteurs tous normés et considérés comme une initialisation de l'algorithme. Les vecteurs  $\{u_i^{(1)}\}$  et  $\{v_j^{(1)}\}$ , étant les  $k_x + k_y$  vecteurs alors calculés lors d'une étape de l'algorithme 3.3, tout le problème consiste donc à établir que

$$\sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \langle X_i u_i^{(0)}, Y_j v_j^{(0)} \rangle^2 < \sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \langle X_i u_i^{(1)}, Y_j v_j^{(1)} \rangle^2. \quad (27)$$

Or, d'après cet algorithme, un vecteur  $v_j^{(1)}$  étant le premier vecteur singulier à gauche de la matrice

$$Y_j' D [X_1 u_1^{(0)} \ X_2 u_2^{(0)} \ \dots \ X_{k_x} u_{k_x}^{(0)}],$$

on a

$$\sum_{i=1}^{k_x} \langle X_i u_i^{(0)}, Y_j v_j^{(0)} \rangle^2 < \sum_{i=1}^{k_x} \langle X_i u_i^{(0)}, Y_j v_j^{(1)} \rangle^2.$$

La sommation sur  $j$  permet alors l'inégalité

$$\sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \langle X_i u_i^{(0)}, Y_j v_j^{(0)} \rangle^2 < \sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \langle X_i u_i^{(0)}, Y_j v_j^{(1)} \rangle^2. \quad (28)$$

La suite de l'algorithme indique qu'un vecteur  $u_i^{(1)}$  est le premier vecteur singulier à gauche de la matrice

$$X_i' D [Y_1 v_1^{(1)} \ Y_2 v_2^{(1)} \ \dots \ Y_{k_y} v_{k_y}^{(1)}].$$

Ainsi on a

$$\sum_{j=1}^{k_y} \langle X_i u_i^{(0)}, Y_j v_j^{(1)} \rangle^2 < \sum_{j=1}^{k_y} \langle X_i u_i^{(1)}, Y_j v_j^{(1)} \rangle^2.$$

La somme sur  $i$  de cette dernière égalité donnant

$$\sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \langle X_i u_i^{(0)}, Y_j v_j^{(1)} \rangle^2 < \sum_{i=1}^{k_x} \sum_{j=1}^{k_y} \langle X_i u_i^{(1)}, Y_j v_j^{(1)} \rangle^2,$$

permet alors de conclure grâce à l'inégalité (28) que l'inégalité (27) est vérifiée.

**Remerciement** : les auteurs remercient E.M. Qannari pour sa lecture attentive de l'article, ses critiques positives ayant permis d'améliorer la rédaction.

## 7. Bibliographie

- BALBI S., ESPOSITO V. (1999), *Rotated Canonical Analysis onto a Reference Subspace*. CSDA Journal, North Holland, vol. 32, 395-410.
- CAZES P., BAUMERDER A., BONNEFOUS S., PAGES J.P. (1977), *Codage et analyse des tableaux logiques. Introduction à la pratique des variables qualitatives*. Cahiers du BUREAU, 27, Paris VI, 3-47.
- ESPOSITO V., SCEPI G. (1999), *A non symmetrical generalised co-structure analysis for inspecting quality control data*. In M. Vichi and O. Optiz a cura di, Classification and Data Analysis. Theory and Application. Springer-Verlag.

- GREENACRE M.J. (1984), *Theory and applications of Correspondence Analysis*. Academic Press Inc.
- HANAFI M. (1997), *Structure de l'ensemble des analyses multivariées des tableaux de données à trois entrées : éléments théoriques et appliqués*. Thèse, Université Lyon I.
- HANAFI M., LAFOSSE R. (2001), *Généralisations de la régression linéaire simple pour analyser la dépendance de K ensembles de variables avec un K+1 ème*. Rev. Stat. Appliquée, 49, 1, 5-30.
- LAFFARGUE et coll. (1983-93), *Expérimentation porte-greffes, greffons du cépage Cot*. Maison des vins de Cahors. Ch. Agr. du Lot. CETA de Luzech.
- LAFOSSE R. (1997), *Analyse de concordance de deux tableaux : monogamies, simultanités et découpages*. Rev. Stat. Appliquée, 45, 3, 45-72.
- LAFOSSE R. et HANAFI M. (1997), *Concordance d'un tableau avec K tableaux : définition de K+1 uplés synthétiques*. Rev. Stat. Appliquée, 45, 4, 111-126.
- Logiciel R et logiciel scilab : <http://www.cict.fr/libre/>
- TEN BERGE J. (1986), *A general solution for the maxbet problem*. J. de Leuw, W. Heiser, J. Meulman & F. Critchley (Eds). Multidimensional Data Analysis.
- TUCKER L. R. (1958), *An interbattery method of factor analysis*. Psychometrika, 23, 111-136.
- VIVIEN M. et SABATIER R. (2001), *Une extension multi-tableaux de la régression PLS*. Rev. Stat. Appliquée, 49, 1, 31-54.
- VIVIEN M. (2002), *Approches PLS linéaires et non linéaires pour la modélisation de multi-tableaux : théorie et applications*. Thèse, Université Montpellier I.