

Université de Montréal

Analyse canonique : généralisation, influence et  
robustesse

par

Abdelaziz Khatouri

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)  
en Statistique

janvier 2002



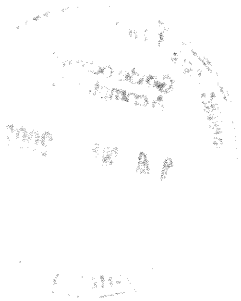
QA

3

U54

2002

v.010



Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Analyse canonique : généralisation, influence et  
robustesse**

présenté par

**Abdelaziz Khatouri**

a été évalué par un jury composé des personnes suivantes :

*Jean-François ANGERS*

---

(président-rapporteur)

*Robert Cléroux*

---

(directeur de recherche)

*Aziz LAZRAQ*

---

(co-directeur)

*Yves LEPAGE*

---

(membre du jury)

Mémoire accepté le:

*19 février 2002*

---

# SOMMAIRE

---

L'analyse canonique classique (AC) est une technique qui permet de réduire la dimension du problème étudié. Elle résume la relation entre deux groupes de variables aléatoires à seulement quelques nouvelles variables. Au lieu de considérer toutes les variables de chaque groupe, Hotelling (1935, 1936) a proposé de se limiter à l'étude de quelques combinaisons linéaires non corrélées des variables du premier groupe à d'autres combinaisons linéaires, aussi non corrélées, des variables du deuxième groupe. Il est un des premiers à chercher à identifier et à mesurer les relations existantes entre deux groupes de variables aléatoires. La première application de l'AC (Hotelling (1935, 1936)) était dans le domaine psychologique. Hotelling y a étudié les relations qui existent entre les tests mentaux (premier groupe de variables aléatoires) et les mesures physiques (deuxième groupe de variables aléatoires). Dans le cadre de ce mémoire on fait le tour de l'AC classique, puis on expose quelques généralisations (analyse canonique à plusieurs groupes de variables aléatoires) tel que celle de Carroll (1968), celle de Lazraq et al. (1992) et la généralisation de Gupta (1981). La dernière partie de ce mémoire sera consacrée à l'étude des fonctions d'influence des variables et corrélations canoniques; on y exposera aussi quelques méthodes d'estimation robuste et leurs applications à l'AC.

mots clés: analyse canonique, généralisation de Carroll, généralisation de Lazraq et al, généralisation de Gupta, fonction d'influence, estimation robuste.

## SUMMARY

---

Canonical analysis (CA) is a statistical method which reduces the dimension of a problem dealing with the relation between two groups of random variables. Instead of considering all the variables, Hotelling (1935, 1936) proposed limiting the problem to a few linear functions of each group. In each group the linear functions are uncorrelated whereas between the groups, they are highly correlated. Hotelling's application was related to mental tests and physical measures. In this master thesis a survey is made of canonical analysis, of its generalizations to several groups of random variables (Carroll (1968), Lazraq et al. (1992) and Gupta (1981)) of the role of influence functions of canonical variables and correlations and finally of the applications of robust estimation in CA.

key-word: Canonical analysis, generalization of Carroll, generalization of Lazraq et al, generalization of Gupta, influence functions, robust estimation.

# REMERCIEMENT

---

À l'issu de ce travail, je tiens à remercier mon directeur, Robert Cléroux, pour ses précieux conseils tout au long de la période de son encadrement. Je le remercie aussi pour son support financier et je le remercie de m'avoir soutenu pour avoir la bourse de fin d'études de l'Université de Montréal. Je remercie le Département de mathématiques et de statistique de l'Université de Montréal de m'avoir permis de mener des études de maîtrise en statistique à l'Université de Montréal. Je remercie Monsieur Aziz Lazraq pour l'aide qu'il ma apportée. Merci à tous les professeurs du Département de mathématiques et de statistique pour les efforts qu'ils ne cessent de déployer afin d'assurer aux étudiants une formation de bonne qualité. Merci à Nathalie Malo, à Stéphane Courchesne, coadministrateurs, pour leur aide et support en informatique. Merci à Monsieur Miguel Chagnon administrateur au Département.

# Table des matières

---

Sommaire.....	3
Summary.....	4
Remerciement.....	5
Liste des tableaux.....	10
Introduction.....	11
<b>Chapitre 1. Analyse canonique à deux groupes de variables aléatoires</b>	
13	
1.1. Formulation du problème.....	14
1.2. Variables et corrélations canoniques au niveau de la population....	15
1.2.1. Les premières variables et corrélation canoniques.....	16
1.2.2. Les deuxièmes variables et corrélation canoniques.....	18
1.2.3. Les jèmes variables et corrélation canoniques.....	21
1.2.4. Ensemble de toutes les variables et corrélations canoniques....	25
1.3. L'analyse canonique au niveau de l'échantillon.....	26
1.4. Propriétés échantillonnales des corrélations canoniques.....	28
1.4.1. La distribution de Wishart.....	28
1.4.2. La distribution de la statistique $\Lambda$ de Wilks.....	29
1.4.3. Distribution des corrélations canoniques.....	30
1.4.3.1. Cas où $\Sigma_{12} = 0$ .....	31
1.4.3.2. Cas où $\Sigma_{12} \neq 0$ .....	31

1.5. Exemple : Données CONSCRITS .....	32
1.5.1. Analyse des résultats de la sortie SAS .....	32
<b>Chapitre 2. L'analyse canonique généralisée .....</b>	<b>36</b>
2.1. Aperçu de la littérature sur sujet .....	36
2.2. Généralisation selon l'approche de Carroll. ....	37
2.2.1. Position du problème et sa résolution. ....	37
2.3. Exemple : Données sur les mouches <i>Leptoconops carteri</i> .....	40
2.4. Rappel des mesures de liaison : RLS, RI, RV et $RV_{reg}$ .....	43
2.4.1. La mesure de liaison de Lingoes et Schonemann : RLS. ....	43
2.4.2. Les propriétés de RLS .....	45
2.4.3. La mesure de liaison de Stewart et Love : $RI$ . ....	45
2.4.4. La mesure d'Escoufier : RV.....	47
2.4.5. Propriétés du RV .....	49
2.4.6. La mesure de Robert et Escoufier : $RV_{reg}$ . ....	49
2.4.7. Comparaisons entre RLS, RI, RV et $RV_{reg}$ .....	50
2.5. Analyse canonique généralisée au cas vectoriel via (RLS,RI) : généralisation de Lazraq, Cléroux et Kiers. ....	50
2.6. Analyse canonique généralisée au cas vectoriel via (RV, $RV_{reg}$ ), $RV - GCA$ ,: généralisation de Lazraq, Cléroux et Kiers.....	52
2.7. La $RV - GCA$ comme une application de la méthode IDIOSCAL appliquée aux matrices de projection .....	53
2.8. Généralisation de l'analyse canonique selon Gupta.....	56
2.8.1. Notation .....	57
2.8.2. Les premières new-GCV et new-GCC.....	58
2.8.3. Les new-GCV et new-GCC d'ordre quelconque. ....	63
2.8.4. Les new-GCV échantillonnales.....	64



<b>Chapitre 3. Analyse canonique: influence et robustesse</b> .....	67
3.1. Perturbation et fonction d'influence au niveau de la population. . .	67
3.1.1. Notion de perturbation.....	67
3.2. Les fonctions d'influence des corrélations et des variables canoniques.	71
3.3. Perturbation et fonction d'influence au niveau de l'échantillon. ....	72
3.3.1. Fonctions d'influence empiriques: IFE.....	73
3.3.2. Fonctions d'influence empiriques restreintes: $IFE_{(i)}$ .....	74
3.3.3. Fonctions d'influence échantillonnales: IFS .....	74
3.4. Méthodes robustes pour l'analyse canonique.....	75
3.4.1. Première approche: utilisation du "Minimum Covariance Determinant": MCD.....	76
3.4.2. Deuxième approche: utilisation des "signes" des données. ....	77
3.4.3. Troisième approche: utilisation de la régression alternée. ....	78
3.4.4. Simulation.....	79
3.4.5. Quatrième approche: M-estimateur .....	81
3.4.6. Cinquième approche: S-estimateur.....	83
<b>Conclusion</b> .....	84
<b>Programmes</b> .....	86
3.5. Programme <i>SAS</i> pour analyser les données CONSCRITS du chapitre 1 .....	86
3.6. Programme <i>Splus</i> pour calculer la solution de Carroll .....	86
3.7. Programme <i>Splus</i> pour la détermination de la solution de Lazraq, Cléroux et Kiers .....	88
3.8. Programme <i>Splus</i> pour calculer la solution de Gupta. ....	92
<b>Annexe</b> .....	96

**Bibliographie** ..... 99

## Liste des tableaux

---

1.5.1	Corrélations entre les variables de $X^{(1)}$ .....	32
1.5.2	Corrélations entre les variables de $X^{(2)}$ .....	32
1.5.3	Corrélations entre les variables de $X^{(1)}$ avec celles de $X^{(2)}$ .....	33
1.5.4	Analyse des corrélation canoniques .....	33
1.5.5	Valeurs-p pour $H_{01}$ , $H_{02}$ et $H_{03}$ .....	33
1.5.6	Corrélation de $U_1$ , $U_2$ et de $U_3$ avec $TAI$ , $TAP$ et $TAM$ . .....	34
1.5.7	Corrélation de $V_1$ , $V_2$ et de $V_3$ avec $PDS$ , $PDP$ et $PDM$ . .....	34
2.3.1	Valeurs propres de la matrice $\mathcal{Q}$ . .....	40
2.3.2	Les vecteurs propres de la matrice $\mathcal{Q}$ correspondants à $\lambda_1$ et $\lambda_2$ . .....	41
2.7.1	La matrice $A_1$ . .....	56
2.7.2	La matrice $A_2$ . .....	56
2.7.3	La matrice $A_3$ . .....	56
2.8.1	Les new-GCC possibles .....	65
2.8.2	Les new-GCV .....	66
3.4.1	Résultats de la simulation (obs.naber: pas de présence d'observations aberrantes; obs.aber: présence d'obsevation aberrantes). .....	80

# INTRODUCTION

---

L'analyse en composante principale étudie les relations existantes au sein d'un même groupe de variables aléatoires. L'analyse canonique quant à elle étudie les relations qui existent entre deux groupes de variables aléatoires  $X^{(1)}: p_1 \times 1$  et  $X^{(2)}: p_2 \times 1$ . Elle consiste à construire, dans chaque groupe, de nouvelles variables indépendantes (des combinaisons linéaires) telles que la première combinaison du premier groupe ait avec la première du deuxième groupe la plus grande corrélation possible. Ces nouvelles variables sont appelées les premières variables canoniques, et leur corrélation est la première corrélation canonique. Les deuxièmes variables canoniques sont les combinaisons linéaires indépendante des premières donnant la plus grande corrélation parmi celles qui restent, et ainsi de suite. Le chapitre 1 traite en détail l'analyse canonique à deux groupes de variables aléatoires.

Plusieurs chercheurs ont travaillé sur la généralisation de l'analyse canonique à plus de deux groupes de variables aléatoires. Ils ont souvent eu recours pour construire les variables et corrélations canoniques à la maximisation d'une fonction de corrélation. Le chapitre 2 sera consacré à trois types de généralisation. Deux sont basées sur la maximisation d'une fonction de corrélation (Carroll (1968), Lazraq et al. (1992)). La troisième généralisation est basée sur la minimisation de la variance généralisée. La généralisation de Carroll (1968) consiste à trouver des combinaisons linéaires des variables de chaque groupe  $A_i X_i$  où  $A_i$  est  $1 \times p_1$ ,  $X_i$   $p_i \times n$  la matrice des données observées du vecteur  $X^{(i)}$ , et à trouver aussi une variable compromis  $Z$  de telle façon que  $\sum w_i r^2(Z, A_i X_i)$  soit maximale, où  $w_i$  est un poids accordé à  $X^{(i)}$  et  $r$  est le coefficient de corrélation linéaire simple. La généralisation de Lazraq et al. (1992) consiste à chercher  $l$  combinaisons linéaires des variables de chaque groupe ainsi que  $l$  variables compromis

les plus liées au sens de la mesure de liaison de Lingos et Schonemann(*RLS*) et au sens de la mesure de Escoufier( *RV*), voir Lazraq & Cléroux (1988A). La généralisation de Gupta est basée sur la minimisation de la variance généralisée. Elle consiste à trouver des combinaisons linéaires de chaque groupe telles que leurs variances généralisées soient minimales.

Au niveau de l'échantillon, l'analyse canonique à deux groupes de variables aléatoires est basée sur l'estimation des paramètres de position  $\mu$  et de dispersion  $\Sigma$ . Toutefois l'utilisation de l'estimateur de maximum de vraisemblance n'est toujours pas un bon choix. En effet en présence d'observations aberrantes, ce choix devient complètement non fiable. C'est pourquoi il faut, dans ce cas, utiliser des estimateurs robustes de  $\mu$  et de  $\Sigma$ . Dans le chapitre 3, nous commencerons par donner les fonctions d'influence des variables et corrélations canoniques. Puis on exposera quelques estimateurs robustes des paramètres de  $\mu$  et de  $\Sigma$  et leur utilisation dans l'analyse canonique. Dans la dernière partie de ce mémoire on retrouve un programmes SAS qui calcule la solution de l'analyse canonique à deux groupes de variables aléatoires, des programmes codés en Splus pour calculer la solution de Carroll (1968), de Lazraq et al. (1992) et celle de Gupta (1981).

# Chapitre 1

---

## ANALYSE CANONIQUE À DEUX GROUPES DE VARIABLES ALÉATOIRES

Par opposition à l'analyse en composantes principales, qui étudie les relations existantes entre les variables d'un même vecteur aléatoire, l'analyse canonique à deux groupes de variables aléatoires, appelée souvent analyse canonique tout court, quant à elle, étudie les relations qui lient deux groupes de variables aléatoires. Au lieu de considérer toutes les variables de chaque groupe, Hotelling (1935, 1936), a proposé de se limiter à l'étude des relations qui lient quelques combinaisons linéaires, non corrélées, des variables du premier groupe à d'autres combinaisons linéaires, aussi non corrélées, des variables du deuxième groupe. Ces combinaisons linéaires sont de nouvelles variables qu'on va construire à l'aide de l'analyse canonique. En fait, l'analyse canonique permet de réduire la dimension du problème qu'on étudie. Si on a un nuage de points ( $n$  observations d'un premier vecteur aléatoire de dimension  $p_1$ ) dans un espace  $\mathbf{R}^{p_1}$  et un autre nuage de points ( $n$  observations d'un deuxième vecteur aléatoire de dimension  $p_2$ ) dans un espace  $\mathbf{R}^{p_2}$  (les vecteurs aléatoires sont souvent observés chez les mêmes individus), l'analyse canonique permet de projeter ces deux nuages sur un seul sous espace  $\mathbf{R}^m$  de dimension  $m \leq \min(p_1, p_2)$  de telle sorte que les deux projections soient rapprochées l'une de l'autre le plus possible. Comme elle permet aussi, si c'est possible, en étudiant les structures de corrélation des deux vecteurs de prédire l'un par l'autre. Les nouvelles variables vont expliquer, aussi bien que les  $p_1 + p_2$  variables initiales, la corrélation entre les deux vecteurs et cela sans perdre

trop d'information.

Dans le cadre de ce chapitre, on verra comment l'analyse canonique à deux vecteurs nous permet de construire les nouvelles variables que ça soit au niveau de la population ou au niveau de l'échantillon, dans ce sens, on définira deux concepts, à savoir "les variables canoniques" et "les corrélations canoniques". On travaillera d'abord au niveau de la population (section 1.2), puis ensuite au niveau de l'échantillon (section 1.3). On verra aussi comment en pratique, on peut utiliser les tests d'hypothèses pour se décider sur le nombre de paires de variables canoniques qui seront suffisantes pour analyser les structures de corrélations (section 1.4). La section 1.5 sera consacrée aux distributions des corrélations canoniques. On conclura ce chapitre par un exemple (section 1.6) qui étudie les corrélations entre le vecteur des tailles et celui des poids de 63 CONSCRITS, de leurs pères et de leurs mères.

### 1.1. FORMULATION DU PROBLÈME

On peut trouver la formulation suivante dans Anderson (1984).

Soit  $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$  un vecteur aléatoire subdivisé en deux sous vecteurs aléatoires  $X^{(1)} : p_1 \times 1$  et  $X^{(2)} : p_2 \times 2$  ( on suppose  $p_1 \leq p_2$  ), avec  $\mathbf{E}(X^{(1)}) = \mu^{(1)}$ ,  $\mathbf{E}(X^{(2)}) = \mu^{(2)}$ , et de matrice de variance-covariance  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$  supposée définie positive, où

$$\Sigma_{11} = \mathbf{E} (X^{(1)} - \mu^{(1)}) (X^{(1)} - \mu^{(1)})'$$

$$\Sigma_{22} = \mathbf{E} (X^{(2)} - \mu^{(2)}) (X^{(2)} - \mu^{(2)})'$$

$$\Sigma'_{21} = \Sigma_{12} = \mathbf{E} (X^{(1)} - \mu^{(1)}) (X^{(2)} - \mu^{(2)})' .$$

On suppose que  $\Sigma_{11}$  et  $\Sigma_{22}$  sont inversibles, et sans perte de généralité on suppose que  $\mu^{(1)} = \mu^{(2)} = 0$ .

En général, chacun des vecteurs  $X^{(1)}$  et  $X^{(2)}$  contient un grand nombre de variables aléatoires, et on s'intéresse à chercher de nouvelles variables, en plus petit nombre, de telle façon qu'elles expliqueront la corrélation entre les deux vecteurs  $X^{(1)}$  et  $X^{(2)}$  aussi bien que les  $p_1 + p_2$  variables initiales.

On cherche donc des combinaisons linéaires  $\mathcal{U}_j = \alpha^{(j)'} X^{(1)}$  et  $\mathcal{V}_j = \gamma^{(j)'} X^{(2)}$ ,  $j \leq p_1$ , où  $\alpha^{(j)} : p_1 \times 1$  et  $\gamma^{(j)} : p_2 \times 1$  de telle sorte que :

- (i) :  $\mathcal{U}_1$  et  $\mathcal{V}_1$  auront la plus grande corrélation simple possible.
- (ii) :  $\mathcal{U}_2$  ne sera pas corrélée à  $\mathcal{U}_1$ ,  $\mathcal{V}_2$  ne sera pas corrélée à  $\mathcal{V}_1$  et  $\mathcal{U}_2$  et  $\mathcal{V}_2$  auront la plus grande corrélation simple possible parmi celles qui restent.
- (iii) :  $\mathcal{U}_j$  ne sera pas corrélée à  $\mathcal{U}_l$  pour tout  $l \leq j - 1$ ,  $\mathcal{V}_j$  ne sera pas corrélée à  $\mathcal{V}_l$ , pour tout  $l \leq j - 1$  et  $\mathcal{U}_j$  et  $\mathcal{V}_j$  auront la plus grande corrélation possible parmi celles qui restent.
- (iv) : ainsi de suite jusqu'à un maximum de  $p_1$  paires de telles combinaisons linéaires.

$\mathcal{U}_j$  et  $\mathcal{V}_j$  seront définies comme les  $j$ èmes variables canoniques, et la corrélation entre  $\mathcal{U}_j$  et  $\mathcal{V}_j$  sera définie, quant à elle, comme la  $j$ ème corrélation canonique.

Si on a de la chance, peu de paires de combinaisons linéaires auront des corrélations linéaires simples importantes, ce qui permettra alors de réduire les dimensions du problème sans perdre trop d'informations (en terme de corrélation).

On commence d'abord à travailler au niveau de la population, puis ensuite au niveau de l'échantillon et on estimera les résultats qu'on aura trouvés au niveau de la population.

## 1.2. VARIABLES ET CORRÉLATIONS CANONIQUES AU NIVEAU DE LA POPULATION.

On commence tout d'abord par chercher les premières et deuxièmes variables et corrélations canoniques, ensuite les  $j$ èmes variables et corrélation canoniques pour  $j \leq p_1$ .



### 1.2.1. Les premières variables et corrélation canoniques.

On veut trouver  $\alpha^{(1)}$  et  $\gamma^{(1)}$  tels que le coefficient de corrélation simple entre  $\mathcal{U}_1 = \alpha^{(1)'} X^{(1)}$  et  $\mathcal{V}_1 = \gamma^{(1)'} X^{(2)}$  qu'on note par  $\rho(\alpha^{(1)}, \gamma^{(1)})$  soit maximal.  $\rho(\alpha^{(1)}, \gamma^{(1)})$  est donné par :

$$\rho(\alpha^{(1)}, \gamma^{(1)}) = \frac{\text{Cov}(\mathcal{U}_1, \mathcal{V}_1)}{[\text{Var}(\mathcal{U}_1) \text{Var}(\mathcal{V}_1)]^{1/2}} = \frac{\alpha^{(1)'} \Sigma_{12} \gamma^{(1)}}{[\alpha^{(1)'} \Sigma_{11} \alpha^{(1)} \gamma^{(1)'} \Sigma_{22} \gamma^{(1)}]^{1/2}} .$$

On utilise la notation  $\rho(\alpha^{(j)}, \gamma^{(j)})$  pour souligner que cette corrélation varie en fonction de  $\alpha^{(j)}$  et de  $\gamma^{(j)}$  .

Le coefficient de corrélation est invariant par un changement d'échelle, on peut donc normaliser  $\alpha^{(1)}$  et  $\gamma^{(1)}$  de telle façon que  $\text{Var}(\mathcal{U}_1) = \text{Var}(\mathcal{V}_1) = 1$  .

$\rho(\alpha^{(1)}, \gamma^{(1)})$  se réduit alors à  $\rho(\alpha^{(1)}, \gamma^{(1)}) = \alpha^{(1)'} \Sigma_{12} \gamma^{(1)}$  .

Donc on a à maximiser  $\alpha^{(1)'} \Sigma_{12} \gamma^{(1)}$  par rapport à  $\alpha^{(1)}$  et à  $\gamma^{(1)}$  sous contraintes  $\alpha^{(1)'} \Sigma_{11} \alpha^{(1)} = \gamma^{(1)'} \Sigma_{22} \gamma^{(1)} = 1$  .

On procède alors par la méthode des multiplicateurs de Lagrange. On a :

$$\mathcal{L}(\alpha^{(1)}, \gamma^{(1)}) = \alpha^{(1)'} \Sigma_{12} \gamma^{(1)} - \frac{\lambda}{2} (\alpha^{(1)'} \Sigma_{11} \alpha^{(1)} - 1) - \frac{\mu}{2} (\gamma^{(1)'} \Sigma_{22} \gamma^{(1)} - 1) , \quad (1.2.1)$$

où  $\lambda$  et  $\mu$  sont les multiplicateurs de Lagrange. On dérive (1.2.1) par rapport à  $\alpha^{(1)}$  et à  $\gamma^{(1)}$  et on doit résoudre le système d'équations

$$\frac{\partial \mathcal{L}(\alpha^{(1)}, \gamma^{(1)})}{\partial \alpha^{(1)}} = \Sigma_{12} \gamma^{(1)} - \lambda \Sigma_{11} \alpha^{(1)} = 0 \quad (1.2.2)$$

$$\frac{\partial \mathcal{L}(\alpha^{(1)}, \gamma^{(1)})}{\partial \gamma^{(1)}} = \Sigma'_{12} \alpha^{(1)} - \mu \Sigma_{22} \gamma^{(1)} = 0 \quad (1.2.3)$$

On multiplie (1.2.2) à gauche par  $\alpha^{(1)'} et (1.2.3) à gauche par  $\gamma^{(1)'$  et on obtient :$

$$\alpha^{(1)'} \Sigma_{12} \gamma^{(1)} - \lambda \alpha^{(1)'} \Sigma_{11} \alpha^{(1)} = 0 \quad (1.2.4)$$

$$\gamma^{(1)'} \Sigma'_{12} \alpha^{(1)} - \mu \gamma^{(1)'} \Sigma_{22} \gamma^{(1)} = 0 \quad (1.2.5)$$

Or  $\alpha^{(1)'} \Sigma_{11} \alpha^{(1)} = \gamma^{(1)'} \Sigma_{22} \gamma^{(1)} = 1$  , on obtient alors :

$\lambda = \alpha^{(1)'} \Sigma_{12} \gamma^{(1)} = \mu = \gamma^{(1)'} \Sigma'_{12} \alpha^{(1)}$  (  $\lambda$  et  $\mu$  sont des scalaires). Ce qui fait que

(1.2.2) et (1.2.3) deviennent, avec  $\Sigma'_{12} = \Sigma_{21}$  :

$$-\lambda \Sigma_{11} \alpha^{(1)} + \Sigma_{12} \gamma^{(1)} = 0, \quad (1.2.6)$$

$$\Sigma_{21} \alpha^{(1)} - \lambda \Sigma_{22} \gamma^{(1)} = 0. \quad (1.2.7)$$

Les équations (1.2.6) et (1.2.7) sont équivalentes à :

$$\begin{pmatrix} -\lambda \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda \Sigma_{22} \end{pmatrix} \begin{pmatrix} \alpha^{(1)} \\ \gamma^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

L'équation (1.2.7) donne :

$$\gamma^{(1)} = \frac{1}{\lambda} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(1)} \quad \text{si } \lambda \neq 0. \quad (1.2.9)$$

On remplace  $\gamma^{(1)}$  dans (1.2.6) , on trouve :

$$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(1)} = \lambda^2 \alpha^{(1)}. \quad (1.2.10)$$

Anderson (1984), page 483, montre que cette equation possède  $p_1$  valeur propres positives.  $\alpha^{(1)}$  est le vecteur propre correspondant à la valeur propre  $\lambda^2$  de la matrice  $\mathcal{M} = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$  . Et puisque on cherche  $\alpha^{(1)'} X^{(1)}$  et  $\gamma^{(1)'} X^{(2)}$  telle que  $\rho(\alpha^{(1)}, \gamma^{(1)})$  soit maximale, et comme on a trouvé que  $\rho(\alpha^{(1)}, \gamma^{(1)}) = \alpha^{(1)'} \Sigma_{12} \gamma^{(1)} = \lambda$ , donc il suffit de prendre  $\lambda = \lambda_1$  égale à la racine carrée positive de la plus grande valeur propre (non nulle)  $\lambda_1^2$  de la matrice  $\mathcal{M}$  .

Une fois qu'on a  $\lambda_1$ , on prend le vecteur propre  $\alpha^{(1)}$  de la matrice  $\mathcal{M}$  correspondant à la valeur propre  $\lambda_1^2$  telle que  $\alpha^{(1)'} \Sigma_{11} \alpha^{(1)} = 1$ , ensuite on aura  $\gamma^{(1)}$  à partir de (1.2.9) qui vérifie nécessairement  $\gamma^{(1)'} \Sigma_{22} \gamma^{(1)} = 1$  .

En effet, ayant normalisé  $\alpha^{(1)}$  pour avoir  $\alpha^{(1)'} \Sigma_{11} \alpha^{(1)} = 1$  on a :

$$\begin{aligned}
 \gamma^{(1)'} \Sigma_{22} \gamma^{(1)} &= \frac{1}{\lambda_1^2} \alpha^{(1)'} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{22} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(1)} \\
 &= \frac{1}{\lambda_1^2} \alpha^{(1)'} \Sigma_{11} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(1)} \\
 &= \frac{1}{\lambda_1^2} \alpha^{(1)'} \Sigma_{11} \mathcal{M} \alpha^{(1)} \\
 &= \frac{1}{\lambda_1^2} \alpha^{(1)'} \Sigma_{11} \lambda_1^2 \alpha^{(1)} \\
 &= 1 \quad .
 \end{aligned}$$

**Définition 1.2.1.** Soient  $X^{(1)} : p_1 \times 1$  ,  $X^{(2)} : p_2 \times 1$  ( $p_1 \leq p_2$ ) deux vecteurs aléatoires, et soient les combinaisons linéaires  $\mathcal{U}_1 = \alpha^{(1)'} X^{(1)}$ ,  $\mathcal{V}_1 = \gamma^{(1)'} X^{(2)}$  à variances égales à 1 et à corrélation maximale .

$\mathcal{U}_1$  et  $\mathcal{V}_1$  sont dites les premières variables canoniques, et leur corrélation est la première corrélation canonique.

**Théorème 1.2.1.** Soient  $X^{(1)} : p_1 \times 1$  ,  $X^{(2)} : p_2 \times 1$  ( $p_1 \leq p_2$ ) deux vecteurs aléatoires, et  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$  leur matrice de variance-covariance qui est supposée définie positive et où  $\Sigma_{11} = \mathbf{Var}(X^{(1)})$  et  $\Sigma_{22} = \mathbf{Var}(X^{(2)})$  sont supposées inversibles et  $\Sigma'_{21} = \Sigma_{12} = \mathbf{Cov}(X^{(1)}, X^{(2)})$ .

La première corrélation canonique de  $X^{(1)}$  et  $X^{(2)}$  est la racine carré positive de la plus grande valeur propre non nulle  $\lambda_1^2$  de la matrice  $\mathcal{M} = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ .  $\mathcal{U}_1 = \alpha^{(1)'} X^{(1)}$  et  $\mathcal{V}_1 = \gamma^{(1)'} X^{(2)}$  où  $\alpha^{(1)}$  est le vecteur propre normalisé associé à la plus grande valeur propre  $\lambda_1^2$  de la matrice  $\mathcal{M}$  et  $\gamma^{(1)}$  est telle que  $\gamma^{(1)} = \frac{1}{\lambda_1} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(1)}$  sont les premières variables canoniques de  $X^{(1)}$  et de  $X^{(2)}$  .

### 1.2.2. Les deuxièmes variables et corrélation canoniques.

Cherchons à présent la deuxième corrélation canonique ainsi que les deuxièmes variables canoniques .

Soient  $\mathcal{U}_2 = \alpha^{(2)'} X^{(1)}$ ,  $\alpha^{(2)}: p_1 \times 1$  et  $\mathcal{V}_2 = \gamma^{(2)'} X^{(2)}$ ,  $\gamma^{(2)}: p_2 \times 1$  des combinaisons linéaires telle que :

$$\mathbf{Var}(\mathcal{U}_2) = \alpha^{(2)'} \Sigma_{11} \alpha^{(2)} = 1 \quad ,$$

$$\mathbf{Var}(\mathcal{V}_2) = \gamma^{(2)'} \Sigma_{22} \gamma^{(2)} = 1 \quad .$$

$$\text{Donc } \rho(\alpha^{(2)}, \gamma^{(2)}) = \frac{\mathbf{Cov}(\alpha^{(2)'} X^{(1)}, \gamma^{(2)'} X^{(2)})}{[\mathbf{Var}(\mathcal{U}_2) \mathbf{Var}(\mathcal{V}_2)]^{\frac{1}{2}}} = \alpha^{(2)'} \Sigma_{12} \gamma^{(2)}.$$

On a avait dit que les deuxièmes combinaisons linéaires qu'on cherche doivent être non corrélées avec les premières variables canoniques, ceci se traduit par :

$$\mathbf{Cov}(\mathcal{U}_2, \mathcal{U}_1) = \mathbf{Cov}(\alpha^{(2)'} X^{(1)}, \alpha^{(1)'} X^{(1)}) = \alpha^{(2)'} \Sigma_{11} \alpha^{(1)} = 0 \quad , \quad (1.2.11)$$

$$\mathbf{Cov}(\mathcal{V}_2, \mathcal{V}_1) = \mathbf{Cov}(\gamma^{(2)'} X^{(2)}, \gamma^{(1)'} X^{(2)}) = \gamma^{(2)'} \Sigma_{22} \gamma^{(1)} = 0 \quad . \quad (1.2.12)$$

On montrera plus tard dans un cas général que (1.2.11) implique aussi que  $\mathcal{U}_2$  et  $\mathcal{V}_1$  sont non corrélées, c'est à dire  $\alpha^{(2)'} \Sigma_{12} \gamma^{(1)} = \gamma^{(1)'} \Sigma_{21} \alpha^{(2)} = 0$  et que (1.2.12) implique à son tour que  $\mathcal{V}_2$  et  $\mathcal{U}_1$  sont non corrélées c'est à dire  $\gamma^{(2)'} \Sigma_{21} \alpha^{(1)} = \alpha^{(1)'} \Sigma_{12} \gamma^{(2)} = 0$ .

Donc on a à maximiser  $\alpha^{(2)'} \Sigma_{12} \gamma^{(2)}$  par rapport à  $\alpha^{(2)}$  et  $\gamma^{(2)}$  sous les contraintes :

$$\alpha^{(2)'} \Sigma_{11} \alpha^{(2)} = 1 \quad (1.2.13)$$

$$\gamma^{(2)'} \Sigma_{22} \gamma^{(2)} = 1 \quad (1.2.14)$$

$$\alpha^{(2)'} \Sigma_{11} \alpha^{(1)} = \alpha^{(1)'} \Sigma_{11} \alpha^{(2)} = 0 \quad (1.2.15)$$

$$\gamma^{(2)'} \Sigma_{22} \gamma^{(1)} = \gamma^{(1)'} \Sigma_{22} \gamma^{(2)} = 0 \quad . \quad (1.2.16)$$

On va résoudre ce système par la méthode des multiplicateurs de Lagrange .

Le lagrangien de ce système s'écrit comme :

$$\begin{aligned} \mathcal{L}(\alpha^{(2)}, \gamma^{(2)}) = & \alpha^{(2)'} \Sigma_{12} \gamma^{(2)} - \frac{\lambda}{2} (\alpha^{(2)'} \Sigma_{11} \alpha^{(2)} - 1) - \\ & \frac{\mu}{2} (\gamma^{(2)'} \Sigma_{22} \gamma^{(2)} - 1) + \nu \alpha^{(2)'} \Sigma_{11} \alpha^{(1)} + \theta \gamma^{(2)'} \Sigma_{22} \gamma^{(1)} \quad . \quad (1.2.17) \end{aligned}$$

On dérive (1.2.17) par rapport à  $\alpha^{(2)}$  et à  $\gamma^{(2)}$  et on égalisera les équations trouvées à zéro pour obtenir  $\alpha^{(2)}$  et  $\gamma^{(2)}$ . Le système d'équations à résoudre est donc :

$$\frac{\partial \mathcal{L}(\alpha^{(2)}, \gamma^{(2)})}{\partial \alpha^{(2)}} = \Sigma_{12} \gamma^{(2)} - \lambda \Sigma_{11} \alpha^{(2)} + \nu \Sigma_{11} \alpha^{(1)} = 0 \quad , \quad (1.2.18)$$

$$\frac{\partial \mathcal{L}(\alpha^{(2)}, \gamma^{(2)})}{\partial \gamma^{(2)}} = \Sigma_{21} \alpha^{(2)} - \mu \Sigma_{22} \gamma^{(2)} + \theta \Sigma_{22} \gamma^{(1)} = 0 \quad . \quad (1.2.19)$$

On multiplie (1.2.18) à gauche par  $\alpha^{(1) \prime}$  et on trouve :

$$\nu \alpha^{(1) \prime} \Sigma_{11} \alpha^{(1)} = \nu = 0 \quad \text{puisque} \quad \alpha^{(1) \prime} \Sigma_{12} \gamma^{(2)} = 0 \quad \text{et} \quad \alpha^{(1) \prime} \Sigma_{11} \alpha^{(2)} = 0 \quad .$$

On multiplie aussi (1.2.19) à gauche par  $\gamma^{(1) \prime}$  et on trouve :

$$\theta \gamma^{(1) \prime} \Sigma_{22} \gamma^{(1)} = \theta = 0 \quad \text{puisque} \quad \gamma^{(1) \prime} \Sigma_{21} \alpha^{(2)} = 0 \quad \text{et} \quad \gamma^{(1) \prime} \Sigma_{22} \gamma^{(2)} = 0 \quad .$$

Les équations (1.2.18) et (1.2.19) deviendront alors :

$$\Sigma_{12} \gamma^{(2)} - \lambda \Sigma_{11} \alpha^{(2)} = 0 \quad , \quad (1.2.20)$$

$$\Sigma_{21} \alpha^{(2)} - \mu \Sigma_{22} \gamma^{(2)} = 0 \quad . \quad (1.2.21)$$

Les équations (1.2.20) et (1.2.21) sont équivalentes aux équations (1.2.2) et (1.2.3), donc  $\alpha^{(2)}$  sera le vecteur propre normalisé correspondant à une valeur propre non nulle de la matrice  $\mathcal{M}$  vérifiant, en plus de la contrainte (1.2.13), la contrainte (1.2.15). Et  $\gamma^{(2)}$  sera  $\gamma^{(2)} = \frac{1}{\lambda_2} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(2)}$  et qui, en plus qu'elle vérifie nécessairement  $\gamma^{(2) \prime} \Sigma_{22} \gamma^{(2)} = 1$ , doit vérifier la contrainte (1.2.16).

On cherche à maximiser la corrélation canonique qui comme auparavant sera la racine carrée positive de la plus grande valeur propre non nulle de la matrice  $\mathcal{M}$ , telle que le vecteur propre lui correspondant  $\alpha^{(2)}$  vérifie la contrainte d'orthogonalité à  $\alpha^{(1)}$ . On sait que les vecteurs propres correspondants à deux valeurs propres distinctes sont orthogonaux. De ce fait,  $\alpha^{(2)}$  sera le vecteur propre normalisé correspondant à la racine carrée positive de la deuxième plus grande valeur

propre de la matrice  $\mathcal{M}$ . Et par conséquent  $\mathcal{U}_2 = \alpha^{(2)'} X^{(1)}$  sera à son tour orthogonale à  $\mathcal{U}_1 = \alpha^{(1)'} X^{(1)}$  et à  $\mathcal{V}_1 = \gamma^{(1)'} X^{(2)}$ ,  $\mathcal{V}_2 = \gamma^{(2)'} X^{(2)}$  sera aussi orthogonale à  $\mathcal{V}_1 = \gamma^{(1)'} X^{(2)}$  et à  $\mathcal{U}_1 = \alpha^{(1)'} X^{(1)}$ .

**Définition 1.2.2.** Soient  $X^{(1)} : p_1 \times 1$ ,  $X^{(2)} : p_2 \times 1$  avec  $p_1 \leq p_2$ , deux vecteurs aléatoires, et soient les combinaisons linéaires  $\mathcal{U}_2 = \alpha^{(2)'} X^{(1)}$  et  $\mathcal{V}_2 = \gamma^{(2)'} X^{(2)}$  telles que  $\mathcal{U}_2$  et  $\mathcal{V}_2$  sont à variances égales à 1 et ne sont pas corrélées aux premières variables canoniques de  $X^{(1)}$  et  $X^{(2)}$  et sont à corrélation maximale.  $\mathcal{U}_2$  et  $\mathcal{V}_2$  sont dites les deuxièmes variables canoniques et leur corrélation est la deuxième corrélation canonique.

**Théorème 1.2.2.** En gardant les mêmes notations qu'au théorème 1.2.1, la deuxième corrélation canonique de  $X^{(1)}$  et  $X^{(2)}$  est la racine carré positive de la deuxième plus grande valeur propre non nulle  $\lambda_2^2$  de la matrice  $\mathcal{M} = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ .  $\mathcal{U}_2 = \alpha^{(2)'} X^{(1)}$  et  $\mathcal{V}_2 = \gamma^{(2)'} X^{(2)}$  où  $\alpha^{(2)}$  est le vecteur propre normalisé de  $\mathcal{M}$  correspondant à la valeur propre  $\lambda_2^2$  et  $\gamma^{(2)}$  est telle que  $\gamma^{(2)} = \frac{1}{\lambda_2} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(2)}$  sont les deuxièmes variables canoniques de  $X^{(1)}$  et  $X^{(2)}$ .

### 1.2.3. Les $j$ èmes variables et corrélation canoniques.

À présent, et après avoir trouvé les premières et les deuxièmes variables et corrélations canoniques, on suppose qu'on a trouvé les  $(j - 1)$  premières variables et corrélations canoniques  $[(\mathcal{U}_1, \mathcal{V}_1), (\mathcal{U}_2, \mathcal{V}_2), \dots, (\mathcal{U}_{j-1}, \mathcal{V}_{j-1}), \lambda_1 > \lambda_2 > \dots > \lambda_{j-1} > 0]$  on va chercher les ( $j$ èmes) variables et corrélation canoniques. Et rappelons nous que pour  $i \leq j - 1$ ,  $\mathcal{U}_i$  et  $\mathcal{V}_i$  ne sont pas corrélés à  $\mathcal{U}_l$  et à  $\mathcal{V}_l \forall l \leq i - 1$ , de ce fait  $\text{Cov}(\mathcal{U}_i, \mathcal{U}_l) = \alpha^{(i)'} \Sigma_{11} \alpha^{(l)} = 0$  et  $\text{Cov}(\mathcal{V}_i, \mathcal{V}_l) = \gamma^{(i)'} \Sigma_{22} \gamma^{(l)} = 0$  pour tout  $l \leq i - 1$ . On cherche deux combinaisons linéaires  $\mathcal{U}_j = \alpha^{(j)'} X^{(1)}$  et  $\mathcal{V}_j = \gamma^{(j)'} X^{(2)}$  à variances égales à 1 et telles que  $\mathcal{U}_j$  et  $\mathcal{V}_j$  ne sont pas corrélés à  $\mathcal{U}_l$  et  $\mathcal{V}_l$  pour tout  $l \leq j - 1$  et que la corrélation  $\rho(\alpha^{(j)}, \gamma^{(j)})$  entre  $\mathcal{U}_j$  et  $\mathcal{V}_j$  est maximale. Ceci se traduit comme suit :

maximiser  $\rho(\alpha^{(j)}, \gamma^{(j)}) = \text{Cov}(\mathcal{U}_j, \mathcal{V}_j) = \alpha^{(j)'} \Sigma_{12} \gamma^{(j)}$  par rapport à  $\alpha^{(j)}$  et  $\gamma^{(j)}$  sous

les contraintes suivantes :

$$\text{Var}(\mathcal{U}_j) = \alpha^{(j)'} \Sigma_{11} \alpha^{(j)} = 1 \quad (1.2.22)$$

$$\text{Var}(\mathcal{V}_j) = \gamma^{(j)'} \Sigma_{22} \gamma^{(j)} = 1 \quad (1.2.23)$$

$$\begin{aligned} \text{Cov}(\mathcal{U}_j, \mathcal{U}_l) &= \text{Cov}(\alpha^{(j)'} X^{(1)}, \alpha^{(l)'} X^{(1)}) = \alpha^{(j)'} \Sigma_{11} \alpha^{(l)} \\ &= \alpha^{(l)'} \Sigma_{11} \alpha^{(j)} = 0 \quad \text{pour } l \leq j-1 \end{aligned} \quad (1.2.24)$$

$$\text{Cov}(\mathcal{V}_j, \mathcal{V}_l) = \text{Cov}(\gamma^{(j)'} X^{(2)}, \gamma^{(l)'} X^{(2)}) = \gamma^{(j)'} \Sigma_{22} \gamma^{(l)} \quad (1.2.25)$$

$$= \gamma^{(l)'} \Sigma_{22} \gamma^{(j)} = 0 \quad \text{pour } l \leq j-1. \quad (1.2.26)$$

Là aussi pour résoudre ce système on procédera par la méthode des multiplicateurs de Lagrange. Mais avant tout, on va démontrer un résultat qu'on a déjà utilisé et qu'on utilisera encore à savoir :

$$\text{pour } \lambda_m \neq 0 \text{ et } m < n \quad \text{Cov}(\mathcal{U}_n, \mathcal{U}_m) = 0 \implies \text{Cov}(\mathcal{U}_n, \mathcal{V}_m) = 0 \quad (1.2.27)$$

$$\text{Cov}(\mathcal{V}_n, \mathcal{V}_m) = 0 \implies \text{Cov}(\mathcal{U}_m, \mathcal{V}_n) = 0 \quad (1.2.28)$$

En effet :

$$\text{Cov}(\mathcal{U}_n, \mathcal{V}_m) = \text{Cov}(\alpha^{(n)'} X^{(1)}, \gamma^{(m)'} X^{(2)}) = \alpha^{(n)'} \Sigma_{12} \gamma^{(m)} \quad .$$

$$\text{Comme } \gamma^{(m)} = \frac{1}{\lambda_m} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(m)}, \quad \text{ansi} \quad (1.2.29)$$

$$\begin{aligned} \text{Cov}(\mathcal{U}_n, \mathcal{V}_m) &= \frac{1}{\lambda_m} \alpha^{(n)'} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(m)} \\ &= \frac{1}{\lambda_m} \alpha^{(n)'} \Sigma_{11} \mathcal{M} \alpha^{(m)} \quad (\text{où } \mathcal{M} = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}) \\ &= \frac{1}{\lambda_m} \alpha^{(n)'} \Sigma_{11} \lambda_m^2 \alpha^{(m)} \quad (\alpha^{(m)} \text{ est vecteur propre correspondant à } \lambda_m^2) \\ &= \lambda_m \alpha^{(n)'} \Sigma_{11} \alpha^{(m)} \\ &= \lambda_m \text{Cov}(\mathcal{U}_n, \mathcal{U}_m) \quad \text{ce qui prouve (1.2.27)}. \end{aligned}$$

De la même façon, nous avons

$$\begin{aligned}
\text{Cov}(\mathcal{V}_n, \mathcal{V}_m) &= \text{Cov}(\gamma^{(n)'} X^{(2)}, \gamma^{(m)'} X^{(2)}) \\
&= \gamma^{(n)'} \Sigma_{22} \gamma^{(m)} \\
\gamma^{(m)} &= \frac{1}{\lambda_m} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(m)} \\
\text{Cov}(\mathcal{V}_n, \mathcal{V}_m) &= \frac{1}{\lambda_m} \gamma^{(n)'} \Sigma_{22} \Sigma_{22}^{-1} \Sigma_{21} \alpha^{(m)} \\
&= \frac{1}{\lambda_m} \gamma^{(n)'} \Sigma_{21} \alpha^{(m)} \\
&= \frac{1}{\lambda_m} \text{Cov}(\gamma^{(n)'} X^{(2)} \alpha^{(m)'} X^{(1)}) \\
&= \frac{1}{\lambda_m} \text{Cov}(\mathcal{V}_n, \mathcal{U}_m) \quad \text{ce qui prouve (1.2.28)}.
\end{aligned}$$

Revenons à présent à la résolution de notre système dont le lagrangien et ses dérivées s'écrivent comme suit :

$$\begin{aligned}
\mathcal{L}(\alpha^{(j)}, \gamma^{(j)}) &= \alpha^{(j)'} \Sigma_{12} \gamma^{(j)} - \frac{\lambda}{2} (\alpha^{(j)'} \Sigma_{11} \alpha^{(j)} - 1) - \frac{\mu}{2} (\gamma^{(j)'} \Sigma_{22} \gamma^{(j)} - 1) \\
&\quad + \sum_{k=1}^{j-1} \nu_k \alpha^{(j)'} \Sigma_{11} \alpha^{(k)} + \sum_{k=1}^{j-1} \theta_k \gamma^{(j)'} \Sigma_{22} \gamma^{(k)} \quad (1.2.30)
\end{aligned}$$

$$\frac{\partial \mathcal{L}(\alpha^{(j)}, \gamma^{(j)})}{\partial \alpha^{(j)}} = \Sigma_{12} \gamma^{(j)} - \lambda \Sigma_{11} \alpha^{(j)} + \sum_{k=1}^{j-1} \nu_k \Sigma_{11} \alpha^{(k)} = 0 \quad (1.2.31)$$

$$\frac{\partial \mathcal{L}(\alpha^{(j)}, \gamma^{(j)})}{\partial \gamma^{(j)}} = \Sigma_{21} \alpha^{(j)} - \mu \Sigma_{22} \gamma^{(j)} + \sum_{k=1}^{j-1} \theta_k \Sigma_{22} \gamma^{(k)} = 0 \quad (1.2.32)$$

On multiplie (1.2.31) à gauche par  $\alpha^{(l)'$  et (1.2.32) par  $\gamma^{(l)'}$  pour  $l \leq j-1$ . Ainsi les équations (1.2.31) et (1.2.32) deviennent en utilisant (1.2.25), (1.2.26) (1.2.27) et (1.2.28) :

$$\nu_l \alpha^{(l)'} \Sigma_{11} \alpha^{(l)} = \nu_l = 0 \quad (1.2.33)$$

$$\theta_l \gamma^{(l)'} \Sigma_{22} \gamma^{(l)} = \theta_l = 0 \quad (1.2.34)$$



Finalement les équations (1.2.31) et (1.2.32) se résument à :

$$\Sigma_{12}\gamma^{(j)} - \lambda\Sigma_{11}\alpha^{(j)} = 0 \quad (1.2.35)$$

$$\Sigma_{21}\alpha^{(j)} - \mu\Sigma_{22}\gamma^{(j)} = 0 \quad . \quad (1.2.36)$$

Les équations (1.2.35) et (1.2.36) sont exactement les équations (1.2.2) et (1.2.3) en prenant  $j = 1$  ou les équations (1.2.20) et (1.2.21) en prenant  $j = 2$ . On multiplie à gauche (1.2.35) et (1.2.36) respectivement par  $\alpha^{(j)'$  et  $\gamma^{(j)'}$  on trouve :

$$\begin{aligned} \alpha^{(j)'}\Sigma_{12}\gamma^{(j)} - \lambda\alpha^{(j)'}\Sigma_{11}\alpha^{(j)} &= \alpha^{(j)'}\Sigma_{12}\gamma^{(j)} - \lambda = 0 \\ \gamma^{(j)'}\Sigma_{21}\alpha^{(j)} - \mu\gamma^{(j)'}\Sigma_{22}\gamma^{(j)} &= \gamma^{(j)'}\Sigma_{21}\alpha^{(j)} - \mu = 0 \quad . \end{aligned}$$

Donc  $\mu = \lambda = \alpha^{(j)'}\Sigma_{12}\gamma^{(j)} = \rho(\alpha^{(j)}, \gamma^{(j)})$ . On remplace  $\mu$  par  $\lambda$  dans (1.2.36) ce qui donne  $\Sigma_{22}\gamma^{(j)} = \frac{1}{\lambda}\Sigma_{21}\alpha^{(j)}$  pour  $\lambda \neq 0$ . L'équation (1.2.35) est équivalente à  $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\gamma^{(j)} = \lambda\Sigma_{11}\alpha^{(j)}$ . On remplace le terme  $\Sigma_{22}\gamma^{(j)}$  par  $\frac{1}{\lambda}\Sigma_{21}\alpha^{(j)}$

$$\begin{aligned} \Sigma_{12}\Sigma_{22}^{-1}\frac{1}{\lambda}\Sigma_{21}\alpha^{(j)} &= \lambda\Sigma_{11}\alpha^{(j)} \\ \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\alpha^{(j)} &= \lambda^2\alpha^{(j)} \quad \text{soit} \quad \mathcal{M}\alpha^{(j)} = \lambda^2\alpha^{(j)} \quad . \end{aligned}$$

Donc  $\alpha^{(j)}$  est le vecteur propre correspondant à la valeur propre  $\lambda^2$  de la matrice  $\mathcal{M}$ . Or on veut maximiser  $\rho(\alpha^{(j)}, \gamma^{(j)})$  de telle façon que  $\mathcal{U}_j = \alpha^{(j)'}X^{(1)}$  soit orthogonale à  $\mathcal{U}_1, \dots, \mathcal{U}_{j-1}$  donc  $\alpha^{(j)}$  serait le vecteur propre normalisé correspondant à la racine carré positive de la  $j$ ème plus grande valeur propre  $\lambda^2$  de la matrice  $\mathcal{M}$ , et  $\gamma^{(j)}$  se déduit de  $\gamma^{(j)} = \frac{1}{\lambda}\Sigma_{22}^{-1}\Sigma_{21}\alpha^{(j)}$  qui vérifiera nécessairement  $\gamma^{(j)'}\Sigma_{22}\gamma^{(j)} = 1$ , et  $\alpha^{(j)}$  et  $\gamma^{(j)}$  sont telles que  $\mathcal{U}_j$  et  $\mathcal{V}_j$  sont non corrélées à  $\mathcal{U}_1, \mathcal{V}_1, \dots, \mathcal{U}_{j-1}, \mathcal{V}_{j-1}$ .

**Définition 1.2.3.** Soient  $X^{(1)} : p_1 \times 1$ ,  $X^{(2)} : p_2 \times 1$  avec ( $p_1 \leq p_2$ ) deux vecteurs aléatoires, et soient les combinaisons linéaires  $\mathcal{U}_r = \alpha^{(r)'}X^{(1)}$  et  $\mathcal{V}_r = \gamma^{(r)'}X^{(2)}$  telles que  $\mathcal{U}_r$  et  $\mathcal{V}_r$  sont à variances égales à un et ne sont pas corrélées avec les  $(r-1)$  premières variables canoniques et la corrélation entre  $\mathcal{U}_r$  et  $\mathcal{V}_r$  est maximale.  $\mathcal{U}_r$  et  $\mathcal{V}_r$  sont dites les  $r$ èmes variables canoniques, et leur corrélation est la  $r$ ème corrélation canonique.

**Théorème 1.2.3.** *En gardant les mêmes notations qu'au théorème (1.2.1), la rième corrélation canonique de  $X^{(1)}$  et  $X^{(2)}$  est la racine carré positive de la rième plus grande valeur propre non nulle  $\lambda_r^2$  de la matrice  $\mathcal{M} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ .  $\mathcal{U}_r = \alpha^{(r)'}X^{(1)}$  et  $\mathcal{V}_r = \gamma^{(r)'}X^{(2)}$  où  $\alpha^{(r)}$  est le vecteur propre normalisé de  $\mathcal{M}$  correspondant à la valeur propre  $\lambda_r^2$ , et  $\gamma^{(r)}$  est telle que  $\gamma^{(r)} = \frac{1}{\lambda_r}\Sigma_{22}^{-1}\Sigma_{21}\alpha^{(r)}$  sont les rièmes variables canoniques de  $X^{(1)}$  et  $X^{(2)}$ .*

#### 1.2.4. Ensemble de toutes les variables et corrélations canoniques.

On peut résumer tout ce qu'on a vu sur les variables et corrélations canoniques sous forme d'écriture matricielle.

Soit  $\mathcal{A} = (\alpha^{(1)}, \dots, \alpha^{(p_1)})$  une matrice  $p_1 \times p_1$  dont les colonnes sont les  $p_1$  vecteurs des coefficients des combinaisons linéaires des composantes de  $X^{(1)}$  telles que  $\mathcal{A}'\Sigma_{11}\mathcal{A} = \mathcal{I}$  où  $\mathcal{I}$  est la matrice identité ( $p_1 \times p_1$ ). Et soit  $\Gamma = (\gamma^{(1)}, \dots, \gamma^{(p_1)})$  une matrice  $p_2 \times p_1$  dont les colonnes sont les  $p_1$  vecteurs des coefficients des combinaisons linéaires des composantes de  $X^{(2)}$  telles que  $\Gamma'\Sigma_{22}\Gamma = \mathcal{I}$ . Et soit  $\Lambda = (\lambda_1, \dots, \lambda_{p_1})$  une matrice diagonale dont les  $\lambda_i$  sont les racines carrées positives des valeurs propres de la matrice  $\mathcal{M} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  et sont telles que  $(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p_1} \geq 0)$ .

Soient  $\mathcal{U} = \mathcal{A}'X^{(1)} : p_1 \times 1$  et  $\mathcal{V} = \Gamma'X^{(2)} : p_1 \times 1$ , si les  $\alpha^{(i)}$   $i = 1, \dots, p_1$  sont les vecteurs propres normalisés correspondants aux racines carrées positives des valeurs propres  $\lambda_i^2$   $i = 1, \dots, p_1$  de la matrice  $\mathcal{M}$ , et les  $\gamma^{(i)}$  sont telles que  $\gamma^{(i)} = \frac{1}{\lambda_i}\Sigma_{22}^{-1}\Sigma_{21}\alpha^{(i)}$  (pour  $\lambda_i \neq 0$ ) alors  $\mathcal{U}$  et  $\mathcal{V}$  regroupent les  $p_1$  variables canoniques respectivement de  $X^{(1)}$  et  $X^{(2)}$ , et les éléments de la diagonale de la matrice  $\mathcal{A}'\Sigma_{12}\Gamma = \Lambda$  sont les  $p_1$  corrélations canoniques.

**Théorème 1.2.4.** *Soient  $\mathcal{U}_i, \mathcal{V}_i$   $i = 1, \dots, p_1$  les  $p_1$  variables canoniques de  $X^{(1)} : p_1 \times 1$  et  $X^{(2)} : p_2 \times 1$  où  $p_1 \leq p_2$ , et soient  $\mathcal{U} = (\mathcal{U}_1, \dots, \mathcal{U}_{p_1})'$ ,  $\mathcal{V} = (\mathcal{V}_1, \dots, \mathcal{V}_{p_1})'$*

et  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{p_1})$  où  $\lambda_1 \leq \dots \leq \lambda_{p_1} \leq 0$  sont les  $p_1$  corrélations canoniques, alors:

$$\mathbf{E} \begin{pmatrix} \mathcal{U} \\ \mathcal{V} \end{pmatrix} (\mathcal{U}' \quad \mathcal{V}') = \begin{pmatrix} \mathcal{I} & \Lambda \\ \Lambda & \mathcal{I} \end{pmatrix} .$$

DÉMONSTRATION. Les  $\mathcal{U}_i, \mathcal{V}_i$  sont les  $p_1$  variables canoniques  $\implies \exists \mathcal{A} : p_1 \times p_1$  et  $\exists \Gamma : p_2 \times p_1$  tel que  $\mathcal{U} = \mathcal{A}'X^{(1)}, \mathcal{V} = \Gamma'X^{(2)}$

$$\mathbf{E}(\mathcal{U}\mathcal{U}') = \mathcal{A}'\Sigma_{11}\mathcal{A} = \mathcal{I}$$

$$\mathbf{E}(\mathcal{V}\mathcal{V}') = \Gamma'\Sigma_{22}\Gamma = \mathcal{I}$$

$$\text{et } \mathbf{E}(\mathcal{U}'\mathcal{V}) = \mathcal{A}'\Sigma_{12}\Gamma = \Lambda$$

$$\text{alors } \mathbf{E} \begin{pmatrix} \mathcal{U} \\ \mathcal{V} \end{pmatrix} (\mathcal{U}' \quad \mathcal{V}') = \mathbf{E} \begin{pmatrix} \mathcal{U}\mathcal{U}' & \mathcal{U}\mathcal{V}' \\ \mathcal{V}\mathcal{U}' & \mathcal{V}\mathcal{V}' \end{pmatrix} = \begin{pmatrix} \mathcal{I} & \Lambda \\ \Lambda & \mathcal{I} \end{pmatrix} .$$

□

### 1.3. L'ANALYSE CANONIQUE AU NIVEAU DE L'ÉCHANTILLON.

Quand on a travaillé au niveau de la population, le paramètre  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$  était présent dans toutes les étapes de la recherche des variables et corrélations canoniques. Malheureusement ce paramètre est souvent inconnu dans beaucoup de situations réelles. Ceci n'empêche pas, pour trouver les variables et corrélations canoniques au niveau de l'échantillon, de procéder de la même manière qu'au niveau de la population tout en substituant la matrice  $\mathcal{S} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$  à la matrice  $\Sigma$  et où  $S_{ij} = \frac{1}{n-1} \sum_{\alpha=1}^n (X_{\alpha}^{(i)} - \bar{X}^{(i)})(X_{\alpha}^{(j)} - \bar{X}^{(j)})'$ ,  $n$  est la taille de l'échantillon.

Supposons que le vecteur aléatoire  $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$ ,  $X^{(1)} : p_1 \times 1$ ,  $X^{(2)} : p_2 \times 1$  est observé chez un échantillon aléatoire de taille  $n$ .

Soient  $x_1, \dots, x_\alpha, \dots, x_n$  où  $x_\alpha = \begin{pmatrix} x_\alpha^{(1)} \\ x_\alpha^{(2)} \end{pmatrix}$  ces observations. On a donc à chercher les combinaisons linéaires  $u_j = a^{(j)'} X^{(1)}$ ,  $a^{(j)} : p_1 \times 1$  et  $v_j = b^{(j)'} X^{(2)}$ ,  $b^{(j)} : p_2 \times 1$  telles que le coefficient de corrélation échantillonnaire (empirique)  $r(a^{(j)}, b^{(j)}) = \frac{S_{u_j v_j}}{\sqrt{S_{u_j u_j} S_{v_j v_j}}} = a^{(j)'} S_{12} b^{(j)}$  soit maximal sous les contraintes  $S_{u_j u_j} = a^{(j)'} S_{11} a^{(j)} = 1$  et  $S_{v_j v_j} = b^{(j)'} S_{22} b^{(j)} = 1$  et que  $u_j$  et  $v_j$  sont non corrélés avec les  $(j-1)$  combinaisons  $(u_1, v_1), \dots, (u_{j-1}, v_{j-1})$ .

On procédera de la même manière qu'au niveau de la population. On trouvera que la corrélation canonique  $l_j$  et les variables canoniques  $u_j$  et  $v_j$  sont telles que  $l_j^2$ ,  $a^{(j)}$  et  $b^{(j)}$  sont solutions du systèmes suivant :

$$S_{12} S_{22}^{-1} S_{21} a^{(j)} = l_j^2 S_{11} a^{(j)} \quad (1.3.1)$$

$$b^{(j)} = \frac{1}{l_j} S_{22}^{-1} S_{21} a^{(j)}. \quad (1.3.2)$$

Donc  $l_j$  est la racine carré positive non nulle de la  $j$ ème plus grande valeur propre  $l_j^2$  de la matrice  $S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}$  et  $a^{(j)}$  est le vecteur propre lui correspondant,  $b^{(j)}$  est telle que  $b^{(j)} = \frac{1}{l_j} S_{22}^{-1} S_{21} a^{(j)}$ .  $u_j$  et  $v_j$  sont les  $j$ èmes variables canoniques échantillonnales (au niveau de l'échantillon) de  $X^{(1)}$  et  $X^{(2)}$ . Remarquons qu'en travaillant au niveau de la population ou au niveau de l'échantillon on n'avait pas besoin de la distribution de  $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$ . Cependant Anderson (1984) a montré que si  $X$  suit une loi multinormale alors  $\hat{\alpha}^{(j)}$ ,  $\hat{\gamma}^{(j)}$  et  $\hat{\lambda}_j$ , données par les équations (1.3.1) et (1.3.2), sont les estimateurs de maximum de vraisemblance (EMV) de,  $\alpha^{(j)}$ ,  $\gamma^{(j)}$  et  $\lambda_j$ .

En effet, on sait que l'EMV de  $\Sigma$  est :

$$\widehat{\Sigma} = \frac{1}{n} \sum_{\alpha=1}^n (x_\alpha - \bar{x})(x_\alpha - \bar{x})',$$

qu'ont peut partitionner de la façon suivante :

$$\begin{aligned} \widehat{\Sigma} &= \begin{pmatrix} \frac{1}{n} \sum_{\alpha=1}^n (x_{\alpha}^{(1)} - \bar{x}^{(1)}) (x_{\alpha}^{(1)} - \bar{x}^{(1)})' & \frac{1}{n} \sum_{\alpha=1}^n (x_{\alpha}^{(1)} - \bar{x}^{(1)}) (x_{\alpha}^{(2)} - \bar{x}^{(2)})' \\ \frac{1}{n} \sum_{\alpha=1}^n (x_{\alpha}^{(2)} - \bar{x}^{(2)}) (x_{\alpha}^{(1)} - \bar{x}^{(1)})' & \frac{1}{n} \sum_{\alpha=1}^n (x_{\alpha}^{(2)} - \bar{x}^{(2)}) (x_{\alpha}^{(2)} - \bar{x}^{(2)})' \end{pmatrix} \\ &= \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} . \end{aligned}$$

En remplaçant dans (1.2.35) et (1.2.36)  $\Sigma_{11}$  ,  $\Sigma_{12}$  ,  $\Sigma_{21}$  et  $\Sigma_{22}$  respectivement par  $\hat{\Sigma}_{11}$  ,  $\hat{\Sigma}_{12}$ ,  $\hat{\Sigma}_{21}$  et  $\hat{\Sigma}_{22}$  ( autrement on remplace  $\Sigma$  par son EMV ) on aura , selon le corollaire 3.2.1 de Anderson (1984), les EMV de  $\alpha^{(j)}$ ,  $\gamma^{(j)}$  et  $\lambda^{(j)}$ .

#### 1.4. PROPRIÉTÉS ÉCHANTILLONNALES DES CORRÉLATIONS CANONIQUES.

Avant d'étudier les propriétés échantillonnales des corrélations canoniques, on tient à définir la loi de Wishart et celle de la statistique  $\Lambda$  de Wilks puisqu'on en aura besoin.

##### 1.4.1. La distribution de Wishart.

**Définition 1.4.1.** Soient  $Z_1, \dots, Z_n$  un échantillon aléatoire issu d'une population  $\mathcal{N}\left(0, \Sigma\right)$  :  $p_1 \times 1$  et  $\Sigma$  est définie positive . Soit  $\mathcal{A} = \sum_{\alpha=1}^n Z_{\alpha} Z_{\alpha}'$ . La distribution de  $\mathcal{A}$  est appelée la distribution de Wishart et est notée par :  $\mathcal{A} \sim \mathcal{W}(\Sigma, n)$ .

**Théorème 1.4.1.** Si  $\mathcal{A} \sim \mathcal{W}(\Sigma, n)$ , alors sa densité est donnée par :

$$f(\mathcal{A}) = \frac{|\mathcal{A}|^{\frac{1}{2}(n-p-1)} e^{\left(\frac{-1}{2} \text{tr} \Sigma^{-1} \mathcal{A}\right)}}{2^{\frac{1}{2}np} \pi^{\frac{p(p-1)}{4}} |\Sigma|^{\frac{1}{2}n} \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]} ,$$

$n \geq p$  pour  $\mathcal{A}$  définie positive.

Pour la preuve de ce théorème on se réfère à Anderson (1984)

### 1.4.2. La distribution de la statistique $\Lambda$ de Wilks.

**Définition 1.4.2.** (tirée de Mardia et al. (1979) page 81 ).

Quand  $\mathcal{A} \sim \mathcal{W}(\mathcal{I}, m)$  et  $\mathcal{B} \sim \mathcal{W}(\mathcal{I}, n)$ , où  $\mathcal{I}$  est la matrice identité  $p \times p$ , sont indépendantes, on dit que :

$$\Lambda = \frac{|\mathcal{A}|}{|\mathcal{A} + \mathcal{B}|} = |\mathcal{I} + \mathcal{A}^{-1}\mathcal{B}|^{-1}$$

a la distribution de la statistique  $\Lambda$  de Wilks de paramètres  $p$ ,  $m$  et  $n$ . Souvent on utilise la notation suivante  $\Lambda \sim \Lambda(p, m, n)$ .

**Théorème 1.4.2.**  $\Lambda(p, m, n) \sim \prod_{i=1}^n U_i$  où  $U_1, \dots, U_n$  sont  $n$  variables indépendantes et telles que :  $U_i \sim \beta\left(\frac{1}{2}(m+i-p), \frac{1}{2}p\right)$ ,  $i = 1, \dots, n$  où  $\beta(\cdot, \cdot)$  est la loi bêta.

Rappelons que l'objectif de l'analyse canonique est de trouver de nouvelles variables (variables canoniques  $\mathcal{U}_j, \mathcal{V}_j$ ) telle que leurs corrélations,  $\lambda_j = \alpha^{(j)'} \Sigma_{12} \gamma^{(j)}$ ,  $j = 1 \dots, p_1$ , soient maximales. Les nouvelles variables expliquent les corrélations qui existent entre  $X^{(1)}$  et  $X^{(2)}$ . Si ces dernières ne sont pas corrélées, c'est-à-dire  $\text{Cov}(X^{(1)}, X^{(2)}) = \Sigma_{12} = 0$ , alors c'est inutile d'aller chercher  $\mathcal{U}_j$  et  $\mathcal{V}_j$  puisque toutes les  $\lambda_j$  sont nulles.

Pour tester l'hypothèse  $\mathcal{H}_0 : \Sigma_{12} = 0$  ( ou bien les  $\lambda_j = 0$ ,  $j = 1, \dots, p_1$  ), sous la normalité de  $X^{(1)}$  et  $X^{(2)}$ , on utilise la statistique suivante (Mardia et al. (1979), page 282):

$$|\mathcal{I} - S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}| = \prod_{i=1}^{p_1} (1 - l_i^2), \quad (1.4.1)$$

où les  $l_i$  sont les corrélations canoniques au niveau de l'échantillon. Cette statistique suit la distribution de Wilks  $\Lambda(p_2, n-1-p_2, p_2)$ .

Pour  $n$  assez grand on utilise l'approximation de Bartlett (Bartlett (1939)) :

$$-\left[ n - \frac{1}{2}(p_1 + p_2 + 3) \right] \log \prod_{i=1}^{p_1} (1 - l_i^2) \sim \chi_{p_1 p_2}^2 \quad (1.4.2)$$

De cette façon, on teste si toutes les corrélations canoniques  $\lambda_1, \dots, \lambda_{p_1}$  sont nulles.

Si  $\Sigma_{12} \neq 0$ , alors la question qui se pose est : combien y a-t-il de corrélations canoniques  $\lambda_j$  qui sont non nulles? ou bien de combien de combinaisons de variables canoniques avons nous besoin pour expliquer les corrélations entre  $X^{(1)}$  et  $X^{(2)}$ ? Pour tester l'hypothèse  $\mathcal{H} : \lambda_{s+1} = \dots = \lambda_{p_1} = 0$ , c'est à dire les  $p_1 - s$  dernières corrélations sont nulles, alors on utilise le test suivant proposé par Bartlett (1939).

$$- \left[ n - \frac{1}{2} (p_1 + p_2 + 3) \right] \log \prod_{i=s+1}^{p_1} (1 - l^2) \sim \chi_{(p_1-s)(p_2-s)}^2. \quad (1.4.3)$$

Afin de déterminer le nombre de corrélations canoniques non nulles on peut appliquer la règle empirique suivante :

- (1): On teste l'hypothèse  $\mathcal{H}_0$  : les  $p_1$  corrélations canoniques sont nulles ( $\Sigma_{12} = 0$  ou  $\lambda_1 = 0$ ).
- (2): Si l'hypothèse en (1) est rejetée, alors on teste :  $\mathcal{H}_0$  : les  $(p_1 - 1)$  petites corrélations canoniques sont nulles, c'est à dire  $\lambda_2 = \dots = \lambda_{p_1} = 0$ .
- (3): et ainsi de suite jusqu'au moment où les  $\lambda_i$  qui restent commencent à être négligeables (statistiquement nulles).

La détermination du niveau globale du test pose un problème.

### 1.4.3. Distribution des corrélations canoniques.

Dans cette section on donnera la densité des corrélations canoniques dans le cas où  $\Sigma_{12} = 0$  et dans le cas où il existe des corrélations canoniques non nulles.

1.4.3.1. *Cas où  $\Sigma_{12} = 0$ .*

Supposons que  $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$  où  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ .

Soient  $l_1, \dots, l_{p_1}$  les corrélations canoniques au niveau de l'échantillon. La distribution de  $l_j$  ( $j = 1, \dots, p_1$ ) est donnée par :

$$f(l) = \pi^{\frac{1}{2}} p_1^2 \frac{\Gamma_{p_1} \left[ \frac{1}{2}(n-1) \right]}{\Gamma_{p_1} \left[ \frac{1}{2}(n-1-p_2) \right] \Gamma_p \left( \frac{1}{2} p_1 \right) \Gamma_p \left( \frac{1}{2} p_2 \right)} \\ \times \prod_{i=1}^{p_1} \left[ l_i^{\frac{1}{2}(p_2-p_1-1)} (1-l_i)^{\frac{1}{2}(n-p_2-p_1-3)} \right] \prod_{i < j}^{p_1} (l_i - l_j) \quad \text{où } p = p_1 + p_2. \quad (1.4.4)$$

Dans le cas où  $p_1 = 1$   $p_2 = p - 1$ , (1.4.4) se résume à :

$$\frac{\Gamma \left[ \frac{1}{2}(n-1) \right]}{\Gamma \left[ \frac{1}{2}(n-1-p_2) \right] \Gamma_p \left[ \frac{1}{2} p_2 \right]} l^{\frac{1}{2}(p_2-2)} (1-l)^{\frac{1}{2}(n-p_2-4)}. \quad (1.4.5)$$

Cette densité n'est autre que celle de la racine carrée du coefficient de corrélation multiple empirique entre  $X^{(1)}$  et  $X^{(2)}$ . Pour plus de détails on peut se référer à (Anderson (1984), section 13.4)

1.4.3.2. *Cas où  $\Sigma_{12} \neq 0$ .*

On suppose que les  $k$  premières corrélations canoniques  $\lambda_1, \dots, \lambda_k$  sont distinctes et non nulles. Et soient :

$$z_i = \sqrt{n} \frac{l_i^2 - \lambda_i^2}{2\lambda_i(1 - \lambda_i^2)} \quad i = 1, \dots, k \\ z_i = n l_i^2 \quad i = k+1, \dots, p_1.$$

Les  $z_i$  pour  $i = 1, \dots, k$ , ont une distribution limite normale centrée et réduite, alors que la distribution de  $z_i$  pour  $i = k+1, \dots, p_1$  est donnée par :

$$\frac{\pi^{\frac{1}{2}(p_1-k)^2} e^{-\frac{1}{2} \sum_{i=k+1}^{p_1} z_i}}{\prod_{i=k+1}^{p_1} z_i^{\frac{1}{2}(p_2-p_1-1)} \prod_{\substack{i,j=k+1 \\ i < j}}^{p_1} (z_i - z_j)}$$



Pour plus de détails on peut se référer à Anderson (1984), section 12.4.

### 1.5. EXEMPLE : DONNÉES CONSCRITS

On étudiera dans le cadre de ce paragraphe les données d'un exemple dit CONSCRITS . Les données de cet exemple sont tirées de Foucart (1984) pages 84 et 85 et sont données en annexe. Elles sont constituées de la taille, du poids et de la peinture de 63 conscrits, de leurs pères et de leurs mères; les tailles sont notées respectivement  $TAI$ ,  $TAP$  et  $TAM$ ; les poids sont notés respectivement  $PDS$ ,  $PDP$  et  $PDM$ ; les peintures sont respectivement notées  $PTU$ ,  $PTP$  et  $PTM$ . On veut étudier les corrélations existantes entre les tailles et les poids. Le premier vecteur de variables aléatoires est celui des tailles:  $X^{(1)} = (TAI, TAP, TAM)'$ ; le deuxième est celui des poids  $X^{(2)} = (PDS, PDP, PDM)'$

#### 1.5.1. Analyse des résultats de la sortie SAS

L'examen des corrélations simples deux à deux que ce soit des variables de  $X^{(1)}$ (tableau 1.5.1) ou celles de  $X^{(2)}$ (tableau 1.5.2) ou des variables de  $X^{(1)}$  avec celles de  $X^{(2)}$  (tableau 1.5.3) ne révèlent pas de corrélation importante.

TAB. 1.5.1. Corrélations entre les variables de  $X^{(1)}$

Variable	$TAI$	$TAP$	$TAM$
$TAI$	1,000	0,532	0,447
$TAP$	0,532	1,000	0,311
$TAM$	0,447	0,311	1,000

TAB. 1.5.2. Corrélations entre les variables de  $X^{(2)}$

Variable	$PDS$	$PDP$	$PDM$
$PDS$	1,000	0,390	0,385
$PDP$	0,390	1,000	0,160
$PDM$	0,385	0,160	1,000

TAB. 1.5.3. Corrélations entre les variables de  $X^{(1)}$  avec celles de  $X^{(2)}$ 

Variable	$TAI$	$TAP$	$TAM$
$PDS$	0,693	0,283	0,356
$PDP$	0,428	0,650	0,173
$PDM$	0,368	0,103	0,562

Le tableau 1.5.4 donne les corrélations canoniques entre  $X^{(1)}$  et  $X^{(2)}$ .

TAB. 1.5.4. Analyse des corrélations canoniques

ordre de la corrélation	corrélations canoniques
1	0,721
2	0,652
3	0,402

Soient les hypothèses  $H_{01} : \lambda_1 = \lambda_2 = \lambda_3 = 0$ ;  $H_{02} : \lambda_2 = \lambda_3 = 0$  et  $H_{03} : \lambda_3 = 0$ . Le tableau 1.5.5 donne les valeurs-p du test de chacune de ces hypothèses versus  $H_{11} : \lambda_1 \neq 0$  ou  $\lambda_2 \neq 0$  ou  $\lambda_3 \neq 0$ ,  $H_{12} : \lambda_2 \neq 0$  ou  $\lambda_3 \neq 0$  et  $H_{13} : \lambda_3 \neq 0$  (respectivement) .

TAB. 1.5.5. Valeurs-p pour  $H_{01}$ ,  $H_{02}$  et  $H_{03}$ 

hypothèse	valeur-p
$H_{01}$	< 0,0001
$H_{02}$	< 0,0001
$H_{03}$	0,0013

Au niveau de signification 0,01 on rejette  $H_{01}$ ,  $H_{02}$  et  $H_{03}$  . Les deux premières variables canoniques sont :

$$U_1 = 0,114TAI + 0,035TAP + 0,047TAM \quad ,$$

$$V_1 = 0,10PDS + 0,02PDP + 0,04PDM \quad .$$

Ces deux variables canoniques sont des variables de tailles, une variable de taille est une moyenne pondérée alors qu'une variable de forme est une variable qui oppose un groupe de variables qui la compose avec un autre. Si on retient l'échelle suivant : une corrélation supérieur à 0,80 est jugée forte, une corrélation entre 0,30 et 0,80 exclusivement est jugée moyenne et enfin une corrélation en deça de 0,30 sera considéré comme faible.  $U_1$  est fortement corrélée à  $TAI$  puisque  $\text{Corr}(U_1, TAI) = 0,95$  et moyennement corrélée à  $TAP$  et à  $TAM$  puisque  $\text{Corr}(U_1, TAP) = 0,69$  et  $\text{Corr}(U_1, TAM) = 0,63$ .  $V_1$  est fortement corrélée à  $PDS$  puisque  $\text{Corr}(V_1, PDS) = 0,95$  et moyennement corrélée à  $PDP$  et à  $PDM$  puisque  $\text{Corr}(V_1, PDP) = 0,52$  et  $\text{Corr}(V_1, PDM) = 0,60$ . Les secondes variables canoniques sont :

$$U_2 = -0,068TAI + 0,171TAP - 0,106TAM \quad ,$$

$$V_2 = -0,032PDS + 0,112PDP - 0,067PDM \quad .$$

Elles sont des variables de formes.  $U_2$  est moyennement corrélée à  $TAP$  et faiblement corrélée à  $TAI$  et  $TAM$ .  $V_2$  est faiblement corrélée à  $PDS$ , à  $PDP$  et à  $PDM$ .

TAB. 1.5.6. Corrélation de  $U_1$ ,  $U_2$  et de  $U_3$  avec  $TAI$ ,  $TAP$  et  $TAM$ .

	$U_1$	$U_2$	$U_3$
$TAI$	0,954	-0,091	-0,286
$TAP$	0,688	0,693	0,214
$TAM$	0,633	-0,387	0,670

TAB. 1.5.7. Corrélation de  $V_1$ ,  $V_2$  et de  $V_3$  avec  $PDS$ ,  $PDP$  et  $PDM$ .

	$V_1$	$V_2$	$V_3$
$PDS$	0,955	-0,045	-0,294
$PDP$	0,525	0,813	0,253
$PDM$	0,598	-0,405	0,692

Les troisièmes variables canoniques sont :

$$U_3 = -0,147TAI + 0,065TAP + 0,192TAM \quad ,$$

$$V_3 = -0,106PDS + 0,048PDP + 0,135PDM \quad .$$

$U_3$  et  $V_3$  sont, respectivement, moyennement corrélées à  $TAM$  et à  $PDM$ . Elles sont, respectivement, faiblement corrélées à  $TAI$  et à  $TAP$ ; à  $PDS$  et à  $PDP$ .

# Chapitre 2

---

## L'ANALYSE CANONIQUE GÉNÉRALISÉE

### 2.1. APERÇU DE LA LITTÉRATURE SUR SUJET

La majorité des chercheurs qui ont travaillé sur la généralisation de l'analyse canonique (AC) à plusieurs groupes de variables aléatoires ont souvent eu recours, pour trouver les variables canoniques, à maximiser une fonction des corrélations de celles-ci. La méthode des corrélations canoniques partielles (Roy (1957), Raja (1969)) considère uniquement deux groupes de variables tandis que les autres groupes sont considérés comme fixes. Anderson (1984) dans le problème 3 page 519, affirme que dans le cas de deux groupes de variables, la recherche des variables canoniques en se basant sur le critère de maximisation du coefficient de corrélation est équivalente à celle basée sur le critère de minimisation de la variance généralisée des variables canoniques, plus encore, il donne une méthode pour trouver les variables canonique dans le cas de trois groupes de variables aléatoires. Mckeon (1965) a défini un coefficient de corrélation  $r_1(X_1, \dots, X_k)$  entre  $k$  variables comme étant  $r_1(X_1, \dots, X_k) = \frac{2 \sum_{i < j} \sigma_{ij}}{(k-1) \sum_i \sigma_i^2}$  sur lequel il s'est basé pour généraliser l'AC, où  $\sigma_{ij} = \mathbf{E}(X_i - \mu(X_i))(X_j - \mu(X_j))$  et  $\sigma_i^2 = \mathbf{E}(X_i - \mu(X_i))^2$  Dans ce cas, il définit la corrélation canonique généralisée comme :

$$\rho(a'_1 X^{(1)}, \dots, a'_k X^{(k)}) = \max_{a_i} r_1(a'_1 X^{(1)}, \dots, a'_k X^{(k)})$$

et les variables canoniques sont les combinaisons linéaires  $a'_i X^{(i)}$   $i = 1, \dots, k$  qui maximisent  $r_1$ . On verra dans la section suivante que Carroll (1968) quant à lui a proposé une autre méthode de généralisation de l'AC basée sur le coefficient

de corrélation empirique. Selon lui, les variables canoniques sont celles qui maximisent une moyenne pondérée des carrés des coefficients de corrélation empiriques entre elles et une variable auxiliaire.

## 2.2. GÉNÉRALISATION SELON L'APPROCHE DE CARROLL.

### 2.2.1. Position du problème et sa résolution.

Soient  $G_1, \dots, G_k$   $k$  groupes de variables aléatoires de tailles respectives  $m_1, \dots, m_k$ . On observe tous ces groupes de variables chez un échantillon aléatoire de taille  $n$ . Les données sont regroupées dans des matrices  $X_1, \dots, X_k$  où  $X_i : m_i \times n$  est formé de  $n$  observations prises sur chacune des  $m_i$  variables du groupe  $G_i, i = 1, \dots, k$ . On suppose que chaque matrice  $X_i$  est centrée (la somme des colonnes de chacune d'elles est un vecteur de zéro) et de plein rang  $m_i, i = 1, \dots, k$ . On cherche des combinaisons linéaires  $\mathcal{A}_i X_i, \mathcal{A}_i : 1 \times m_i, i = 1, 2, \dots, k$  et un vecteur  $Z : 1 \times n$  composé de  $n$  observations prises sur la variable compromis (inconnue) telle que la quantité :

$$\mathcal{R}^2 = \sum_{i=1}^k w_i r^2(Z, \mathcal{A}_i X_i) \quad ,$$

soit maximale, où  $w_i$  est un poids associé à  $G_i$  et où  $r$  est le coefficient de corrélation simple empirique entre  $Z$  et  $\mathcal{A}_i X_i$  (on suppose que  $Z$  est centré et  $Z Z' = 1$ ). Pour maximiser  $\mathcal{R}^2$ , il suffit de maximiser  $r^2(Z, \mathcal{A}_i X_i)$ .

$$r^2(Z, \mathcal{A}_i X_i) = \frac{[Z X_i' \mathcal{A}_i']^2}{Z Z' \mathcal{A}_i X_i X_i' \mathcal{A}_i'} \quad . \quad (2.2.1)$$

Dans une première étape, on maximise (2.2.1) par rapport à  $\mathcal{A}_i$ , puis par la suite on maximise par rapport à  $Z$ .

Pour maximiser  $r^2(Z, \mathcal{A}_i X_i)$  par rapport à  $\mathcal{A}_i$ , il suffit de maximiser :

$$r^{*2}(Z, \mathcal{A}_i X_i) = \frac{[Z X_i' \mathcal{A}_i']^2}{\mathcal{A}_i X_i X_i' \mathcal{A}_i'} \quad ,$$

pour y arriver on utilise le théorème et le corollaire suivants (tirés de Mardia et al. (1979) pages 479-481) :

**Théorème 2.2.1.** *Soient  $\mathcal{A}$  et  $\mathcal{B}$  deux matrices symétriques. On suppose que la matrice  $\mathcal{B}$  est définie positive. Alors le maximum (minimum) de  $X'AX$  sous contrainte  $X'BX = 1$  est atteint quand  $X$  est le vecteur propre de  $\mathcal{B}^{-1}\mathcal{A}$  correspondant à la plus grande (petite) valeur propre  $\lambda_1(\lambda_p)$  de  $\mathcal{B}^{-1}\mathcal{A}$ .*

**Corollaire 2.2.1.**

$$\max_X \frac{(a'X)^2}{X'BX} = a' B^{-1} a,$$

et ce maximum est atteint quand

$$X = \frac{B^{-1}a}{a' B^{-1} a} .$$

Retournons à la maximisation de  $r^{*2}$ . Nous avons

$$r^{*2}(Z, \mathcal{A}_i X_i) = \frac{[ZX'_i \mathcal{A}'_i]^2}{\mathcal{A}_i X_i X'_i \mathcal{A}'_i} = \frac{[a' \mathcal{A}'_i]^2}{\mathcal{A}_i \mathcal{B} \mathcal{A}'_i}$$

où  $a = X_i Z'$  et  $\mathcal{B} = X_i X'_i$ .

D'après le corollaire précédent :

$$\max_{\mathcal{A}_i} r^{*2}(Z, \mathcal{A}_i X_i) = ZX'_i (X_i X'_i)^{-1} X_i Z' ,$$

et ce maximum est atteint quand

$$\mathcal{A}'_i = \frac{(X_i X'_i)^{-1} X_i Z'}{ZX'_i (X_i X'_i)^{-1} X_i Z'} , \quad (2.2.7)$$

soit

$$\mathcal{A}_i = \frac{ZX'_i (X_i X'_i)^{-1}}{ZX'_i (X_i X'_i)^{-1} X_i Z'} . \quad (2.2.8)$$

La matrice  $\mathcal{H} = X'_i (X_i X'_i)^{-1} X_i$  est idempotente et centrante (centring matrix). À partir des propriétés de ce type de matrices (voir Mardia et al. (1979) section

A.3.3 page 462) on peut écrire en posant  $T = Z'$  :

$$\begin{aligned} Z\mathcal{H}Z' &= T'\mathcal{H}T = T' \left[ T - \frac{1}{n} \sum_i T_i \right] \\ &= T'T = ZZ' = 1 \quad \left( \sum_i T_i = 0 \quad \text{puisque } Z \text{ est centré} \right), \end{aligned}$$

par conséquent  $\mathcal{A}_i$  se réduit à :

$$\mathcal{A}_i = ZX'_i(X_iX'_i)^{-1} . \quad (2.2.9)$$

D'où

$$\max_{\mathcal{A}_i} r^2(Z, \mathcal{A}_i X_i) = \frac{ZX'_i(X_iX'_i)^{-1}X_iZ'}{ZZ'} ,$$

et

$$\begin{aligned} \max_{\mathcal{A}_i} \mathcal{R}^2 &= \sum_{i=1}^k w_i \frac{ZX'_i(X_iX'_i)^{-1}X_iZ'}{ZZ'} \\ &= \frac{1}{ZZ'} Z \left[ \sum_{i=1}^k w_i X'_i(X_iX'_i)^{-1} X_i \right] Z' \\ &= \frac{ZQZ'}{ZZ'} \quad \text{où} \quad Q = \sum_{i=1}^k w_i X'_i(X_iX'_i)^{-1} X_i. \quad (2.2.10) \end{aligned}$$

Si tous les  $w_i$  sont positifs, alors  $Q$  est au moins semi-définie positive. Il suffit maintenant de trouver le maximum de  $ZQZ'$ . D'après le théorème (2.2.1) ce maximum est atteint quand  $Z$  est le vecteur propre de  $Q$  correspondant à la plus grande valeur propre de  $Q$ .

Comme on peut le voir à partir de (2.2.10), la valeur maximale de  $\mathcal{R}^2$  serait alors la plus grande valeur propre de  $Q$ . Donc une fois que  $Z$  est trouvé, on le remplace dans (2.2.9) pour trouver  $\mathcal{A}_i$ .

Pour trouver les deuxièmes combinaisons linéaires  $\mathcal{A}_i$  orthogonales au premières et une deuxième variable  $Z$  orthogonale à la première, il suffit de prendre



$Z$  comme vecteur propre de  $\mathcal{Q}$  associé à la deuxième plus grande valeur propre de  $\mathcal{Q}$ , et remplacer  $Z$  dans (2.2.9) pour trouver les  $\mathcal{A}_i$ , et ainsi de suite.

### 2.3. EXEMPLE : DONNÉES SUR LES MOUCHES LEPTOCONOPS CARTERI.

Dans l'exemple suivant, on s'intéresse à étudier les structures de corrélations qui existent entre les grandeurs de l'aile:  $X^{(1)}$ , les grandeurs de palpe:  $X^{(2)}$  et les grandeurs de l'antenne:  $X^{(3)}$  de 35 mouches de types leptoconops carteri. Les données utilisées dans cet exemple sont tirées de Johnson & Wichern (1982). Les variables mesurées sont  $X_1^{(1)}$ : la longueur de l'aile,  $X_2^{(1)}$ : la largeur de l'aile,  $X_1^{(2)}$ : la longueur de la troisième palpe,  $X_2^{(2)}$ : la largeur de la troisième palpe,  $X_3^{(2)}$ : la longueur de la quatrième palpe,  $X_1^{(3)}$ : la longueur du 12 ième segment de l'antenne et  $X_2^{(3)}$ : la longueur du 13 ième segment de l'antenne. On forme les trois matrices suivantes:  $X_1 : 2 \times 35$  contient les données sur  $X^{(1)}$ ,  $X_2 : 3 \times 35$  contient les données sur  $X^{(2)}$  et  $X_3 : 2 \times 35$  contenant les données sur  $X^{(3)}$ . Les poids des vecteurs  $X^{(1)}$ ,  $X^{(2)}$  et  $X^{(3)}$  sont considérés égaux à 1. Après exécution de l'algorithme, codé en Splus(voir annexe1), qui calcule la solution de Carroll, les résultats sont comme suit.

Seules les sept premières valeurs propres sont non nulles. Le tableau 2.3.1 les donne par ordre décroissant.

TAB. 2.3.1. Valeurs propres de la matrice  $\mathcal{Q}$ .

$i$	1	2	3	4	5	6	7
$\lambda_i$	2,25	1,48	1,01	0,92	0,66	0,50	0,21

Limitons nous aux deux premières valeurs propres. Les vecteurs propres  $Z_i$  leurs correspondants sont donnés dans tableau 2.3.2.

TAB. 2.3.2 Les vecteurs propres de la matrice  $\mathcal{Q}$  correspondants à  $\lambda_1$  et  $\lambda_2$ .

$Z_1$	$Z_2$
0,373194055	0,79808844
0,281260226	0,21490302
-0,161446540	-0,03682801
0,801006719	-0,11750189
-0,184874277	0,14727664
-0,002845713	0,10013188
-0,378908031	0,21097475
-0,093281796	-0,26137419
-0,177882745	0,15538861
-0,257507869	0,02885862
-0,637114629	-0,03465702
0,169432743	-0,24214203
0,112801772	-0,35935212
-0,051519259	0,01982990
0,204349682	-0,06916512
-0,168187437	0,26580975
0,233884700	0,14905367
0,019387433	-0,06588097
-0,078333494	-0,15349412
-0,051545860	-0,13044032
-0,006288148	-0,07077284
-0,281313597	0,13833199
0,193754689	-0,24321462
0,161962260	0,24997910
0,726977456	-0,01352004

0,029214586	0,07121957
-0,540298401	0,09304140
0,038468581	-0,20658897
-0,147714545	0,08375808
0,313918569	-0,19668588
-0,220542388	-0,09020074
-0,288788133	-0,04745660
0,142869676	-0,06028405
-0,274776249	-0,29631689
0,200685963	-0,03076900

Après avoir centré les  $Z_i$ , on les transforme de telle sorte que  $Z_i Z_i' = 1$ . On calcule les  $A_i$   $i = 1, 2, 3$  d'ordre 1 et 2.

Les  $A_i$  d'ordre 1 sont :

$$\begin{cases} A_1 = -0.032X_1^{(1)} + 0.007X_2^{(1)} \\ A_2 = -0.038X_1^{(2)} - 0.012X_2^{(2)} - 0.013X_3^{(2)} \\ A_3 = 0.087X_1^{(3)} - 0.020X_2^{(3)} \end{cases} .$$

Les  $A_i$  d'ordre 2 sont :

$$\begin{cases} A_1 = 0.013X_1^{(1)} - 0.034X_2^{(1)} \\ A_2 = 0.035X_1^{(2)} - 0.076X_2^{(2)} - 0.006X_3^{(2)} \\ A_3 = -0.132X_1^{(3)} - 0.122X_2^{(3)} \end{cases} .$$

Le maximum de  $\mathcal{R}$  est égale à la première valeur propre de  $\mathcal{Q}$  à savoir 2,25, et il se réalise quand les  $A_i$  sont ceux du premier ordre et  $Z = Z_1$ . La deuxième plus grande valeur de  $\mathcal{R}$  est la deuxième plus grande valeur propre de  $\mathcal{Q}$  à savoir 1.48, et elle se réalise quand les  $A_i$  sont ceux du deuxième ordre et  $Z = Z_2$ .

## 2.4. RAPPEL DES MESURES DE LIAISON : RLS, RI, RV ET $RV_{reg}$ .

On considère deux vecteurs de variables aléatoires  $X^{(1)} : p_1 \times 1$  et  $X^{(2)} : p_2 \times 1$  observés chez un échantillon aléatoire de taille  $n$ . Soient  $X_\alpha^{(1)}, X_\alpha^{(2)}$   $\alpha = 1, \dots, n$  ces observations. Et soient :

$$Y_1 = (X_1^{(1)} - \bar{X}^{(1)}, X_2^{(1)} - \bar{X}^{(1)}, \dots, X_n^{(1)} - \bar{X}^{(1)}) \quad : p_1 \times n$$

$$Y_2 = (X_1^{(2)} - \bar{X}^{(2)}, X_2^{(2)} - \bar{X}^{(2)}, \dots, X_n^{(2)} - \bar{X}^{(2)}) \quad : p_2 \times n$$

On rappelle que pour toute matrice  $\mathcal{A}$ , la fonction  $\|\mathcal{A}\| = \sqrt{\text{tr} \mathcal{A} \mathcal{A}'}$  est une norme et que la distance entre deux matrices  $\mathcal{A}$  et  $\mathcal{B}$  de même dimension est  $\text{dist}(\mathcal{A}, \mathcal{B}) = \|\mathcal{A} - \mathcal{B}\|$ .

### 2.4.1. La mesure de liaison de Lingoies et Schonemann : RLS.

La mesure de Lingoies et Schonemann RLS (Lazraq & Cléroux (1988A)) est une mesure de liaison qui évalue la proximité entre deux tableaux de données, elle est définie par :

$$RLS(Y_1, Y_2) = \frac{\text{tr}(S_{12} S_{21})^{\frac{1}{2}}}{\sqrt{\text{tr} S_{11} \text{tr} S_{22}}}$$

et elle a la propriété suivante :

$$\min_{T(\perp)} \left\| \frac{Y_1}{\sqrt{\text{tr} Y_1' Y_1}} - T \frac{Y_2}{\sqrt{\text{tr} Y_2' Y_2}} \right\| = \sqrt{2} \sqrt{1 - RLS},$$

où  $T$  est une transformation linéaire orthogonale.

En effet, si on suppose que  $p_1 < p_2$ , alors on ajoute  $p_2 - p_1$  lignes de zéros à la matrice  $Y_1$ , dans ce cas  $Y_1$  et  $Y_2$  sont de même dimension et on peut alors écrire :

$$\begin{aligned} \text{dist}^2(Y_1, TY_2) &= \|Y_1 - TY_2\|^2 = \text{tr}(Y_1 - TY_2)(Y_1 - TY_2)' \\ &= \text{tr} Y_1 Y_1' + \text{tr} Y_2 Y_2' - 2 \text{tr}(TY_2 Y_1'). \end{aligned} \quad (2.4.3)$$

Minimiser (2.4.3) par rapport à  $T$  revient à maximiser  $\text{tr}(TY_2Y_1')$  par rapport à  $T$ . Pour cela, on utilisera le lemme suivant énoncé sans preuve (voir Green (1969)).

**Lemme 2.4.1.** *Soit  $\mathcal{A} : r \times r$  une matrice carré et  $\mathcal{A} = U\Lambda V'$  sa décomposition singulière où  $U$  et  $V$  sont orthogonales et  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$  avec  $\lambda_i \geq 0$  pour  $i = 1, 2, \dots, r$ . Alors, pour toute matrice  $T$  orthogonale,  $\text{tr}T\mathcal{A} \leq \text{tr}(\mathcal{A}'\mathcal{A})^{\frac{1}{2}}$  et l'égalité tient quand  $T = \mathcal{V}\mathcal{U}'$ .*

On pose  $\mathcal{A} = Y_2Y_1'$ , on obtient alors,  $\max_{T(\perp)} \text{tr}(TY_2Y_1') = \text{tr}(Y_1Y_2'Y_2Y_1')^{\frac{1}{2}}$  et le minimum de (2.4.3) devient :

$$\min_{T(\perp)} \text{dist}^2(Y_1, TY_2) = \text{tr}Y_1Y_1' + \text{tr}Y_2Y_2' - 2\text{tr}(Y_1Y_2'Y_2Y_1')^{\frac{1}{2}}. \quad (2.4.4)$$

Si on remplace respectivement  $Y_1$  et  $Y_2$  dans (2.4.4) par  $Y_1^*$  et  $Y_2^*$  où

$$Y_1^* = \frac{Y_1}{\sqrt{\text{tr}Y_1'Y_1}}$$

$$Y_2^* = \frac{Y_2}{\sqrt{\text{tr}Y_2'Y_2}},$$

on obtient, pour la même transformation  $T$ ,

$$\begin{aligned} \min_{T(\perp)} \text{dist}^2(Y_1^*, TY_2^*) &= \text{tr}Y_1^*Y_1^{*\prime} + \text{tr}Y_2^*Y_2^{*\prime} - 2\text{tr}(Y_1^*Y_2^{*\prime}Y_2^*Y_1^{*\prime})^{\frac{1}{2}} \\ &= \text{tr} \left[ \frac{Y_1Y_1'}{\text{tr}Y_1'Y_1} \right] + \text{tr} \left[ \frac{Y_2Y_2'}{\text{tr}Y_2'Y_2} \right] - 2\text{tr} \left[ \frac{Y_1Y_2'Y_2Y_1'}{\text{tr}Y_1'Y_1\text{tr}Y_2'Y_2} \right]^{\frac{1}{2}} \\ &= 2 - 2 \frac{\text{tr}(Y_1Y_2'Y_2Y_1')^{\frac{1}{2}}}{[\text{tr}Y_1'Y_1\text{tr}Y_2'Y_2]^{\frac{1}{2}}} \end{aligned} \quad (2.4.5)$$

or :

$$Y_1Y_1' = (n-1)S_{11} \quad Y_2Y_2' = (n-1)S_{22}$$

$$Y_1Y_2' = (n-1)S_{12} \quad Y_2Y_1' = (n-1)S_{21}.$$

On remplace dans (2.4.5), on trouve alors :

$$\begin{aligned} \min_{T(\perp)} dist^2(Y_1^*, TY_2^*) &= 2 \left[ 1 - \frac{\text{tr}(S_{12}S_{21})^{\frac{1}{2}}}{\sqrt{\text{tr}S_{11}\text{tr}S_{22}}} \right] \\ &= 2(1 - RLS) \quad . \end{aligned}$$

Ce qui donne  $\min_{T(\perp)} dist(Y_1^*, TY_2^*) = \sqrt{2}\sqrt{1 - RLS}$ .

$RLS = 1$  si et seulement s'il existe une transformation linéaire orthogonale  $T$ , qui fait coïncider les deux tableaux de données  $Y_1$  et  $Y_2$ .

Si  $\mathcal{A} = Y_2 Y_1'$  est non singulière, alors  $T\mathcal{A} = (\mathcal{A}\mathcal{A}')^{\frac{1}{2}}$  possède une solution unique  $T = (\mathcal{A}'\mathcal{A})^{\frac{1}{2}}\mathcal{A}^{-1} = (\mathcal{A}'\mathcal{A})^{-\frac{1}{2}}\mathcal{A}' = (Y_1 Y_2' Y_2 Y_1')^{-\frac{1}{2}} Y_1 Y_2'$ .

La rotation  $T$  qui transforme  $Y_2$  en  $TY_2$  est appelée rotation procruste de  $Y_2$  relativement à  $Y_1$ .

### 2.4.2. Les propriétés de RLS

Les propriétés de RLS sont les suivantes :

- (1) : Si  $p_1 = p_2 = 1$  alors  $RLS = |r|$  où  $r$  est le coefficient de corrélation simple empirique;
- (2) :  $0 \leq RLS \leq 1$  et  $RLS = 0$  si et seulement si  $S_{12} = 0$ ;
- (3) :  $RLS(Y_1, Y_2) = RLS(Y_2, Y_1)$ ;
- (4) : Si  $\mathcal{H} : p_1 \times p_1$  est telle que  $\mathcal{H}'\mathcal{H} = k\mathcal{I}$  où  $k$  est un scalaire strictement positif et  $\mathcal{I}$  est la matrice identité, alors  $RLS(\mathcal{H}Y_1, Y_2) = RLS(Y_1, Y_2)$ .

### 2.4.3. La mesure de liaison de Stewart et Love : $RI$ .

La mesure de liaison de Stewart et Love (Lazraq & Cléroux (1988A)) est définie comme étant le maximum de  $RLS^2(Y_1, M'Y_2)$  par rapport à  $M'$ , où  $M : p_2 \times p_1$  est une transformation linéaire. Ce maximum est atteint quand  $M' = S_{12}S_{22}^{-1}$ . Dans ce cas  $RLS(Y_1, M'Y_2)$  serait :

$$RLS^2(Y_1, M'Y_2) = \frac{\text{tr}S_{12}S_{22}^{-1}S_{21}}{\text{tr}S_{11}} = RI(Y_1, Y_2) = RI \quad . \quad (2.4.6)$$

En effet, on peut écrire

$$RLS(Y_1, M' Y_2) = \max_T \frac{\text{tr} Y_1' T M' Y_2}{(\text{tr} Y_1 Y_1')^{\frac{1}{2}} (\text{tr} M' Y_2 Y_2' M)^{\frac{1}{2}}}, \quad (2.4.7)$$

où  $T$  est contrainte à l'orthonormalité. Par l'inégalité de Cauchy-Schwartz on a :

$$\begin{aligned} \text{tr} Y_1' T M' Y_2 &= \text{tr} Y_2 Y_1' T M' = \text{tr} Y_2 Y_2' (Y_2 Y_2')^{-1} Y_2 Y_1' T M' = \text{tr} Y_2' (Y_2 Y_1')^{-1} Y_2 Y_1' T M' Y_2 \\ \text{tr} Y_1' T M' Y_2 &\leq \text{tr} \left[ \left[ Y_2' (Y_2 Y_2')^{-1} Y_2 Y_1' \right] \left[ Y_2' (Y_2 Y_2')^{-1} Y_2 Y_1' \right]' \right]^{\frac{1}{2}} \text{tr} \left[ \left[ T M' Y_2 \right] \left[ T M' Y_2 \right]' \right]^{\frac{1}{2}} \\ &= \text{tr} \left[ Y_2' (Y_2 Y_2')^{-1} Y_2 Y_1' Y_1 Y_2' (Y_2 Y_2')^{-1} Y_2 \right]^{\frac{1}{2}} \text{tr} \left[ T M' Y_2 Y_2' M T \right]^{\frac{1}{2}} \\ &= \text{tr} \left[ Y_2 Y_2' (Y_2 Y_2')^{-1} Y_2 Y_1' Y_1 Y_2' (Y_2 Y_2')^{-1} \right]^{\frac{1}{2}} \text{tr} \left[ M' Y_2 Y_2' M T' T \right]^{\frac{1}{2}} \\ &= \text{tr} \left[ Y_2 Y_1' Y_1 Y_2' (Y_2 Y_2')^{-1} \right]^{\frac{1}{2}} \text{tr} \left[ M' Y_2 Y_2' M \right]^{\frac{1}{2}} \\ &= \text{tr} \left[ Y_1 Y_2' (Y_2 Y_2')^{-1} Y_2 Y_1' \right]^{\frac{1}{2}} \text{tr} \left[ M' Y_2 Y_2' M \right]^{\frac{1}{2}} \\ &= \text{tr} \left[ S_{12} S_{22}^{-1} S_{21} \right]^{\frac{1}{2}} \text{tr} \left[ M' S_{22} M \right]^{\frac{1}{2}}. \end{aligned} \quad (2.4.9)$$

Revenons à (2.4.7) et si on tient compte de (2.4.9) on aura :

$$\begin{aligned} RLS(Y_1, M' Y_2) &\leq \frac{(\text{tr} S_{12} S_{22}^{-1} S_{21})^{\frac{1}{2}} (\text{tr} M' S_{22} M)^{\frac{1}{2}}}{(\text{tr} S_{11})^{\frac{1}{2}} (\text{tr} M' S_{22} M)^{\frac{1}{2}}} \\ &= \frac{(\text{tr} S_{12} S_{22}^{-1} S_{21})^{\frac{1}{2}}}{(\text{tr} S_{11})^{\frac{1}{2}}}. \end{aligned} \quad (2.4.10)$$

L'équation (2.4.10) fournit ainsi une borne supérieure pour  $RLS(Y_1, M' Y_2)$  qui ne dépend ni de  $M$  ni de  $T$ . Pour montrer que cette borne supérieure est atteinte quand  $M' = S_{12} S_{22}^{-1}$ , il suffit de substituer  $M' = S_{12} S_{22}^{-1}$  dans (2.4.7) :

$$\begin{aligned} RLS(Y_1, M' Y_2) &= \max_T \frac{\text{tr} Y_1' T S_{12} S_{22}^{-1} Y_2}{[\text{tr} S_{11}]^{\frac{1}{2}} [\text{tr} S_{12} S_{22}^{-1} S_{22} S_{22}^{-1} S_{12}]^{\frac{1}{2}}} \\ &= \max_T \frac{\text{tr} T S_{12} S_{22}^{-1} S_{21}}{[\text{tr} S_{11}]^{\frac{1}{2}} [\text{tr} S_{12} S_{22}^{-1} S_{21}]^{\frac{1}{2}}}. \end{aligned}$$

Le maximum est atteint pour  $T = \mathcal{I}$  puisqu'alors on obtient la borne supérieure fournie par (2.4.10). On obtient finalement :

$$RLS(Y_1, S_{12}S_{22}^{-1}Y_2) = \frac{(\text{tr}S_{12}S_{22}^{-1}S_{21})^{\frac{1}{2}}}{(\text{tr}S_{11})^{\frac{1}{2}}} = \sqrt{RI(Y_1, Y_2)} \quad .$$

RI est appelé indice de redondance de Stewart et Love. Si on se situe dans le contexte de la régression linéaire multivariée, la matrice de covariance résiduelle après la prédiction de  $X^{(1)}$  par  $X^{(2)}$ , ou de  $Y_1$  par  $Y_2$  est donné par  $S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$ . Si on calcule la trace de  $S_{11.2}$ , on trouve :

$$\begin{aligned} \text{tr}S_{11.2} &= \text{tr}S_{11} - \text{tr}S_{12}S_{22}^{-1}S_{21} \\ &= \text{tr}S_{11} - RI\text{tr}S_{11} \\ &= (1 - RI)\text{tr}S_{11} \\ (n - 1)\text{tr}S_{11.2} &= (1 - RI)(n - 1)\text{tr}S_{11} \\ \|Y_1 - M'Y_2\|^2 &= (1 - RI)\|Y_1\|^2 \quad . \end{aligned}$$

La mesure RI a été utilisée par Lazraq & Cléroux (1988B) dans une procédure pas à pas (de type stepwise) de sélection de variables en régression linéaire multivariée ou en ACPVI (analyse en composantes principales par rapport à des variables instrumentales).

#### 2.4.4. La mesure d'Escoufier : RV.

La mesure de liaison d'Escoufier entre deux tableaux de données  $Y_1, Y_2$  est définie par :

$$RV = RV(Y_1, Y_2) = \frac{\text{tr}S_{12}S_{21}}{\sqrt{\text{tr}S_{11}^2 \text{tr}S_{22}^2}}.$$



Cette mesure est basée sur une certaine distance entre  $Y_1$  et  $Y_2$   $d(Y_1, Y_2)$ .

En effet, si on pose :

$$\mathcal{A} = \frac{Y_1' Y_1}{\sqrt{\text{tr}(Y_1' Y_1)^2}}$$

$$\mathcal{B} = \frac{Y_2' Y_2}{\sqrt{\text{tr}(Y_2' Y_2)^2}},$$

une distance induite entre  $Y_1$  et  $Y_2$  est donnée par :

$$\begin{aligned} d(Y_1, Y_2) &= \text{dist}(\mathcal{A}, \mathcal{B}) \\ &= \left\| \frac{Y_1' Y_1}{\sqrt{\text{tr}(Y_1' Y_1)^2}} - \frac{Y_2' Y_2}{\sqrt{\text{tr}(Y_2' Y_2)^2}} \right\|. \end{aligned} \quad (2.4.12)$$

On développe (2.4.12) :

$$\begin{aligned} d^2(Y_1, Y_2) &= \text{tr} \left[ \left( \frac{Y_1' Y_1}{\sqrt{\text{tr}(Y_1' Y_1)^2}} - \frac{Y_2' Y_2}{\sqrt{\text{tr}(Y_2' Y_2)^2}} \right) \left( \frac{Y_1' Y_1}{\sqrt{\text{tr}(Y_1' Y_1)^2}} - \frac{Y_2' Y_2}{\sqrt{\text{tr}(Y_2' Y_2)^2}} \right)' \right] \\ &= \text{tr} \left( \frac{(Y_1' Y_1)^2}{\text{tr}(Y_1' Y_1)^2} + \frac{(Y_2' Y_2)^2}{\text{tr}(Y_2' Y_2)^2} - \frac{Y_1' Y_1 Y_2' Y_2}{\sqrt{\text{tr}(Y_1' Y_1)^2 \text{tr}(Y_2' Y_2)^2}} - \frac{Y_2' Y_2 Y_1' Y_1}{\sqrt{\text{tr}(Y_2' Y_2)^2 \text{tr}(Y_1' Y_1)^2}} \right) \\ &= 2 - \frac{\text{tr} S_{12} S_{21}}{\sqrt{\text{tr} S_{11}^2 \text{tr} S_{22}^2}} - \frac{\text{tr} S_{12} S_{21}}{\sqrt{\text{tr} S_{22}^2 \text{tr} S_{11}^2}} \\ &= 2 \left[ 1 - \frac{\text{tr} S_{12} S_{21}}{\sqrt{\text{tr} S_{11}^2 \text{tr} S_{22}^2}} \right] = 2 [1 - RV] \end{aligned}$$

Ainsi,

$$d(Y_1, Y_2) = \sqrt{2} \sqrt{1 - RV} \quad \text{où} \quad RV = \frac{\text{tr} S_{12} S_{21}}{\sqrt{\text{tr} S_{11}^2 \text{tr} S_{22}^2}}.$$

$RV = 1$  si et seulement si les positions relatives du premier nuage de points dans  $\mathbf{R}^{p_1}$  et ceux du deuxième nuage dans  $\mathbf{R}^{p_2}$  sont identiques.

### 2.4.5. Propriétés du RV

Les propriétés de RV sont les suivantes :

- (1) : Si  $p_1 = p_2 = 1$  alors  $RV^2 = r^2$  où  $r$  est le coefficient de corrélation simple empirique;
- (2) :  $0 \leq RV \leq 1$  et  $RV = 0$  si et seulement si  $S_{12} = 0$ ;
- (3) :  $RV(Y_1, Y_2) = RV(Y_2, Y_1)$ ;
- (4) : Si  $\mathcal{H} : p_1 \times p_1$  est telle que  $\mathcal{H}'\mathcal{H} = k\mathcal{I}$  où  $k$  est un scalaire strictement positive et  $\mathcal{I}$  est la matrice identité, alors  $RV(\mathcal{H}Y_1, Y_2) = RV(Y_1, Y_2)$ .

### 2.4.6. La mesure de Robert et Escoufier : $RV_{reg}$ .

La mesure de liaison de Robert et Escoufier entre deux tableaux de données  $Y_1$  et  $Y_2$  est définie comme :

$$\begin{aligned}
 RV_{reg} &= RV_{reg}(Y_1, Y_2) = \max_{M'} RV(Y_1, M'Y_2) \\
 &= \left[ \frac{\text{tr}(S_{12}S_{22}^{-1}S_{21})^2}{\text{tr}S_{11}^2} \right]^{\frac{1}{2}} \quad \text{où } M' : p_2 \times p_1 \text{ est une transformation linéaire.}
 \end{aligned}
 \tag{2.4.13}$$

Ce maximum est atteint quand  $M' = S_{12}S_{22}^{-1}$  ( voir Robert & Escoufier (1976)). Pour cette valeur de  $M'$ , on a  $(Y_1 - M'Y_2)Y_2' = 0$ .

On peut facilement démontrer que :

$$\|Y_1'Y_1 - Y_2'MM'Y_2\|^2 = \|Y_1'Y_1\|^2(1 - RV_{reg}^2).
 \tag{2.4.14}$$

En effet :

$$\begin{aligned}
 \|Y_1'Y_1 - Y_2'MM'Y_2\|^2 &= \text{tr} \left[ (Y_1'Y_1 - Y_2'MM'Y_2) (Y_1'Y_1 - Y_2'MM'Y_2)' \right] \\
 &= \text{tr}(Y_1'Y_1)^2 + \text{tr}(MM'Y_2Y_2')^2 - \text{tr}(Y_1Y_2'MM'Y_2Y_1') - \text{tr}(MM'Y_2Y_1'Y_1Y_2') \\
 &= \text{tr}(Y_1'Y_1)^2 + \text{tr}(S_{22}^{-1}S_{21}S_{12}S_{22}^{-1}S_{22})^2(n-1)^2 - \text{tr}(S_{12}S_{22}^{-1}S_{21}S_{12}S_{22}^{-1}S_{21})(n-1)^2 \\
 &\quad - \text{tr}(S_{22}^{-1}S_{21}S_{12}S_{22}^{-1}S_{21}S_{12})(n-1)^2
 \end{aligned}$$

$$\begin{aligned}
&= (n-1)^2 \left[ \text{tr} S_{11}^2 + \text{tr} (S_{12} S_{22}^{-1} S_{21})^2 - \text{tr} (S_{12} S_{22}^{-1} S_{21})^2 - \text{tr} (S_{12} S_{22}^{-1} S_{21})^2 \right] \\
&= (n-1)^2 \left[ \text{tr} S_{11}^2 - \text{tr} (S_{12} S_{22}^{-1} S_{21})^2 \right] \\
&= (n-1)^2 \text{tr} S_{11}^2 \left[ 1 - \frac{\text{tr} (S_{12} S_{22}^{-1} S_{21})^2}{\text{tr} S_{11}^2} \right] \\
&= \text{tr} (Y_1 Y_1')^2 [1 - RV_{reg}^2(Y_1, Y_2)] \\
&= \|Y_1 Y_1'\|^2 [1 - RV_{reg}^2(Y_1, Y_2)] \quad .
\end{aligned}$$

#### 2.4.7. Comparaisons entre RLS, RI, RV et $RV_{reg}$ .

On peut facilement voir que  $RLS^2 \leq RI$  et  $RV \leq RV_{reg}$ . On peut aussi démontrer que (voir Lazraq & Cléroux (1988A))

$$\frac{1}{\sqrt{p_1}} RI \leq RV_{reg} \leq \sqrt{p_1} RI \quad .$$

Comme on peut démontrer que (voir Lazraq et al. (1992))

$$\frac{1}{\sqrt{p_1 p_2}} RLS^2 \leq RV \leq \sqrt{p_1 p_2} RLS^2 \quad .$$

### 2.5. ANALYSE CANONIQUE GÉNÉRALISÉE AU CAS VECTORIEL VIA (RLS, RI) : GÉNÉRALISATION DE LAZRAQ, CLÉROUX ET KIERS.

L'analyse canonique généralisée au sens de Lazraq et al. (1992) consiste à chercher  $l$  combinaisons linéaires des variables de chaque groupe ainsi que  $l$  variables compromises les plus liées, au sens de la mesure RLS, avec l'ensemble des groupes de  $l$  combinaisons linéaires. Au niveau de l'échantillon, il s'agit de chercher  $k$  groupes de combinaisons linéaires  $\mathcal{A}_i X_i$  où  $\mathcal{A}_i : l \times m_i$   $i = 1, 2, \dots, k$  et une matrice  $Z : l \times n$  dont les lignes sont normalisées ( $ZZ' = \mathcal{I}$ ) telles que la quantité

$$\mathcal{Q}_l^2 = \sum_{i=1}^k w_i RLS^2(Z, \mathcal{A}_i X_i)$$

soit maximale. Comme à la section 2, pour maximiser  $\mathcal{Q}_i^2$ , dans un premier temps on fixe  $Z$  et on maximise par rapport à  $\mathcal{A}_i$ , puis après on maximise par rapport à  $Z$ .

Maximiser  $\mathcal{Q}_i^2$  revient à maximiser  $RLS^2(Z, \mathcal{A}_i X_i) \forall i = 1, 2, \dots, k$ . Le maximum de  $RLS^2(Z, \mathcal{A}_i X_i)$  est atteint quand  $\mathcal{A}_i = ZX'_i(X_i X'_i)^{-1}$  (voir section 2.2) et la valeur maximale de  $RLS^2(Z, \mathcal{A}_i X_i)$  est

$$RI = \frac{\text{tr} ZX'_i(X_i X'_i)^{-1} X_i Z'}{\text{tr} ZZ'} = \frac{\text{tr} S_{zi} S_{ii}^{-1} S_{iz}}{\text{tr} S_{zz}}$$

où  $(n-1)S_{zi} = ZX'_i$ ,  $(n-1)S_{ii} = X_i X'_i$ ,  $(n-1)S_{zz} = ZZ' = \mathcal{I}$  sont des matrices de covariances empirique. Et  $\mathcal{Q}_i^2$  serait alors

$$\begin{aligned} \mathcal{Q}_i^2 &= \sum_{i=1}^k w_i \frac{\text{tr} ZX'_i(X_i X'_i)^{-1} X_i Z'}{\text{tr} ZZ'} \\ &= \frac{\text{tr} Z \left[ \sum_{i=1}^k w_i X'_i(X_i X'_i)^{-1} X_i \right] Z'}{\text{tr} ZZ'} \\ &= \frac{\text{tr} Z Q Z'}{\text{tr} ZZ'} \\ &= \frac{\text{tr} Z Q Z'}{l} \quad \text{où} \quad Q = \sum_{i=1}^k w_i X'_i(X_i X'_i)^{-1} X_i \end{aligned}$$

Maximiser  $\mathcal{Q}_i^2$  par rapport à  $Z$  revient alors à maximiser  $\text{tr}(Z Q Z')$  par rapport à  $Z$ . La matrice  $Z$  qui maximise  $\text{tr}(Z Q Z')$  est celle dont les lignes sont les  $l$  vecteurs propres de  $Q$ , correspondant aux  $l$  premières valeurs propres de  $Q$ , observées chez l'échantillon aléatoire, et les corrélations canoniques sont les  $l$  premières valeurs propres de  $Q$ . On retrouve alors la solution de Carroll.

**Exemple 2.5.1.** On revient à l'exemple de la section 2.2 sur les mouches *Leptocnops carteri*. On a vu que seules les sept premières valeurs propres de la matrice  $Q$  sont non nulles. Donc si on s'intéresse à trouver  $l = 7$  combinaisons linéaires de chaque groupe de variables aléatoire  $X^{(i)}$  les plus corrélées avec  $l = 7$  variables compromis  $Z_i$   $i = 1, \dots, 7$ ; il suffit alors de prendre comme  $Z_i$  les vecteurs

propres de  $\mathcal{Q}$  correspondants aux sept valeurs propres non nulles, on remplace dans l'expression de  $A_i$  et on retrouve la solution de Carrol jusqu'à l'ordre sept.

## 2.6. ANALYSE CANONIQUE GÉNÉRALISÉE AU CAS VECTORIEL VIA $(RV, RV_{reg}), RV - GCA, :$ GÉNÉRALISATION DE LAZRAQ, CLÉROUX ET KIERS.

Dans cette section, on cherche  $l$  combinaisons linéaires  $\mathcal{A}_i X_i$  et  $l$  variables compromis telle que la quantité :

$$T^2 = \sum_{i=1}^k w_i RV^2(Z, \mathcal{A}_i X_i) \quad (2.6.1)$$

soit maximale sous la contrainte  $ZZ' = \mathcal{I}$ . Il suffit donc de chercher le maximum de  $RV^2(Z, \mathcal{A}_i X_i)$ . Pour  $Z$  fixe, ce maximum est atteint quand  $\mathcal{A}_i = ZX'_i(X_i X'_i)^{-1}$  et est donné par

$$RV_{reg}^2 = \frac{\text{tr}(ZX'_i(X_i X'_i)^{-1} X_i Z')^2}{\text{tr}(ZZ')^2} = \frac{\text{tr}(S_{zi} S_{ii}^{-1} S_{iz})^2}{\text{tr} S_{zz}^2} \quad (2.6.2)$$

où  $S_{zi}, S_{ii}, S_{zz}$  sont données dans la section 2.4 .  $RV_{reg}$  peut être écrit autrement, en effet:

$$RV_{reg}^2 = \frac{\text{tr} ZX'_i(X_i X'_i)^{-1} X_i Z' Z X'_i(X_i X'_i)^{-1} X_i Z'}{\text{tr} ZZ' ZZ'} = \frac{\text{tr}(Z \mathcal{P}_i Z')^2}{\text{tr} \mathcal{I}} = \frac{\text{tr}(\mathcal{P}_i \mathcal{H}_z)^2}{l} \quad ,$$

où  $\mathcal{H}_z = Z' Z : n \times n$  est une matrice de produit scalaire et  $\mathcal{P}_i = X'_i(X_i X'_i)^{-1} X_i : n \times n$  est semi défini positive.  $\mathcal{P}_i$  est dite une matrice de projection.

Donc on a à maximiser

$$\sum_{i=1}^k w_i \text{tr}(Z \mathcal{P}_i Z')^2 \quad \text{sous contrainte } ZZ' = \mathcal{I} \quad .$$

Cette quantité peut être maximisée de façon itérative en prenant, comme mise à jour de  $Z$ , la matrice  $UV'$  formé à partir de la décomposition singulière

$$\sum_{i=1}^k \mathcal{P}_i Z' Z \mathcal{P}_i Z' = UDV'$$

(voir Kiers et al. (1994)). Cette procédure assure une convergence vers un maximum au moins local de  $\sum_{i=1}^k w_i \text{tr}(Z \mathcal{P}_i Z')$ . Lazraq, Cléroux et Kiers ont élaboré un algorithme codé en PCMATLAB pour calculer la solution  $Z$  obtenue par la méthode de Carroll (1968) ainsi que la solution  $Z_1, Z_2, \dots, Z_n$  de cette section.

## 2.7. LA $RV - GCA$ COMME UNE APPLICATION DE LA MÉTHODE IDIOSCAL APPLIQUÉE AUX MATRICES DE PROJECTION

La méthode IDIOSCAL (voir Carroll & Chang (1970) et Carroll & Chang (1972)) consiste à minimiser la fonction

$$\sigma_1(Z, C_1, \dots, C_K) = \sum_{k=1}^K \left\| S_k - Z C_k Z' \right\|^2, \quad (2.7.1)$$

par rapport aux matrices  $Z : n \times r$  et  $C_k : r \times r$ ,  $k = 1, \dots, K$  où  $C_k$  est défini semi positive(p.s.d), et sans perdre de généralité,  $Z$  est supposée orthonormale.  $\sigma_1$  peut être écrit sous forme

$$\sigma_1(Z, C_1, \dots, C_k) = \sum_{k=1}^K \left\| S_k - Z Z' S_k Z Z' \right\|^2 + \sum_{k=1}^K \left\| C_k - Z' S_k Z \right\|^2, \quad (2.7.2)$$

voir(Kiers et al. (1994)).  $\sigma_1$  atteint son minimum en prenant  $C_k = Z' S_k Z$ , pour tout  $k = 1, \dots, K$ . Si  $S_k$  est p.s.d, alors la matrice  $C_k$  l'est aussi. Remplaçant  $C_k$ , par l'expression ci-dessus, dans (2.7.2), le problème à minimiser devient

$$\begin{aligned} \sigma_1(Z, *) &= \sum_{k=1}^K \left\| S_k - Z Z' S_k Z Z' \right\|^2 \\ &= \sum_{k=1}^K \|S_k\|^2 - \sum_{k=1}^K \text{tr}(Z' S_k Z)^2. \end{aligned} \quad (2.7.3)$$

Minimiser (2.7.3) sous contraintes  $Z' Z = I_r$  (où  $I_r$  est la matrice d'identité  $r \times r$ ) revient à maximiser

$$K(Z) = \sum_{k=1}^K \text{tr}(Z' S_k Z)^2, \text{ par rapport à } Z \text{ sous contrainte } Z' Z = I_r. \quad (2.7.4)$$

On a vu que la solution de la  $RV - GCA$  est celle qui maximise

$$\sum_k^K w_k \text{tr}(Z\mathcal{P}_k Z')^2 \text{ sous contrainte } ZZ' = I_r, \quad (2.7.5)$$

où  $\mathcal{P}_k = X'_k(X_k X'_k)^{-1} X_k$ . L'équation (2.7.5) peut être écrite comme  $\sum_{k=1}^K \text{tr} \left[ Z' (w_k^{\frac{1}{2}} \mathcal{P}_k) Z \right]^2$ . Il s'en suit que la  $RV - GCA$  est équivalente à l'application de la méthode IDIOS-CAL aux matrices  $S_k = w_k^{\frac{1}{2}} \mathcal{P}_k$ . On note que le  $Z$  dans (2.7.4) est le transposé de celui dans (2.7.5)

**Exemple 2.7.1.** *Toujours avec le même exemple des mouches, on cherche les coefficients des variables canoniques par la méthode  $RV - GCA$ . Comme on a supposé que les vecteurs  $X^{(1)}$ ,  $X^{(2)}$  et  $X^{(3)}$  sont de même poids, alors la matrice  $Z$  donnée par la méthode  $RV - GCA$  est aussi celle donnée par la méthode IDIOS-CAL tout en prenant  $S_k = \mathcal{P}_k$ . Une fonction *Splus* (programme 3) permettra de calculer la matrice  $Z$ . Le résultat trouvé est donné dans le tableau suivant (la sortie est celle de *Splus*). On est allé jusqu'à l'ordre 5. La colonne  $[,j]$  désigne  $Z_j$ .*

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	-0.216187446	0.64287136	0.1784160040	0.089954218	0.442603176
[2,]	-0.153672103	0.17709684	-0.0737096925	-0.447381907	-0.088768602
[3,]	0.084234886	-0.03647529	0.1348973650	-0.087472660	-0.125531753
[4,]	-0.474619809	-0.10209814	0.0842340711	0.083818496	-0.242474971
[5,]	0.119325405	0.12731849	-0.0002977815	-0.056199469	-0.032619672
[6,]	0.008380318	0.08187910	-0.2248834593	0.024857675	0.030089168
[7,]	0.200210950	0.16169066	0.0200164951	0.174991633	-0.233622968
[8,]	0.056604727	-0.20335936	-0.1611392898	0.146639680	0.344214100
[9,]	0.093195607	0.12398962	-0.0046333544	0.029175787	-0.006707763
[10,]	0.138128554	0.02416364	-0.0005412001	-0.027132448	-0.064531549
[11,]	0.382727907	-0.01508429	0.2420634215	0.193505577	0.084716283
[12,]	-0.089891908	-0.19266308	0.1648997547	0.026865961	-0.078795072
[13,]	-0.040818543	-0.26863838	-0.3080115035	0.020627325	0.209846170

[14,]	0.031699933	0.01589189	-0.0827480168	0.360792767	-0.184594857
[15,]	-0.106277584	-0.05392470	0.1250780381	0.125332294	-0.051253231
[16,]	0.091488310	0.20770247	-0.5023417709	-0.014876551	-0.329911925
[17,]	-0.104412310	0.13453603	0.3240256970	0.043752793	-0.048721590
[18,]	-0.023612419	-0.06579106	-0.0130652794	-0.057817796	0.060651338
[19,]	0.048232654	-0.11923532	-0.1678195080	-0.141521368	0.086557705
[20,]	0.009345790	-0.11442696	0.0512793905	-0.101258455	-0.099808633
[21,]	0.016581905	-0.04553534	0.0739795596	0.140115785	-0.131464055
[22,]	0.155806445	0.11132251	-0.0558277782	0.079890184	-0.196707126
[23,]	-0.105215199	-0.19173019	0.0915280788	0.024233653	0.078801431
[24,]	-0.094394327	0.19784510	-0.2997776511	0.140701290	-0.001132904
[25,]	-0.412319862	-0.01509921	-0.0575774902	0.145348060	0.066958840
[26,]	-0.013975546	0.05829959	0.1288163618	-0.152859551	0.041602361
[27,]	0.311846199	0.07584848	-0.0861263891	-0.324639551	0.222059187
[28,]	-0.037402433	-0.17667780	-0.0686738062	-0.155194272	0.314895235
[29,]	0.082446865	0.06320743	0.1231531825	-0.071571685	0.109971372
[30,]	-0.179762807	-0.15904518	-0.0549139821	-0.257652957	-0.027366164
[31,]	0.099992821	-0.08698947	0.2977987360	-0.118780085	-0.049731718
[32,]	0.144653659	-0.04995044	0.0462186441	0.007872296	-0.024855264
[33,]	-0.083016186	-0.05394061	-0.0877789796	0.275464260	0.137222846
[34,]	0.150767680	-0.24148769	0.0752524416	0.175033433	0.039011000
[35,]	-0.090092133	-0.01151070	0.0882096910	-0.294614413	-0.250600399

*Les matrices  $A_1$ ,  $A_2$  et  $A_3$  données par le programme 4 en annexe1 sont données dans les tableaux suivants.*



TAB. 2.7.1. La matrice  $A_1$ .

<i>ordre</i>	<i>coef de <math>X_1^{(1)}</math></i>	<i>coef de <math>X_2^{(1)}</math></i>
1	0,033	-0,007
2	0,014	-0,034
3	0,001	-0,003
4	0,001	-0,004
5	0,012	-0,024

TAB. 2.7.2. La matrice  $A_2$ .

<i>ordre</i>	<i>coef de <math>X_1^{(2)}</math></i>	<i>coef de <math>X_2^{(2)}</math></i>	<i>coef de <math>X_3^{(2)}</math></i>
1	0,039	0,013	0,013
2	0,035	-0,075	-0,006
3	-0,004	-0,83	0,033
4	0,021	-0,0121	-0,017
5	-0,034	0,055	-0,001

TAB. 2.7.3. La matrice  $A_3$ .

<i>ordre</i>	<i>coef de <math>X_1^{(3)}</math></i>	<i>coef de <math>X_2^{(3)}</math></i>
1	-0,081	0,193
2	-0,129	0,114
3	0,087	-0,060
4	0,193	-0,135
5	0,032	-0,010

## 2.8. GÉNÉRALISATION DE L'ANALYSE CANONIQUE SELON GUPTA.

Les généralisations qu'on a vues jusqu'à présent sont basées sur le critère de maximisation d'une fonction de corrélation. La généralisation de Gupta (1981), quant à elle, est basée sur le critère de minimisation de la variance généralisée des variables canoniques. Dans cette section on travaille avec les mêmes appellations

utilisées par Gupta (1981), à savoir “new-GCV” qui désigne “nouvelle variable canonique” et “new-GCC” qui désigne “nouvelle corrélation canonique”. Le principe de cette approche consiste à chercher, dans une étape donnée, des combinaisons linéaires (variables canoniques) des variables de chaque groupe telle que la variance généralisée de ces variables canoniques soit minimale sous contrainte que les variables canoniques sont éuicorrélées, sont à variances égales à 1 et sont indépendantes des variables canoniques des étapes précédentes. On commencera par préciser les notations qui seront utilisées dans cette section et ce sera l’objet de la section 2.8.1. Comme auparavant, on commencera d’abord par travailler au niveau de la population, la section 2.8.2 sera consacrée à cet effet, dans laquelle on va chercher dans un premier temps les premières new-GCV et new-GCC et dans un deuxième temps on va chercher les new-GCV et new-GCC d’ordre quelconque. Par la suite on va travailler au niveau de l’échantillon et ce sera l’objet de la section 2.8.4.

### 2.8.1. Notation

Soient  $X = (X^{(1)'}, X^{(2)'}, \dots, X^{(k)'})' : p \times 1$  où  $X^{(i)} : p_i \times 1$   $i = 1, 2, \dots, k$  et  $\sum_{i=1}^k p_i = p$  un vecteur aléatoire réparti en  $k$  groupes de variables aléatoires  $X^{(i)}$ ,  $i = 1, 2, \dots, k$ .

On suppose sans perte de généralité que  $\mathbf{E}(X) = \mathbf{0}$  et soit  $\Sigma$  la matrice de variance-covariance de  $X$  définie par :

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \Sigma_{k1} & \cdots & \cdots & \Sigma_{kk} \end{pmatrix} : p \times p, \quad (2.8.1)$$

et où  $\Sigma_{ij} = \text{Cov}(X^{(i)}, X^{(j)}) : p_i \times p_j$   $i, j = 1, 2, \dots, k$ .

Soit  $Y_s = (Y_s^{(1)}, Y_s^{(2)}, \dots, Y_s^{(k)})$  la  $s^{ieme}$  new-GCV avec  $Y_s^{(i)} = \alpha_s^{(i)'} X^{(i)}$  où  $\alpha_s^{(i)}$  est un vecteur de coefficients des composantes de  $X^{(i)}$  à l’étape  $s$ .

Soit  $\alpha = (\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(k)})$  un vecteur arbitraire qui sera utilisé pour trouver les new-GCV  $Y^{(1)}, Y^{(2)}, \dots, Y^{(k)}$ ,  $Y^{(i)} = \alpha^{(i)'} X^{(i)}$ , à n'importe quelle étape. Les  $\alpha^{(i)}$  sont tels que  $Y^{(1)}, Y^{(2)}, \dots, Y^{(k)}$  sont équadcorrélées. Si on pose  $Y = (Y^{(1)}, Y^{(2)}, \dots, Y^{(k)})$  ceci se traduit par :

$$\begin{aligned} \text{Var}(Y) &= \begin{pmatrix} \alpha^{(1)'} \sum_{11} \alpha^{(1)} & \alpha^{(1)'} \sum_{12} \alpha^{(2)} & \dots & \alpha^{(1)'} \sum_{1k} \alpha^{(k)} \\ \alpha^{(2)'} \sum_{21} \alpha^{(1)} & \alpha^{(2)'} \sum_{22} \alpha^{(2)} & \dots & \alpha^{(2)'} \sum_{2k} \alpha^{(k)} \\ \vdots & \vdots & \dots & \vdots \\ \alpha^{(k)'} \sum_{k1} \alpha^{(1)} & \dots & \dots & \alpha^{(k)'} \sum_{kk} \alpha^{(k)} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \rho & \dots & \dots & \rho \\ \rho & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \rho \\ \rho & \dots & \dots & \rho & 1 \end{pmatrix} \cdot \end{aligned} \quad (2.8.2)$$

Cette matrice on la note  $\sum_{\rho}$ , où  $\rho = \text{Corr}(Y^{(i)}, Y^{(j)})$  pour tout  $i, j = 1, 2, \dots, k$  désigne le coefficient de corrélation simple entre les variables new-GCV  $Y^{(i)}$  et  $Y^{(j)}$  trouvées à l'étape  $s$ .

### 2.8.2. Les premières new-GCV et new-GCC.

**Définition 2.8.1.** *La première new-GCV est le vecteur  $Y_1 = (Y_1^{(1)}, Y_1^{(2)}, \dots, Y_1^{(k)}) = (\alpha_1^{(1)'} X^{(1)}, \alpha_1^{(2)'} X^{(2)}, \dots, \alpha_1^{(k)'} X^{(k)})$  où  $\alpha_1^{(i)}$   $i = 1, 2, \dots, k$ , sont choisis de telle façon que les  $Y_1^{(i)}$   $i = 1, 2, \dots, k$  sont équadcorrélées, à variances égales à 1 et que la variance généralisée de  $Y_1$  soit minimale.*

Pour trouver les premières new-GCV, on doit minimiser  $|\sum_{\rho}|$  sous les contraintes que les new-GCV sont à variances égales à 1 et sont équadcorrélées.

Comme on a fait auparavant, pour résoudre ce problème, on utilisera la méthode

de Lagrange. Le lagrangien de ce programme est donné par :

$$\begin{aligned} \mathbf{L} = \mathbf{L}(\lambda_i, \nu_{ij} \quad i, j = 1, 2, \dots, k) &= \left| \sum_{\rho} \right| + \sum_{i=1}^k \frac{\lambda_i}{k(k-1)} \left( \alpha^{(i)'} \Sigma_{ii} \alpha^{(i)} - 1 \right) \\ &- \sum_{i, j (i \neq j)} \frac{\nu_{ij}}{2k(k-1)} \left[ k(k-1) \alpha^{(i)'} \Sigma_{ij} \alpha^{(j)} - \sum_{i, j (i \neq j)} \alpha^{(i)'} \Sigma_{ij} \alpha^{(j)} \right], \end{aligned} \quad (2.8.3)$$

où  $\sum_{i, j (i \neq j)}$  désigne une double somme par rapport à  $i$  et  $j$  pour tout  $i \neq j$ .

Le dernier terme de (2.8.3) provient du fait que les  $\text{Corr}(Y^{(i)}, Y^{(j)})$  sont toutes identiques et égales à leur moyenne :

$$\begin{aligned} \text{Corr}(Y^{(i)}, Y^{(j)}) &= \alpha^{(i)'} \Sigma_{ij} \alpha^{(j)} = \frac{1}{k(k-1)} \sum_{i, j (i \neq j)} \text{Corr}(Y^{(i)}, Y^{(j)}) \\ &= \frac{\sum_{i, j (i \neq j)} \alpha^{(i)'} \Sigma_{ij} \alpha^{(j)}}{k(k-1)} = \rho. \end{aligned} \quad (2.8.4)$$

Les  $\lambda_i$  et  $\nu_{ij}$  sont les multiplicateurs de Lagrange. En dérivant (2.8.3) par rapport à  $\alpha^{(t)}$  on trouve :

$$\begin{aligned} \frac{\partial \mathbf{L}}{\partial \alpha^{(t)}} &= \mathbf{C}(\rho) \frac{2}{k(k-1)} \sum_{j(j \neq t)} \Sigma_{tj} \alpha^{(j)} + \frac{2\lambda_t}{k(k-1)} \Sigma_{tt} \alpha^{(t)} \\ &- \frac{1}{2k(k-1)} \sum_{j(j \neq t)} \nu_{tj} \left[ k(k-1) \Sigma_{tj} \alpha^{(j)} - 2 \sum_{j(j \neq t)} \Sigma_{tj} \alpha^{(j)} \right] = 0 \end{aligned} \quad (2.8.5)$$

pour tout  $t = 1, 2, \dots, k$  où  $\mathbf{C}(\rho) = \frac{\partial |\sum_{\rho}|}{\partial \rho}$ . Dans le dernier terme de l'équation (2.8.5) on a utilisé le fait que

$$\alpha^{(i)'} \Sigma_{ij} \alpha^{(j)} = \alpha^{(j)'} \Sigma_{ji} \alpha^{(i)} \quad \forall i, j = 1, 2, \dots, k.$$

En multipliant (2.8.5) à gauche par  $\alpha^{(t)'}$  on trouve :

$$\begin{aligned} \mathbf{C}(\rho) \frac{2}{k(k-1)} \sum_{j(j \neq t)} \alpha^{(t)'} \Sigma_{tj} \alpha^{(j)} + \frac{2\lambda_t}{k(k-1)} \alpha^{(t)'} \Sigma_{tt} \alpha^{(t)} \\ - \frac{1}{k(k-1)} \sum_{j(j \neq t)} \nu_{tj} \left[ k(k-1) \alpha^{(t)'} \Sigma_{tj} \alpha^{(j)} - 2 \sum_{j(j \neq t)} \alpha^{(t)'} \Sigma_{tj} \alpha^{(j)} \right] = 0 \end{aligned} \quad (2.8.6)$$

En tenant compte du fait que  $\alpha^{(t)'} \Sigma_{tj} \alpha^{(j)} = \rho \quad \forall t \neq j \quad t, j = 1, 2, \dots, k$  et  $\alpha^{(t)'} \Sigma_{tt} \alpha^{(t)} = 1$  (2.8.6) devient :

$$\mathbf{C}(\rho) \frac{2}{k(k-1)} \sum_{j(j \neq t)} \rho + \frac{2\lambda_t}{k(k-1)} - \frac{1}{k(k-1)} \sum_{j(j \neq t)} \nu_{tj} \left[ k(k-1)\rho - 2 \sum_{j(j \neq t)} \rho \right] = 0$$

$$\mathbf{C}(\rho) \frac{2(k-1)\rho}{k(k-1)} + \frac{2\lambda_t}{k(k-1)} - \frac{1}{k(k-1)} \sum_{j(j \neq t)} \nu_{tj} [k(k-1)\rho - 2(k-1)\rho] = 0$$

$$\mathbf{C}(\rho) \frac{2\rho}{k} + \frac{2\lambda_t}{k(k-1)} - \frac{1}{k(k-1)} (k-1)(k-2)\rho \sum_{j(j \neq t)} \nu_{tj} = 0$$

$$\mathbf{C}(\rho) 2\rho(k-1) + 2\lambda_t - (k-1)(k-2)\rho \sum_{j(j \neq t)} \nu_{tj} = 0 \quad (2.8.7)$$

$$\mathbf{C}(\rho) 2\rho(k-1) + 2\lambda_{t+1} - (k-1)(k-2)\rho \sum_{j(j \neq t)} \nu_{t+1j} = 0 \quad (2.8.8)$$

Ans (2.8.7) – (2.8.8) donne :

$$2(\lambda_t - \lambda_{t+1}) + (k-1)(k-2)\rho \left[ \sum_{j(j \neq t+1)} \nu_{t+1j} - \sum_{j(j \neq t)} \nu_{tj} \right] = 0$$

$$\lambda_t - \lambda_{t+1} = \frac{(k-1)(k-2)\rho}{2} \left[ \sum_{j(j \neq t)} \nu_{tj} - \sum_{j(j \neq t)} \nu_{t+1j} \right] \quad (2.8.9)$$

Le dernier terme de (2.8.3) correspond aux contraintes d'équicorrélation entre les variables canoniques. En fait ces contraintes sont identiques, de là on conclut que les  $\nu_{ij}$  sont tous identiques, et par conséquent (2.8.9) donnera  $\lambda_t = \lambda_{t+1} = \lambda \quad \forall t = 1, 2, \dots, k-1$ . Le dernier terme de (2.8.3) sera donc nul et (2.8.5) deviendra alors :

$$\frac{\partial \mathbf{L}}{\partial \alpha^{(t)}} = \mathbf{C}(\rho) \frac{2}{k(k-1)} \sum_{j(j \neq t)} \Sigma_{tj} \alpha^{(j)} + 2 \frac{\lambda}{k(k-1)} \Sigma_{tt} \alpha^{(t)} = 0 \quad (2.8.10)$$

En multipliant à gauche (2.8.10) par  $\alpha^{(t)'}$ , on obtient :

$$\mathbf{C}(\rho) \frac{2}{k(k-1)} \sum_{j(j \neq t)} \alpha^{(t)' \Sigma_{tj}} \alpha^{(j)} + \frac{2\lambda}{k(k-1)} \alpha^{(t)' \Sigma_{tt}} \alpha^{(t)} = 0, \quad (2.8.11)$$

et en utilisant le fait que  $\alpha^{(t)' \Sigma_{tj}} \alpha^{(j)} = \rho \forall j \neq t$  et  $\alpha^{(t)' \Sigma_{tt}} \alpha^{(t)} = 1$  (2.8.11) devient :

$$\mathbf{C}(\rho) \frac{2}{k(k-1)} (k-1)\rho + \frac{2\lambda}{k(k-1)} = 0,$$

ce qui donne  $\lambda = -(k-1)\rho \mathbf{C}(\rho)$ .

Remplaçant  $\lambda$  dans (2.8.10), et en supposant que  $\mathbf{C}(\rho) \neq 0$  on trouve :

$$\sum_{j(j \neq t)} \Sigma_{tj} \alpha^{(j)} - (k-1)\rho \Sigma_{tt} \alpha^{(t)} = 0 \quad \text{pour } t = 1, 2, \dots, k. \quad (2.8.12)$$

En écriture matricielle, (2.8.12) peut être écrite sous forme :

$$\begin{pmatrix} -(k-1)\rho \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & -(k-1)\rho \Sigma_{22} & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & -(k-1)\rho \Sigma_{kk} \end{pmatrix} \begin{pmatrix} \alpha^{(1)} \\ \alpha^{(2)} \\ \vdots \\ \alpha^{(k)} \end{pmatrix} = 0 \quad (2.8.13)$$

$$\mathcal{A}\alpha = 0, \quad (2.8.14)$$

où  $\mathcal{A}$  est la matrice de gauche et  $\alpha$  est le vecteur des  $\alpha^{(j)}$  dans (2.8.13).

Une solution non triviale de (2.8.14) qui vérifie nécessairement  $\alpha^{(j)' \Sigma_{jj}} \alpha^{(j)} = 1$  pour tout  $j = 1, 2, \dots, k$  existe si et seulement si la matrice  $\mathcal{A}$  est singulière. Cela se traduit par :

$$\left| \sum_{od} -(k-1)\rho \sum_d \right| = 0 \quad \text{ou} \quad \left| \sum -\lambda^* \sum_d \right| = 0, \quad (2.8.15)$$

$$\text{avec } \sum_{od} = \begin{pmatrix} 0 & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & 0 & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & 0 \end{pmatrix},$$

$$\sum_d = \begin{pmatrix} \Sigma_{11} & 0 & \cdots & 0 \\ 0 & \Sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \Sigma_{kk} \end{pmatrix},$$

et

$$\lambda^* = 1 + (k - 1)\rho \quad (2.8.16)$$

On résout (2.8.15) pour trouver les racines en fonction des  $\lambda_i^*$ . L'équation (2.8.16) donnera alors les  $p$   $\rho_i$  correspondants. Et on choisira  $\rho_i$  pour lequel  $|\sum_{\rho_i}|$  est minimum. On peut écrire (2.8.14) autrement :

$$\begin{aligned} \mathcal{A} &= \sum_{od} -(k - 1)\rho \sum_d \\ \mathcal{A}\alpha &= \left( \sum_{od} -(k - 1)\rho \sum_d \right) \alpha = 0 \\ \sum_d^{-1} \sum_{od} \alpha &= (k - 1)\rho \alpha, \end{aligned} \quad (2.8.17)$$

donc  $(k - 1)\rho$  est une valeur propre de  $\sum_d^{-1} \sum_{od}$  et  $\alpha$  est le vecteur propre de  $\sum_d^{-1} \sum_{od}$  correspondant à la valeur propre  $(k - 1)\rho$ .

D'après l'expression de  $\sum_{\rho}$  donnée en (2.8.2), nous avons que

$$\begin{aligned} \left| \sum_{\rho} \right| &= (1 - \rho)^{k-1} [\rho(k - 1) + 1], \\ \text{et } \frac{\partial \left| \sum_{\rho} \right|}{\partial \rho} &= -k(k - 1)\rho(1 - \rho)^{k-2} = \mathbf{C}(\rho) \end{aligned} \quad (2.8.18)$$

À partir de (2.8.18) on déduit que  $|\sum_{\rho}|$  croît pour les valeurs négatives de  $\rho$  et décroît pour les valeurs positives. On a à considérer les valeurs positives et négatives de  $\rho$  pour choisir parmi elles celle, disons  $\rho^{(1)}$ , qui minimise  $|\sum_{\rho}|$ . Les vecteurs des coefficients  $\alpha^{(1)} = (\alpha_1^{(1)}, \alpha_1^{(2)}, \dots, \alpha_1^{(k)})$  correspondants à  $\rho^{(1)}$  sont

alors déterminés à partir de (2.8.13), et le premier vecteur des variables canoniques new-GCV est donné par  $Y_1 = (Y_1^{(1)}, Y_1^{(2)}, \dots, Y_1^{(k)})$  où  $Y_1^{(i)} = \alpha_1^{(i)'} X^{(i)}$  pour tout  $i = 1, 2, \dots, k$ . On a supposé auparavant que  $C(\rho) \neq 0$ , ce qui signifie d'après (2.8.18) que  $\rho$  doit être différent de 0 et de 1. Si  $|\sum_{od}| \neq 0$  alors  $\rho \neq 0$  et si  $|\sum| \neq 0$  alors  $\rho \neq 1$ .

### 2.8.3. Les new-GCV et new-GCC d'ordre quelconque.

Pour trouver les new-GCV d'ordre quelconque, d'autres contraintes s'ajouteront à celles de la section précédente. Les nouvelles contraintes vont exprimer la non corrélation des new-GCV de l'étape à laquelle on est à celles des étapes précédentes tout en supposant que les new-GCV des étapes précédentes sont indépendantes entre elles (d'une étape à l'autre). Autrement : supposons qu'on est à l'étape  $r + 1$ . Alors les nouvelles contraintes sont :

$$\text{Corr} \left( Y_{r+1}^{(i)}, Y_l^{(j)} \right) = \alpha_{r+1}^{(i)'} \Sigma_{ij} \alpha_l^{(j)} = 0 \quad (2.8.19)$$

pour  $l = 1, 2, \dots, r$ ,  $i, j = 1, 2, \dots, k$ . Le lagrangien est alors :

$$\mathbf{L}_{r+1} = \mathbf{L} + \sum_{i,j,s}^{k,k,r} \theta_{ijs} \alpha^{(i)'} \Sigma_{tj} \alpha_s^{(j)} = 0, \quad (2.8.20)$$

où  $\mathbf{L}$  est donné en (2.8.3). Le symbole  $\sum_{i,j,s}^{k,k,r}$  désigne la triple somme par rapport aux indices  $i, j$  et  $s$ ,  $i$  et  $j$  vont de 1 jusqu'à  $k$  alors que  $s$  va de 1 jusqu'à  $r$ .

On dérive  $\mathbf{L}_{r+1}$  par rapport à  $\alpha^{(t)}$  et on trouve :

$$\frac{\partial \mathbf{L}_{r+1}}{\partial \alpha^{(t)}} = \frac{\partial \mathbf{L}}{\partial \alpha^{(t)}} + \sum_{j,s}^{k,r} \theta_{tjs} \Sigma_{tj} \alpha_s^{(j)} = 0 \quad . \quad (2.8.21)$$



On multiplie à gauche par  $\alpha_l^{(t) \prime}$  et en tenant compte de (2.8.19), (2.8.21) devient :

$$\begin{aligned}
& \sum_{j,s}^{k,r} \theta_{tjs} \alpha_l^{(t) \prime} \Sigma_{tj} \alpha_s^{(j)} = 0 \\
& \sum_{j=1}^k \theta_{tjl} \alpha_l^{(t) \prime} \Sigma_{tj} \alpha_l^{(j)} + \sum_{j,s(s \neq l)}^{k,r} \theta_{tjs} \alpha_l^{(t) \prime} \Sigma_{tj} \alpha_s^{(j)} = 0 \\
& \theta_{tll} \alpha_l^{(t) \prime} \Sigma_{tt} \alpha_l^{(t)} + \rho \sum_{j(j \neq t)} \theta_{tjl} = 0 \\
& \theta_{tll} + \rho \sum_{j(j \neq t)} \theta_{tjl} = 0 \quad . \quad (2.8.22)
\end{aligned}$$

Dans  $\frac{\partial \mathbf{L}}{\partial \alpha^{(t)}}$  on a les quantités  $\Sigma_{tj} \alpha_{r+1}^{(j)}$ ;  $\Sigma_{tt} \alpha_{r+1}^{(t)}$ ;  $\Sigma_{tj} \alpha_{r+1}^{(j)}$  et  $\Sigma_{tj} \alpha_{r+1}^{(l)}$ . Si on multiplie à gauche tous ces termes par  $\alpha_l^{(t) \prime}$  où  $l < r + 1$  alors ils s'annulent tous. Ce qui donne  $\alpha_l^{(t) \prime} \frac{\partial \mathbf{L}}{\partial \alpha^{(t)}} = 0$ . En fait comme on l'a fait au chapitre 1, on peut aussi démontrer que chacune des contraintes dans (2.8.19) implique les autres, donc une seule suffit, autrement on prend tous les paramètres  $\theta_{ijs}$  égaux, et soit  $\theta$  leur valeur commune. De ce fait, (2.8.22) devient:

$$\theta [1 + (k - 1)\rho] = 0 \quad . \quad (2.8.23)$$

Supposons que  $\rho \neq \frac{-1}{k-1}$ , ce qui donne alors  $\theta = 0$ , et par conséquent (2.8.20) se résume à (2.8.3). On aura alors un système d'équations représenté par (2.8.13). Donc à l'étape  $(r + 1)$ , on choisira la racine de (2.8.15), disons  $\rho^{(r+1)}$ , qui donne la  $(r + 1)^{ieme}$  plus petite valeur de  $\left| \sum_{\rho} \right|$  qu'on appellera la  $(r + 1)$  new-GCC (new generalised canonical correlation), et  $Y^{(r+1)} = \left( \alpha_{r+1}^{(1) \prime} X^{(1)}, \alpha_{r+1}^{(2) \prime} X^{(2)}, \dots, \alpha_{r+1}^{(k) \prime} X^{(k)} \right)$  où  $\alpha_{r+1}^{(1)}, \alpha_{r+1}^{(2)}, \dots, \alpha_{r+1}^{(k)}$  sont les vecteurs des coefficients correspondants à la  $(r + 1)^{ieme}$  new-GCV.  $Y^{(r+1)}$  est appelé la  $(r + 1)^{ieme}$  new-GCV.

#### 2.8.4. Les new-GCV échantillonnales.

Souvent en pratique, on ne dispose pas des paramètres de population  $\mu$  et  $\Sigma$ . Comme on a vu, les new-GCV et les new-GCC dépendent de  $\mu$  et  $\Sigma$  qu'on ignore. Donc on est amené à estimer  $\rho_s$  et  $\alpha_s = \left( \alpha_s^{(1)}, \alpha_s^{(2)}, \dots, \alpha_s^{(k)} \right)$ . Pour simplifier le

problème, on suppose que  $X \sim \mathbf{N}_p(\mu, \Sigma)$ , et soit  $X_\alpha = \left( X_\alpha^{(1)'}, X_\alpha^{(2)'}, \dots, X_\alpha^{(k)'} \right)$   $\alpha = 1, 2, \dots, n$  un échantillon aléatoire issu de  $X \sim \mathbf{N}_p(\mu, \Sigma)$ . Et soit  $\bar{X}$  et  $\mathbf{S}$  la moyenne et la variance échantillonnale. Les estimateurs de maximum de vraisemblance  $\hat{\Sigma}_{ij}$ ,  $\hat{\Sigma}$  respectivement de  $\Sigma_{ij}$  et  $\Sigma$  sont  $\mathbf{S}_{ij}$  et  $\mathbf{S}$ ,  $\mathbf{S}$  est donnée en (2.8.24).

Puisque  $\rho_s$  et  $\Sigma$  sont exprimés en fonction de  $\Sigma_d$ ,  $\Sigma$ , les estimateurs de maximum de vraisemblance  $\hat{\rho}_s = r_s$  et  $\hat{\alpha}_s = a_s$  de  $\rho_s$  et  $\alpha_s$  (resp) sont obtenus en remplaçant les  $\Sigma_{ij}$  dans  $\Sigma_d$  par les  $S_{ij}$  et  $\Sigma$  par  $\mathbf{S}$  dans les expressions de  $\rho_s$  et  $\alpha_s$  où

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{pmatrix}. \quad (2.8.24)$$

**Exemple 2.8.1.** *En annexe1, on retrouve le programme 5 codé en Splus. Il calcule les new-GCC et les new-GCV de l'exemple sur les mouches. Le tableau 2.8.1 donne  $|\Sigma_\rho|$  pour chaque valeur de  $\rho$  trouvée.*

TAB. 2.8.1. Les new-GCC possibles

$\rho$	-0.83	0.53	0.33	-0.30	0.23	0.05	-0.01
$ \Sigma_\rho $	-2.22	0.46	0.74	0.68	0.87	0.99	1

La première new-GCC est celle qui donne le minimum de  $|\Sigma_\rho|$ , sa valeur est  $\rho^{(1)} = -0.83$ . La second new-GCC est celle qui donne le minimum de  $|\Sigma_\rho|$  parmi ceux qui restent, sa valeur est 0.53. Les  $\alpha^{(s)}$ : coefficients des new-GCV correspondant à ces new-GCC sont donnés dans le tableau 2.8.2. L'ordre est celui du tableau 2.8.1.

TAB. 2.8.2. Les new-GCV

$\alpha^{(s)} \setminus rang$	1	2	3	4	5	6	7
$\alpha^{(1)}$	-0.334	0.733	0.136	0.132	-0.180	0.037	-0.049
	0.071	-0.068	-0.185	-0.335	0.356	-0.105	0.132
$\alpha^{(2)}$	-0.401	-0.555	0.257	0.345	0.560	0.586	0.168
	-0.130	-0.570	0.234	-0.754	-0.836	-0.350	0.168
	-0.137	-0.255	0.111	-0.060	0.022	-0.490	3.218
$\alpha^{(3)}$	0.909	0.670	1.420	-1.316	-0.296	5.480	-3.347
	-2.146	-2.071	-2.725	1.208	-0.222	-3.823	2.300

# Chapitre 3

---

## ANALYSE CANONIQUE: INFLUENCE ET ROBUSTESSE

### 3.1. PERTURBATION ET FONCTION D'INFLUENCE AU NIVEAU DE LA POPULATION.

#### 3.1.1. Notion de perturbation.

Soit  $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} : p \times 1$  où  $X^{(1)} : p_1 \times 1$  et  $X^{(2)} : p_2 \times 1$  ( $p_1 \leq p_2$ ). Supposons que  $X$  suit une distribution cumulative  $\mathbf{F}$  de moyenne  $\mu(\mathbf{F}) = \mu$  et de matrice de covariance  $\Sigma(\mathbf{F}) = \Sigma$ , c'est à dire

$$\mu(\mathbf{F}) = \mu = \int x d\mathbf{F}(x) = \begin{pmatrix} \mu_1(\mathbf{F}) \\ \mu_2(\mathbf{F}) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

et

$$\begin{aligned} \Sigma(\mathbf{F}) = \Sigma &= \int [x - \mu(\mathbf{F})][x - \mu(\mathbf{F})]' d\mathbf{F}(x) \\ &= \begin{pmatrix} \Sigma_{11}(\mathbf{F}) & \Sigma_{12}(\mathbf{F}) \\ \Sigma_{21}(\mathbf{F}) & \Sigma_{22}(\mathbf{F}) \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \end{aligned}$$

où  $\mu_i = \mu_i(\mathbf{F}) = \mathbf{E}(X^{(i)})$  et  $\Sigma_{ij} = \Sigma_{ij}(\mathbf{F}) = \mathbf{Cov}(X^{(i)}, X^{(j)}) \quad i, j = 1, 2$ .

Soient  $\rho_j^2(\mathbf{F}) = \rho_j^2 \quad j = 1, 2, \dots, p_1$  les carrés des corrélations canoniques entre  $X^{(1)}$  et  $X^{(2)}$ , qui sont les racines de l'équation :

$$|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^2\Sigma_{11}| = 0.$$

Et soient  $\alpha_j, \beta_j$  des vecteurs canoniques satisfaisant les équations :

$$\begin{aligned} (\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho_j^2\Sigma_{11}) \alpha_j &= 0 \\ (\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \rho_j^2\Sigma_{11}) \beta_j &= 0 \\ \alpha_j' \Sigma_{11} \alpha_j &= \beta_j' \Sigma_{22} \beta_j = 1 \\ \alpha_j' \Sigma_{11} \alpha_{j'} &= \beta_j' \Sigma_{22} \beta_{j'} = 0 \quad \text{pour tout } j \neq j'. \end{aligned}$$

Rappelons que  $\alpha_j X^{(1)}$  et  $\beta_j X^{(2)}$  sont les  $j^{ieme}$  variables canoniques.

Soit  $\delta_z$  une distribution mettant toute sa masse au point  $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ ,  $z_j \in \mathbf{R}^{p_j}$

$j = 1, 2$ . Et soit  $w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = z - \mu$ ,  $w_j = z_j - \mu_j$   $j = 1, 2$ .

Soit  $\epsilon$  un scalaire tel que  $0 < \epsilon < 1$ . Une façon standard pour perturber  $\mathbf{F}$  est de lui ajouter une petite contribution à partir de la distribution  $\delta_z$ . Soit  $\tilde{\mathbf{F}} = (1-\epsilon)\mathbf{F} + \epsilon\delta_z$  cette perturbation. Alors :

$$\begin{aligned} \mu [(1-\epsilon)\mathbf{F} + \epsilon\delta_z] &= \int x d[(1-\epsilon)\mathbf{F} + \epsilon\delta_z](x) \\ &= (1-\epsilon) \int x d(\mathbf{F}) + \epsilon \int x d[\delta_z(x)] \\ &= (1-\epsilon)\mu + \epsilon z = \mu + \epsilon(z - \mu) = \mu + \epsilon w \quad (3.1.2) \\ &= \begin{pmatrix} \mu_1 + \epsilon w_1 \\ \mu_2 + \epsilon w_2 \end{pmatrix}, \end{aligned}$$

et

$$\begin{aligned}
\sum [(1 - \epsilon)\mathbf{F} + \epsilon\delta_z] &= \int (x - \mu - \epsilon w)(x - \mu - \epsilon w)' [(1 - \epsilon)d\mathbf{F}(x) + \epsilon d\delta_z(x)] \\
&= (1 - \epsilon) \int [(x - \mu) - \epsilon w] [(x - \mu)' - \epsilon w'] d\mathbf{F}(x) \\
&\quad + \epsilon \int [(x - \mu) - \epsilon w] [(x - \mu)' - \epsilon w'] d\delta_z(x) \\
&= (1 - \epsilon) \left[ \int (x - \mu)(x - \mu)' d\mathbf{F}(x) - \epsilon \int (x - \mu)w' d\mathbf{F}(x) \right. \\
&\quad \left. - \epsilon w \int (x - \mu)' d\mathbf{F}(x) + \epsilon^2 w w' \int d\mathbf{F}(x) \right] \\
&\quad + \epsilon \int (x - \mu)(x - \mu)' d\delta_z(x) - \epsilon^2 \int (x - \mu)w' d\delta_z(x) \\
&\quad - \epsilon^2 w \int (x - \mu)' d\delta_z(x) \quad + \epsilon^3 w w' \int d\delta_z(x) \\
&= (1 - \epsilon) \left[ \sum + \epsilon^2 w w' \right] + \epsilon w w' - 2\epsilon^2 w w' + \epsilon^3 w w' \\
&= \sum + \epsilon \left( w w' - \sum \right) - \epsilon^2 w w' \\
&= \begin{pmatrix} \Sigma_{11}(\epsilon) & \Sigma_{12}(\epsilon) \\ \Sigma_{21}(\epsilon) & \Sigma_{22}(\epsilon) \end{pmatrix}, \tag{3.1.3}
\end{aligned}$$

où

$$\Sigma_{ij}(\epsilon) = \Sigma_{ij} + \epsilon(w_i w_j' - \Sigma_{ij}) - \epsilon^2 w_i w_j', \quad i, j = 1, 2 \tag{3.1.4}$$

pour tout  $z \in \mathbf{R}^p$  et  $0 < \epsilon < 1$ ,  $\Sigma_{jj}(\epsilon)$  est définie positive. Un développement à l'ordre 2 de  $\Sigma_{jj}^{-1}(\epsilon)$  est donné par Romanazzi (1992)

$$\begin{aligned}
\Sigma_{jj}^{-1}(\epsilon) &= (1 - \epsilon)^{-1} \left[ \Sigma_{jj}^{-1} - \epsilon \left( 1 + \epsilon w_j' \Sigma_{jj}^{-1} w_j \right)^{-1} \Sigma_{jj}^{-1} w_j w_j' \Sigma_{jj}^{-1} \right] \\
&= \Sigma_{jj}^{-1} + \epsilon \left( \Sigma_{jj}^{-1} - \Sigma_{jj}^{-1} w_j w_j' \Sigma_{jj}^{-1} \right) \\
&\quad + \epsilon^2 \left[ \Sigma_{jj}^{-1} - (1 - k_j) \Sigma_{jj}^{-1} w_j w_j' \Sigma_{jj}^{-1} \right] + \mathbf{O}(\epsilon^3) \tag{3.1.5}
\end{aligned}$$

où  $k_j = w_j' \Sigma_{jj}^{-1} w_j$   $j = 1, 2$ . Soient  $Q_j = \Sigma_{ij} \Sigma_{jj}^{-1} \Sigma_{ji}$  et  $\hat{w}_i = \Sigma_{ij} \Sigma_{jj}^{-1} w_j$ ,  $i, j = 1, 2$  et  $i \neq j$ . En utilisant (3.1.4) et (3.1.5) on obtient une approximation de la

perturbation  $Q_j(\epsilon)$  de  $Q_j$ . c'est à dire

$$\begin{aligned} Q_j(\epsilon) &= Q_j + \epsilon(\hat{w}_i w_i' + w_i \hat{w}_i' - \hat{w}_i \hat{w}_i' - Q_j) \\ &+ \epsilon^2 \left[ (1 + k_j)(\hat{w}_i \hat{w}_i' - \hat{w}_i w_i - w_i \hat{w}_i') + k_j w_i w_i' \right] + \mathbf{O}(\epsilon^3). \end{aligned} \quad (3.1.6)$$

Les  $\rho_j^2, j = 1, 2, \dots, p_1$  sont les racines de  $|Q_i - \rho^2 \Sigma_{jj}| = 0$ . Alors les perturbations de  $\rho_j^2, \alpha_j$  et  $\beta_j$  sont données par (Romanazzi (1989)), Théorème 2 et remarque 3) et nous avons :

$$\begin{cases} \rho_k(\epsilon)^2 = \rho_k^2 + \epsilon \gamma_k + \frac{1}{2} \epsilon^2 \varphi_k + \mathbf{O}(\epsilon^3) \\ \alpha_k(\epsilon) = \alpha_k + \epsilon \pi_k + \frac{1}{2} \epsilon^2 \zeta_k + \mathbf{O}(\epsilon^3) \\ \beta_k(\epsilon) = \beta_k + \epsilon \tau_k + \frac{1}{2} \epsilon^2 \xi_k + \mathbf{O}(\epsilon^3) \end{cases} \quad (3.1.7)$$

où  $\gamma_k, \pi_k$  et  $\tau_k$  sont donnés à l'équation (3.1.8) et  $\varphi_k, \zeta_k, \xi_k$  sont donnés aux équations (3.1.9) à (3.1.11). Pour simplifier la notation on pose  $G_j = \Sigma_{11}^{-\frac{1}{2}} Q_2 \Sigma_{11}^{-\frac{1}{2}} - \rho_j^2 \mathcal{I}_{p_1}$ ,  $M_j = \Sigma_{11}^{-\frac{1}{2}} G_j^+ \Sigma_{11}^{-\frac{1}{2}}$  et  $K_j = \Sigma_{22}^{-\frac{1}{2}} Q_1 \Sigma_{22}^{-\frac{1}{2}} - \rho_j^2 \mathcal{I}_{p_2}$ ,  $N_j = \Sigma_{22}^{-\frac{1}{2}} K_j^+ \Sigma_{22}^{-\frac{1}{2}}$  où  $G^+$  et  $K^+$  sont les pseudo-inverses (appelées aussi l'inverse de Moore et Penrose) des matrices  $G_j$  et  $K_j$ .

Rappelons que si  $A^+$  est la pseudo-inverse de  $A$ , alors  $A^+$  vérifie :

$$\begin{aligned} (a) \quad AA^+A &= A & (c) \quad A^+AA^+ &= A^+ \\ (c) \quad (AA^+)' &= AA^+ & (d) \quad (A^+A)' &= A^+A, \end{aligned}$$

la matrice  $A^+$  existe toujours et elle est unique. On pose aussi  $u_j = \alpha_j'(z^{(1)} - \mu_1)$  et  $v_j = \beta_j'(z^{(2)} - \mu_2)$ . Les coefficients de  $\epsilon$  en (3.1.7) sont :

$$\begin{cases} \gamma_k = -\rho_k^2 u_k^2 + 2\rho_k u_k v_k - \rho_k^2 v_k^2, \\ \pi_k = -\rho_k(v_k - \rho_k u_k) M_k w_1 - (u_k - \rho_k v_k) M_k \hat{w}_1 + \frac{1}{2}(1 - u_k^2) \alpha_k, \\ \tau_k = -\rho_k(u_k - \rho_k v_k) N_k w_2 - (v_k - \rho_k u_k) N_k \hat{w}_2 + \frac{1}{2}(1 - v_k^2) \beta_k \end{cases} \quad (3.1.8)$$

et les coefficients de  $\epsilon^2$  sont :

$$\varphi_k = 2 \left[ k_2(u_k - \rho_k v_k)^2 - \gamma_k u_k^2 - \pi'_k \Sigma_{11}^{\frac{1}{2}} G_k \Sigma_{11}^{\frac{1}{2}} \pi_k \right] \quad (3.1.9)$$

$$\begin{aligned} \zeta_k &= 2 \left[ \gamma_k u_k - k_2(u_k - \rho_k v_k) + \pi'_k(\rho_k^2 w_1 - \hat{w}_1) \right] M_k w_1 \\ &\quad - 2\gamma_k(u_k - \rho_k v_k) M_k \Sigma_{11} M_k \hat{w}_1 - 2\rho_k \gamma_k (v_k - \rho_k u_k) M_k \Sigma_{11} M_k w_1 \\ &\quad + 2 \left[ k_2(u_k - \rho_k v_k) - \pi'_k(w_1 - \hat{w}_1) \right] M_k \hat{w}_1 \\ &\quad + (1 - 2u_k \pi'_k w_1 - \pi'_k \Sigma_{11} \pi_k) \alpha_k \end{aligned} \quad (3.1.10)$$

$$\begin{aligned} \xi_k &= 2 \left[ \gamma_k v_k - k_1(v_k - \rho_k u_k) + \tau'_k(\rho_k^2 w_2 - \hat{w}_2) \right] N_k w_2 \\ &\quad - 2\gamma_k(v_k - \rho_k u_k) N_k \Sigma_{22} N_k \hat{w}_2 - 2\rho_k \gamma_k (u_k - \rho_k v_k) N_k \Sigma_{22} N_k w_2 \\ &\quad + 2 \left[ k_1(v_k - \rho_k u_k) - \tau'_k(w_2 - \hat{w}_2) \right] N_k \hat{w}_2 \\ &\quad + (1 - 2v_k \tau'_k w_2 - \tau'_k \Sigma_{22} \tau_k) \beta_k. \end{aligned} \quad (3.1.11)$$

### 3.2. LES FONCTIONS D'INFLUENCE DES CORRÉLATIONS ET DES VARIABLES CANONIQUES.

**Définition 3.2.1.** Soit  $\Gamma = \Gamma(\mathbf{F})$  un paramètre exprimé comme une fonctionnelle de la distribution cumulative  $\mathbf{F}$ .

La fonction d'influence de  $\Gamma$  au point  $z$  en la distribution  $\mathbf{F}$  est défini par :

$$IF[z, \Gamma(\mathbf{F})] = \lim_{\epsilon \rightarrow 0^+} \frac{\Gamma[(1 - \epsilon)\mathbf{F} + \epsilon\delta_z] - \Gamma(\mathbf{F})}{\epsilon}, \quad (3.2.1)$$

où  $\delta_z$  est la distribution de Dirac qui met tout sa masse au point  $z$ .

La fonction d'influence  $IF[z, \Gamma(\mathbf{F})]$  donne une mesure de l'influence du point  $z$  sur la fonction  $\Gamma$  si la distribution sous-jacente est  $\mathbf{F}$ . On s'attend pour un estimateur robuste à une fonction d'influence bornée.



On peut aussi écrire (3.2.1) autrement, en effet :

$$\begin{aligned}
IF[z, \Gamma(\mathbf{F})] &= \lim_{\epsilon \rightarrow 0^+} \frac{\Gamma[(1-\epsilon)\mathbf{F} + \epsilon\delta_z] - \Gamma(\mathbf{F})}{\epsilon} \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\Gamma[\tilde{\mathbf{F}}(\epsilon)] - \Gamma[\tilde{\mathbf{F}}(0)]}{\epsilon - 0} \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\Gamma[\tilde{\mathbf{F}}(\epsilon)] - \Gamma[\tilde{\mathbf{F}}(0)]}{\tilde{\mathbf{F}}(\epsilon) - \tilde{\mathbf{F}}(0)} \frac{\tilde{\mathbf{F}}(\epsilon) - \tilde{\mathbf{F}}(0)}{\epsilon - 0} \\
&= \left( \frac{\partial [\Gamma(\tilde{\mathbf{F}})]}{\partial \tilde{\mathbf{F}}} \frac{\partial \tilde{\mathbf{F}}}{\partial \epsilon} \right) (0) = \frac{\partial \Gamma(\tilde{\mathbf{F}})}{\partial \epsilon} (0) .
\end{aligned}$$

Donc  $IF[z, \Gamma(\mathbf{F})]$  est la dérivée de la fonctionnelle  $\Gamma[\tilde{\mathbf{F}}(\epsilon)]$  au point  $\epsilon = 0$ , rappelons que  $\tilde{\mathbf{F}}$  est une perturbation de  $\mathbf{F}$ . Donc pour avoir les fonctions d'influence des corrélations et variables canoniques (qui sont des fonctionnelles de  $\mathbf{F}$ ) il suffit de dériver leurs perturbations qui sont des fonctionnelles de  $\tilde{\mathbf{F}}$ .

En dérivant les équations en (3.1.7) par rapport à  $\epsilon$  et en calculant ses dérivées au point  $\epsilon = 0$ , on trouve:

$$IF[z, \rho_k^2] = \gamma_k,$$

$$IF[z, \alpha_k] = \pi_k,$$

$$IF[z, \beta_k] = \tau_k,$$

où  $\Gamma_k$ ,  $\pi_k$  et  $\tau_k$  sont données en (3.1.8).

### 3.3. PERTURBATION ET FONCTION D'INFLUENCE AU NIVEAU DE L'ÉCHANTILLON.

Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire issu de la distribution  $\mathbf{F}$ , et soit  $\hat{\mathbf{F}}$  sa distribution cumulative empirique. Dans le cadre de cette section en verra trois types de fonction d'influence échantillonnale, à savoir :

(i) : la fonction d'influence empirique: IFE .

(ii) : la fonction d'influence empirique restreinte: IFE<sub>(i)</sub>.

(iii) : la fonction d'influence échantillonnale : IFS.

La fonction d'influence empirique IFE est obtenue à partir de la définition en (3.2.1) en remplaçant  $\mathbf{F}$  par  $\hat{\mathbf{F}}$ . La fonction d'influence empirique restreinte  $IFE_{(i)}$  évaluée au point  $X_i$  est obtenue en remplaçant  $\mathbf{F}$  dans  $IF[z, \Gamma(\mathbf{F})]$  par  $\hat{\mathbf{F}}_{(i)}$  où  $\hat{\mathbf{F}}_{(i)}$  est la fonction de distribution cumulative empirique obtenue en excluant de l'échantillon l'observation  $X_i$ . Ces deux fonctions d'influence sont utilisées dans le cas d'échantillons de grandes tailles.

Si la taille de l'échantillon est petite, on utilise la fonction d'influence échantillonnale IFS définie par :

$$IFS[X_i, \Gamma(\mathbf{F})] = -(n-1) \left[ \Gamma(\hat{\mathbf{F}}_{(i)}) - \Gamma(\hat{\mathbf{F}}) \right] \quad i = 1, 2, \dots, n. \quad (3.3.1)$$

Notons qu'au niveau de l'échantillon, les corrélations et variables canoniques ( $r_j$ ,  $a_j$  et  $b_j$ ) ont les mêmes expressions qu'au niveau de la population à la seule différence que les  $\Sigma_{ij}$  sont remplacées par les  $S_{ij}$ .

Les  $r_j^2$  sont les racines de  $|S_{12}S_{22}^{-1}S_{21} - \rho^2 S_{11}| = 0$  et  $a_j$  et  $b_j$  sont respectivement les solutions de  $(S_{12}S_{22}^{-1}S_{21} - r_j^2 S_{11}) a_j = 0$  et  $(S_{21}S_{11}^{-1}S_{12} - r_j^2 S_{22}) b_j = 0$  et elles vérifient les contraintes de normalisation et d'indépendance.

### 3.3.1. Fonctions d'influence empiriques : IFE

À partir de la définition de la fonction d'influence empirique, il s'en suit que :

$$\begin{cases} IFE \left[ z, \rho_k^2 \left( \hat{\mathbf{F}} \right) \right] = \hat{\gamma}_k \\ IFE \left[ z, \alpha_k \left( \hat{\mathbf{F}} \right) \right] = \hat{\pi}_k \\ IFE \left[ z, \beta_k \left( \hat{\mathbf{F}} \right) \right] = \hat{\tau}_k \end{cases} \quad (3.3.2)$$

où  $\hat{\gamma}_k$ ,  $\hat{\pi}_k$  et  $\hat{\tau}_k$  sont obtenues en remplaçant  $\mu$ ,  $\Sigma$ ,  $\rho_j$ ,  $\alpha_j$  et  $\beta_j$  respectivement par  $\bar{X}$ ,  $S$ ,  $r_j$ ,  $a_j$  et  $b_j$ . Les valeurs des fonctions en (3.3.2) au point  $X_i$  sont

respectivement  $\hat{\gamma}_{ik}$ ,  $\hat{\pi}_{ik}$  et  $\hat{\tau}_{ik}$  qui sont données par :

$$\hat{\gamma}_{ik} = -r_k^2 u_{ik}^2 + 2r_k u_{ik} v_{ik} - r_k^2 v_{ik}^2,$$

$$\hat{\pi}_{ik} = (u_{ik} - r_k v_{ik}) \hat{a}_k [(-1, 0, 1)] + r_k (v_{ik} - r_k u_{ik}) \hat{a}_k [(-1, 1, 0)] + \frac{1}{2} (1 - u_{ik}^2) a_k,$$

$$\hat{\tau}_{ik} = (v_{ik} - r_k u_{ik}) \hat{b}_k [(-1, 0, 1)] + r_k (u_{ik} - r_k v_{ik}) \hat{b}_k [(-1, 1, 0)] + \frac{1}{2} (1 - v_{ik}^2) b_k,$$

où

$$\hat{a}_k [(l, s, t)] = \sum_{j=1, j \neq k}^{p_1} (r_k^2 - r_j^2)^l (a'_k w_{i1})^s (a'_j \hat{w}_{i1})^t a_j,$$

$$\hat{b}_k [(l, s, t)] = \sum_{j=1, j \neq k}^{p_2} (r_k^2 - r_j^2)^l (b'_k w_{i2})^s (b'_j \hat{w}_{i2})^t b_j,$$

avec  $w_{ig} = X_i^{(g)} - \bar{X}^{(g)}$  et  $\hat{w}_{ij} = S_{g'g} S_{gg}^{-1} w_{ig}$ ,  $g, g' \in \{1, 2\}$  et  $g \neq g'$ .

### 3.3.2. Fonctions d'influence empiriques restreintes : $IFE_{(i)}$ .

La fonction d'influence empirique restreinte a la même définition que IFE, sauf que maintenant on travaille avec un échantillon de taille  $(n - 1)$  après avoir éliminé l'observation  $X_i$ .

Romanazzi(1992) a donné une approximation des  $IFE_{(i)} [X_i, \cdot]$  exprimées en fonction des termes de  $IFE [X_i, \cdot]$ . Ces approximations sont :

$$IFE_{(i)} [X_i, \rho_k^2(\hat{\mathbf{F}}_{(i)})] = \hat{\gamma}_{(i)k} = \hat{\gamma}_{ik} - (n - 1)^{-1} (\hat{\varphi}_{ik} - \hat{\gamma}_{ik}) + \mathbf{O} [(n - 1)^2],$$

$$IFE_{(i)} [X_i, \alpha_k^2(\hat{\mathbf{F}}_{(i)})] = \hat{\pi}_{(i)k} = \hat{\pi}_{ik} - (n - 1)^{-1} (\hat{\zeta}_{ik} - \hat{\pi}_{ik}) + \mathbf{O} [(n - 1)^2],$$

$$IFE_{(i)} [X_i, \beta_k^2(\hat{\mathbf{F}}_{(i)})] = \hat{\tau}_{(i)k} = \hat{\tau}_{ik} - (n - 1)^{-1} (\hat{\xi}_{ik} - \hat{\tau}_{ik}) + \mathbf{O} [(n - 1)^2].$$

### 3.3.3. Fonctions d'influence échantillonnales : IFS .

On rappelle que les fonctions d'influence échantillonnales sont utilisées quand la taille de l'échantillon n'est pas assez grande, et elles sont déduites à partir de (3.3.1). On se limitera à donner leurs expressions qui sont aussi exprimées en

fonction des termes des  $IFE [X_{i,\cdot}]$  .

$$\begin{aligned} IFS [X_{i,\hat{\rho}_k}(\hat{\mathbf{F}})] &= -(n-1) (\hat{\rho}_{(i)k} - \hat{\rho}_k) = \hat{\gamma}_{ik} - \frac{1}{2}(n-1)^{-1}\hat{\varphi}_{ik} + \mathbf{O} [(n-1)^{-2}] \quad , \\ IFS [X_{i,\hat{\alpha}_k}(\hat{\mathbf{F}})] &= -(n-1) (\hat{\alpha}_{(i)k} - \hat{\alpha}_k) = \hat{\pi}_{ik} - \frac{1}{2}(n-1)^{-1}\hat{\zeta}_{ik} + \mathbf{O} [(n-1)^{-2}] \quad , \\ IFS [X_{i,\hat{\beta}_k}(\hat{\mathbf{F}})] &= -(n-1) (\hat{\beta}_{(i)k} - \hat{\beta}_k) = \hat{\tau}_{ik} - \frac{1}{2}(n-1)^{-1}\hat{\xi}_{ik} + \mathbf{O} [(n-1)^{-2}] \quad . \end{aligned}$$

Quand  $n \rightarrow +\infty$ ,  $IFE_{(i)}$  et IFS donnent le même résultat que IFE, ce qui n'est pas le cas si  $n$  est petit.

Pour plus de détails sur ces fonctions on peut se référer à Romanazzi(1992) .

### 3.4. MÉTHODES ROBUSTES POUR L'ANALYSE CANONIQUE.

Dans la section précédente on a donné les fonctions d'influence des variables et corrélations canoniques qui, comme on l'a vu au niveau de l'échantillon, utilisent des estimateurs des paramètres  $\Sigma$  et  $\mu$ . Toutefois on n'a pas parlé de la nature de ces estimateurs et de leurs qualité(robustesse). Dans le cadre de cette section, on donnera une série de cinq estimateurs robustes sans les comparer. Dehon et al. (2000) ont appliqué quatre approches robustes à l'analyse canonique de deux groupes de variables aléatoires, et ils ont comparé, à l'aide d'une simulation, les résultats de ces approches à l'analyse canonique classique. On présentera dans cette section les trois des approches en question, à savoir :

- 1** : l'utilisation du "Minimum Covariance Determinant": MCD;
- 2** : l'utilisation des "signes"des données;
- 3** : l'utilisation de la régression alternée robuste.

et les deux dernières approches sont le M-estimateur et le S-estimateur.

### 3.4.1. Première approche : utilisation du “Minimum Covariance Determinant”: MCD.

Soient  $X^{(1)} : p_1 \times 1$  et  $X^{(2)} : p_2 \times 1$  deux vecteurs aléatoires et soit  $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} : p \times 1, p = p_1 + p_2$ .  $X$  est observé sur un échantillon aléatoire de taille  $n$  et soient  $X_\alpha = \begin{pmatrix} X_\alpha^{(1)} \\ X_\alpha^{(2)} \end{pmatrix} \alpha = 1, 2, \dots, n$  ces observations. L'estimateur MCD, introduit par Rousseeuw (1985), consiste à trouver un sous-échantillon, de taille  $h$  ( $1 \leq h \leq n$ ), de l'échantillon de départ tel que la variance généralisée (le déterminant de la matrice de covariance) calculée à partir de ce sous-échantillon est le minimum des déterminants des matrices de variance-covariance de tous les sous-échantillons possibles de taille  $h$ . Soit  $(X_{i_1}, \dots, X_{i_h})$  les observations du vecteur  $X$  sur le sous-échantillon qui donne le plus petit déterminant recherché. L'indice  $i_k$   $k = 1, 2, \dots, h$  est utilisé uniquement pour éviter toute confusion. L'estimateur MCD de position est défini par

$$T_n = \frac{1}{h} \sum_{k=1}^h X_{i_k},$$

et l'estimateur MCD de la matrice de variance-covariance est défini par

$$C_n = c_1 \frac{1}{h} \sum_{k=1}^h (X_{i_k} - T_n)(X_{i_k} - T_n)' \quad \text{où} \quad c_1 = \frac{1 - \alpha}{\mathcal{F}_{\chi_{p+2}^2}(q_\alpha)},$$

$\mathcal{F}_{\chi_{p+2}^2}$  étant la fonction de répartition de la loi du khi-deux à  $p + 2$  degrés de liberté.  $c_1$  est un facteur de convergence pour une distribution normale, et  $q_\alpha = \chi_{p, 1-\alpha}^2$  est le quantile d'ordre  $(1 - \alpha)$  de la distribution khi-deux à  $p$  degrés de liberté. Généralement,  $h$  prend les valeurs  $h = n(1 - \alpha)$ , avec  $\alpha = 0.5$  ou  $\alpha = 0.25$ . Dans le premier cas, l'estimateur aura le plus haut point de rupture (50%) et dans le second cas on aura un meilleur compromis entre efficacité et haut point de rupture (qui sera 25%). On rappelle que le point de rupture d'un estimateur est défini comme étant la plus grande proportion de points aberrants

dans les données ne faisant pas exploser ou imploser l'estimateur. Des propriétés théoriques de l'estimateur MCD ont été étudiées par Butler et al. (1993). La fonction d'influence de l'estimateur MCD d'échelle multivarié à été trouvé par Rousseeuw & Van Driessen (1999). Pour améliorer l'efficacité des estimateurs à haut point de rupture, il est préférable d'utiliser leur version pondérée en une étape, celle-ci garde le point de rupture initial. Les estimateurs RMCD( MCD pondérés à une étape) sont définis par :

$$T_n^R = \frac{\sum_{\alpha=1}^n w_\alpha X_\alpha}{\sum_{\alpha=1}^n w_\alpha}$$

$$C_n^R = c_2 \frac{\sum_{\alpha=1}^n w_\alpha (X_\alpha - T_n^R)(X_\alpha - T_n^R)'}{\sum_{\alpha=1}^n w_\alpha}$$

où

$$c_2 = \frac{1 - \delta}{\mathcal{F}_{\chi_{p+2}^2}(q\delta)},$$

est un facteur de convergence. Les poids sont calculés comme suit :

$$w_\alpha = \begin{cases} 1 & \text{si } (X_\alpha - T_n)' C_n^{-1} (X_\alpha - T_n) \leq q\delta, \\ 0 & \text{sinon} \end{cases}$$

où  $T_n$  et  $C_n$  sont les estimateurs MCD non pondérés.

### 3.4.2. Deuxième approche: utilisation des "signes" des données.

On peut aussi tenir compte de l'influence des données aberrantes en travaillant avec "les signes" des données plutôt qu'avec les données originales. Une fois les signes des données calculés, on calcule la matrice des covariances ordinaire à partir de ces signes ( voir Visuri et al. (2000)). On entend par "signe" de  $X_\alpha^{(i)}$  la quantité :

$$S_\alpha^i = S(X_\alpha^{(i)}) = \frac{X_\alpha^{(i)} - m_{X^{(i)}}}{\|X_\alpha^{(i)} - m_{X^{(i)}}\|} : p_i \times 1 \quad i = 1, 2$$

où  $m_{X^{(i)}}$  est la médiane du nuage de points  $X_\alpha^{(i)}$   $\alpha = 1, 2, \dots, n$  définie par

$$m_{X^{(i)}} = \arg \min_m \sum_{\alpha=1}^n \|X_\alpha^{(i)} - m\| \quad : p_i \times 1 \quad i = 1, 2.$$

Donc une fois  $m_{X^{(i)}}$ ,  $S_\alpha^i$  pour  $i = 1, 2, \dots, n$  calculés, on calcule à partir des signes les matrices de covariances ordinaires  $S_{g_{ij}}$   $i = 1, 2$  définies par :

$$S_{g_{ij}} = \frac{A_{ij}}{n}$$

où

$$A_{ij} = \sum_{\alpha=1}^n (S_\alpha^i - \bar{S}^i)(S_\alpha^j - \bar{S}^j)' \quad i, j = 1, 2$$

et

$$\bar{S}^i = \frac{1}{n} \sum_{\alpha=1}^n S_\alpha^i \quad i = 1, 2 \quad .$$

Les variables et corrélations canoniques seront calculées à partir de la matrice  $\mathcal{M} = S_{g_{11}}^{-1} S_{g_{12}} S_{g_{22}}^{-1} S_{g_{21}}$

### 3.4.3. Troisième approche : utilisation de la régression alternée.

La régression alternée proposé par Wold (1966) peut aussi être utiliser pour l'estimation des variables canoniques. Soient  $\alpha' X^{(1)}$ ,  $\beta' X^{(2)}$  les variables canoniques recherchées. On sait que pour un  $\alpha$  donné tel que  $\alpha' X^{(1)}$  est une variable canonique,

$$\beta = \arg \max_b \left| \text{Corr}(\alpha' X^{(1)}, b' X^{(2)}) \right|.$$

Dans le contexte de la régression multiple,  $\beta$  est proportionnel au coefficient  $b$  dans le modèle suivant :

$$\alpha' X^{(1)} = b' X^{(2)} + \gamma_1 + \varepsilon_1. \quad (3.4.6)$$

De la même façon, pour  $\beta$  donnée telle que  $\beta' X^{(2)}$  est variable canonique,  $\alpha$  est telle que :

$$\alpha = \arg \max_a \left| \text{Corr}(a' X^{(1)}, \beta' X^{(2)}) \right| \quad (3.4.7)$$

De la même manière que précédemment,  $\alpha$  est proportionnel au coefficient  $a$  dans le modèle suivant :

$$\beta' X^{(2)} = a' X^{(1)} + \gamma_2 + \varepsilon_2 \quad (3.4.8)$$

Soit  $\alpha_0$  une valeur initiale de  $\alpha$  (qu'on peut par exemple obtenir de l'analyse en composantes principales sur la matrice de données  $(X_1^{(1)}, X_2^{(1)}, \dots, X_n^{(1)})$ ), alors d'après (3.4.6) on obtient une première valeur  $\beta_1$  de  $\beta$  en faisant la régression linéaire de  $\alpha_0' X^{(1)}$  sur  $X^{(2)}$ . On utilisant par la suite (3.4.8) on obtient une deuxième valeur  $\alpha_2$  de  $\alpha$  en faisant la régression de  $\beta_1' X^{(2)}$  sur  $X^{(1)}$ , et ainsi de suite jusqu'à la convergence de cette procédure. On note qu'à chaque étape les coefficients obtenus  $\alpha$  et  $\beta$  sont normalisés et qu'une estimation robuste de la corrélation canonique est obtenue en calculant la corrélation linéaire simple entre les variables canoniques.

#### 3.4.4. Simulation

Cette section est consacrée à la comparaison des méthodes robustes des sections 3.4.1, 3.4.2 et 3.4.3 avec l'analyse canonique classique à l'aide de la simulation faite par Dehon et al. (2000).

On génère  $X_1^{(1)m}, X_2^{(1)m}, \dots, X_n^{(1)m} \in \mathbf{R}^{p_1}$  et  $X_1^{(2)m}, X_2^{(2)m}, \dots, X_n^{(2)m} \in \mathbf{R}^{p_2}$   $m = 1, 2, \dots, M$  à partir d'une distribution multimodale.  $m$  désigne le numéro de simulation, on fera  $M = 200$  simulations. On prend  $n = 50$ ,  $p_1 = 2$  et  $p_2 = 3$ . Les paramètres estimés sont notés  $\hat{\rho}^m$ ,  $\hat{\alpha}^m$  et  $\hat{\beta}^m$  et sont comparés avec les vrais paramètres  $\rho$ ,  $\alpha$  et  $\beta$  qu'on calcule en faisant l'analyse canonique classique sur  $X^{(1)}$  et  $X^{(2)}$  au niveau de la population ( $\Sigma$  est connu). Soit  $\phi$  la transformation de Fisher définie par  $\phi(\rho) = \tanh^{-1}(\rho)$ . On calcule la variance de  $\phi(\hat{\rho}^m)$  donnée par :

$$MSE(\hat{\rho}) = \frac{1}{M} \sum_{m=1}^M (\phi(\hat{\rho}^m) - \phi(\rho))^2. \quad (3.4.9)$$



On calcule aussi l'angle moyen entre  $\hat{\alpha}^m$  et  $\alpha$ , entre  $\hat{\beta}^m$  et  $\beta$  à l'aide des quantités suivantes:

$$MSE(\hat{\alpha}) = \frac{1}{M} \sum_{m=1}^M \cos^{-1} \left( \frac{|\alpha' \hat{\alpha}^m|}{\|\alpha\| \|\hat{\alpha}^m\|} \right) , \quad (3.4.10)$$

$$MSE(\hat{\beta}) = \frac{1}{M} \sum_{m=1}^M \cos^{-1} \left( \frac{|\beta' \hat{\beta}^m|}{\|\beta\| \|\hat{\beta}^m\|} \right) . \quad (3.4.11)$$

Puisqu' on est intéressé au comportement de  $\hat{\rho}^m$ ,  $\hat{\alpha}^m$  et  $\hat{\beta}^m$  quand des observations aberrantes sont présentes parmi les données, on génère dans une seconde étape 10% des données c'est-à-dire cinq observations disons  $X_1^{(1)m}, \dots, X_5^{(1)m}$  à partir de la distribution  $\mathcal{N}(0, 50\mathcal{I}_{p_1})$  et le reste des données à partir de la distribution précédente  $\mathcal{N}(0, \Sigma)$ . Et on calcule les quantités (3.4.9), (3.4.10) et (3.4.11) dans la nouvelle situation qu'on compare à celles calculées lorsqu'il n'y avait pas de données aberrantes. Le tableau 3.4.1 résume les résultats de cette simulation.

TAB. 3.4.1. Résultats de la simulation (obs.naber: pas de présence d'observations aberrantes; obs.aber: présence d'observations aberrantes).

Type d'AC	$MSE(\hat{\rho})$		$MSE(\hat{\alpha})$		$MSE(\hat{\beta})$	
	obs.naber	obs.aber	obs.naber	obs.aber	obs.naber	obs.aber
classique	0,02	1,33	0,14	0,99	0,28	0,92
basé sur MCD	0,06	0,05	0,23	0,22	0,40	0,40
basé sur les signes	0,12	0,25	0,18	0,19	0,32	0,40
basé sur la régression	0,06	0,05	0,20	0,22	0,34	0,37

On voit bien que dans le cas d'absence d'observations aberrantes, l'AC classique est la plus précise. Alors que dans le cas où des observations aberrantes sont présentes, l'AC basé sur MCD et celle basée sur la régression alternée sont les plus précises.

### 3.4.5. Quatrième approche : M-estimateur .

Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire issu d'une loi  $f$  continue qui s'exprime comme:

$$f(X) = |\mathbf{V}|^{-\frac{1}{2}} h \left[ (X - \mu)' \mathbf{V}^{-1} (X - \mu) \right],$$

où  $\mu : p \times 1$   $\mathbf{V} : p \times p$  sont respectivement les paramètres de position et d'échelle de  $X$ ,  $h(|\cdot|) : [0, \infty) \rightarrow [0, \infty)$  une densité dans  $\mathbf{R}^p$  qui n'est connue qu'approximativement ( donc la distribution de  $f$  est inconnu ), et  $|\cdot|$  est la norme Euclidienne.  $f$  est dite une distribution elliptique et on note  $X \sim \mathcal{E}_p(\mu, \mathbf{V})$ . Notons que  $(X - \mu)' \mathbf{V}^{-1} (X - \mu) = d(X, \mu, \mathbf{V})$  est la distance statistique entre  $X$  et  $\mu$  mesurée dans la métrique induite par la matrice définie positive  $\mathbf{V}$ .

Le M-estimateur (Maronna (1976)) du paramètre  $\mu$  ainsi que celui de  $\mathbf{V}$  sont définis comme étant la solution du système d'équations:

$$\text{ave} [u_1(t_i)(X_i - \mu_n)] = 0 \quad (3.4.12)$$

$$\text{ave} \left[ u_2(t_i^2)(X_i - \mu_n)(X_i - \mu_n)' \right] = \mathbf{V}_n, \quad (3.4.13)$$

où  $t_i = [(X_i - \mu_n)' \mathbf{V}_n^{-1} (X_i - \mu_n)]$ .

Le symbole  $\text{ave}$  désigné la moyenne.  $u_1, u_2$  et  $\psi$  (défini par  $\psi_i(s) = s u_i(s)$ ) sont des fonctions qui vérifient les conditions suivantes:

**A** :  $u_1$  et  $u_2$ , qui sont définies pour tout  $s$  positif, sont des fonctions non négatives, non croissantes et sont continues dans  $[0, \infty)$ ;

**B** :  $\psi_1$  et  $\psi_2$  sont bornées ( dans une bande );

**C** :  $\psi_2$  est non décroissante, et est strictement croissante dans l'intervalle où  $\psi_2 \leq K_2$  avec  $K_i = \sup_{s \geq 0} \psi_i(s)$  pour  $i = 1, 2$ ;

**D** : il existe  $s_0$  tel que  $\psi_2(s_0^2) > p$ , et que  $u_1(s) > 0$  pour tout  $s \leq s_0$  ( et par conséquent  $K_2 > p$ );

**E** : il existe  $a_0$  tel que pour tout hyperplan  $\mathcal{H}$   $\dim \mathcal{H} \leq p - 1$

$$P(\mathcal{H}) \leq 1 - \frac{p}{K_2} - a.$$

La condition E est indispensable pour l'existence, l'unicité et la convergence des estimateurs  $\mu_n$  et  $\mathbf{V}_n$ . Pour des échantillons de grande taille ( $n \gg p$ ), la condition E est vérifiée. Cependant si  $n$  est petit,  $K_2$  devrait être plus grand que  $\frac{pn}{n-p}$ . Mais on verra plus tard qu'on perd de la robustesse (la robustesse diminue) avec des valeurs assez grandes de  $K_2$ .

Si la distribution empirique est donnée par :

$$P(X_i) = \frac{1}{n} \quad \text{pour } i = 1, 2, \dots, n,$$

alors (3.4.12) et (3.4.13) seront équivalentes à :

$$\frac{1}{n} \sum_{i=1}^n u_1(t_i)(X_i - \mu_n) = 0 \quad (3.4.15)$$

$$\frac{1}{n} \sum_{i=1}^n u_2(t_i^2)(X_i - \mu_n)(X_i - \mu_n)' = \mathbf{V}_n. \quad (3.4.16)$$

**Théorème 3.4.1.** *Soit  $f$  une distribution dans  $\mathbf{R}^p$  qui satisfait la condition (E) et pour laquelle  $(\mu, \mathbf{V})$  est la solution unique du système d'équations :*

$$E_f [u_1(t)(X - \mu)] = 0 \quad (3.4.17)$$

$$E_f [u_2(t^2)(X - \mu)(X - \mu)'] = \mathbf{V}, \quad (3.4.18)$$

où  $t = [(X - \mu)' \mathbf{V}^{-1} (X - \mu)]^{\frac{1}{2}}$ .

Soient  $X_1, X_2, \dots, X_n$  un échantillon aléatoire issu de la distribution  $f$ , et soient  $P_n$  la distribution empirique correspondant à cet échantillon, et  $(\mu_n, \mathbf{V}_n)$  la solution de (3.4.17) et (3.4.18) quand la mesure est  $P_n$  alors  $\lim_{n \rightarrow \infty} (\mu_n, \mathbf{V}_n) = (\mu, \mathbf{V})$  avec probabilité égale à 1

Pour la preuve de ce théorème voir Maronna (1976).

Le point de rupture  $\delta^*$  du M-estimateur est telle que :

$$\delta^* \leq \min \left( \frac{1}{K_2}, 1 - \frac{p}{K_2} \right)$$

Donc plus  $K_2$  est grand plus  $\delta^*$  est faible plus le M-estimateur perd de sa robustesse. Maronna(1976) a montré que puisque  $K_2 > p$ , alors pour tout M-estimateur

$\delta^* \leq \frac{1}{p+1}$ , ce qui veut dire que plus la dimension augmente moins le M-estimateur est robuste.

### 3.4.6. Cinquième approche : S-estimateur.

Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire issu d'une population  $X: p \times 1$  où  $X \sim \mathcal{E}_p(\mu, \mathbf{V})$  et soit  $t = \left[ (X - \mu)' \mathbf{V}^{-1} (X - \mu) \right]^{\frac{1}{2}} = |z|$  où  $z = \mathbf{V}^{-\frac{1}{2}}(X - \mu)$ ,  $z \sim \mathcal{E}_p(0, \mathcal{I})$ . Soit  $\rho$  une fonction qui vérifie les conditions suivantes :

**C1** :  $\rho$  est symétrique, sa dérivée  $\psi$  est continue et  $\rho(0) = 0$ .

**C2** : il existe une constante  $c_0 > 0$  finie telle que  $\rho$  est strictement croissante dans  $[0, c_0)$  et est constante dans  $[c_0, \infty)$ .

Rousseeuw & Yohai (1984) ont trouvé, en se basant sur la fonction  $\rho$ , un estimateur robuste, qu'ils ont appelé le S-estimateur, pour estimer dans le cadre de la régression un paramètre  $\theta : p \times 1$ .

Le S-estimateur  $(\mu_n, \mathbf{V}_n)$  de  $(\mu, \mathbf{V})$  est défini comme étant la solution optimale du problème suivant :

$$\left\{ \begin{array}{l} \min_{\mu_n, \mathbf{V}_n > 0} |\mathbf{V}_n| \\ \text{Sous contrainte} \\ \frac{1}{n} \sum_{i=1}^n \rho(t_i) = b_0 \end{array} \right.$$

où  $t_i = [(X_i - \mu_n)' \mathbf{V}_n^{-1} (X_i - \mu_n)]^{\frac{1}{2}}$

$\mathbf{V}_n > 0$  signifie que  $\mathbf{V}_n$  est une matrice défini positive.  $b_0$  est une constante dans  $]0, a_0[$  ( $a_0 = \sup \rho$ ). Une fonction  $\rho$  bien particulière est celle de Tukey défini par :

$$\rho(t) = \begin{cases} \frac{t^2}{2} - \frac{t^4}{2c_0^2} + \frac{t^6}{6c_0^4} & \text{si } |t| \leq c_0, \\ \frac{c_0^2}{6} & \text{si } |t| \geq c_0. \end{cases}$$

Si  $b_0$  est choisie de telle sorte  $0 < \frac{b_0}{a_0} \equiv r \leq \frac{n-p}{2n}$ , alors le point de rupture du S-estimateur serait  $\delta_n^* = \frac{[nr]}{n}$ . Le point de rupture  $\delta^*$  atteint sont plus haut niveau quand  $r = \frac{n-p}{2n}$  et sa valeur est  $\lfloor \frac{n-p+1}{2} \rfloor \frac{1}{n}$  (asymptotiquement 50%).

## CONCLUSION

---

Dans ce mémoire, on trouve l'essentiel de ce qui se rapporte à l'analyse canonique, de l'analyse canonique à deux groupes de variables aléatoires ( $X^{(1)} : p_1 \times 1, X^{(2)} : p_2 \times 1$ ) jusqu'à l'application de la robustesse à celle-ci tout en passant par trois types de généralisation à plusieurs groupes de variables aléatoires ( $X^{(1)}, \dots, X^{(k)}$ ), à savoir la généralisation de Carroll (1968), la généralisation de Lazraq et al. (1992) et enfin celle de Gupta (1981). On a vu dans le chapitre 1 que les corrélations canoniques entre  $X^{(1)}$  et  $X^{(2)}$  ne sont autres que les valeurs propres de la matrice  $\mathcal{M} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ . La première corrélation canonique est la plus grande valeur propre de la matrice  $\mathcal{M}$ , la deuxième corrélation canonique est la deuxième plus grande valeur propre de la matrice  $\mathcal{M}$  et ainsi de suite. Les variables canoniques sont les vecteurs propres correspondants aux valeurs propres de la matrice  $\mathcal{M}$ . Dans le deuxième chapitre, on a présenté les trois généralisations citées ci-dessus. On y a vu que la généralisation de Carroll (1968) consiste à maximiser  $\sum w_i r^2(Z, A_i X_i)$ . La solution était de prendre comme variables canoniques les quantités  $A_i = ZX'_i(X_i X'_i)^{-1}$ ,  $i = 1, 2, \dots, k$  où  $X_i$  est la matrice contenant les données sur le vecteur observé  $X^{(i)}$ .  $Z$ , variable compromis, est le vecteur propre de la matrice  $\mathcal{Q} = \sum_{i=1}^k w_i X'_i(X_i X'_i)^{-1} X_i$ . La solution de Lazraq et al. (1992) basée sur la mesure de Lingos et Schonemann (*RLS*) coïncide avec la solution de Carroll (1968). La généralisation de Lazraq et al. (1992) basée sur la mesure d'Escoufier (*RV*) consiste aussi à prendre comme variables canoniques les quantités  $A_i = ZX'_i(X_i X'_i)^{-1}$ ,  $i = 1, 2, \dots, k$  et où  $Z$  est la variable compromis qui maximise  $\sum_{i=1}^k w_i \text{tr}(Z\mathcal{P}Z')$ . La généralisation de Gupta (1981), à l'opposée de la généralisation de Carroll (1968) et celle de Lazraq et al. (1992) qui sont basées sur la maximisation d'une fonction de corrélation, est basée sur la

minimisation de la variance généralisée des variables canoniques tout en supposant que celles-ci sont éuicorrélées. La solution de Gupta est de prendre comme corrélations canoniques  $\rho$  les solutions de  $|\sum_{od} - (k - 1)\rho \sum_d| = 0$  où  $\sum_{od}$  et  $\sum_d$  sont données par (2.8.16). Les variables canoniques sont les vecteurs propres correspondants aux valeurs propres  $(k - 1)\rho$  de la matrice  $\sum_d^{-1} \sum_{od}$ . Le chapitre 3 est consacré à quelques méthodes robustes de l'analyse canonique. On y a exposé les fonctions d'influence des variables et corrélations canoniques. Comme on y a exposé quelques techniques robustes appliquées à l'analyse canonique telles que l'utilisation du MCD, l'utilisation des signes des données ou l'utilisation de la régression alternée robuste. Et dans la dernière partie de ce travail on a donné les programmes codés soit en SAS soit en Splus qui ont servi pour le calcul des solutions des deux premiers chapitres.

# PROGRAMMES

---

## 3.5. PROGRAMME SAS POUR ANALYSER LES DONNÉES CONSCRITS DU CHAPITRE 1

Programme 1 :

=====

```
options nodate linesize = 80 formdlim ='=';
filename donnees "exp.dat"; /* SAS lira les données dans le fichier exp.dat */

data expl;

    infile 'exp.dat ' ;

    input IND TAI PDS PTU TAP PDP PTP TAM PDM PTM;

run;

proc cancorr all;

    title1 "Analyse canonique sur les donnees CONSCRITS ";

    var TAI TAP TAM;

    with PDS PDP PDM;

run;
```

## 3.6. PROGRAMME SPLUS POUR CALCULER LA SOLUTION DE CAR- ROLL

#  $X_1$ ,  $X_2$  et  $X_3$  sont les trois matrices,  $m_i \times n$ , observées des vecteurs  $X^{(1)}$ ,  
#  $X^{(2)}$  et  $X^{(3)}$ . Ces vecteurs ont été observés sur  $n$  individus. Ces matrices seront

‡ centrées à l'aide de la fonction `Scale` toute on utilisant l'option `scale = F`.

Les

‡ poids de chacun de ces vecteurs seront considérés égaux à 1 ( $w_1 = w_2 = w_3 = 1$ ).

Programme 2 :

=====

$X1 = t(scale(t(X1), center = T, scale = F))$

$X2 = t(scale(t(X2), center = T, scale = F))$

$X3 = t(scale(t(X3), center = T, scale = F))$

$X1.p = t(X1)$

$s1 = solve(X1 \% * \%X1.p)$

$P1 = X1.p \% * \%s1 \% * \%X1$

$X2.p = t(X2)$

$s2 = solve(X2 \% * \%X2.p)$

$P2 = X2.p \% * \%s2 \% * \%X2$

$X3.p = t(X3)$

$s3 = solve(X3 \% * \%X3.p)$

$P3 = X3.p \% * \%s3 \% * \%X3$

$Q = w1 * P1 + w2 * P2 + w3 * P3$

$val.p = eigen.default(Q, only.value = F)$value$



‡ Seules les cinq premières valeurs propres de  $Q$  sont non nulles. Donc on ne  
 ‡ retiendra que les vecteurs propres correspondants à ces valeurs propres.

$$val.p = val.p[1 : 5]$$

$$Z = eigen.default(Q, only.value = F)$vector$$

$$Z = Z[, 1 : 5]$$

‡ la  $i$ ème colonne de  $Z$  est le vecteur propre correspondant à la  $i$ ème valeur  
 ‡ propre de  $Q$ .

‡ Pour centrer  $Z$  on utilise:  $sacle(Z, center = T, scale = F)$ .

$$Z = scale(Z, center = T, scale = F)$$

‡  $reduction$  est une fonction qui transforme un vecteur  $v$  en un autre  $u$  telle  
 ‡ que  $t(u) \% * \%u = 1$ , on a besoin de cette fonction pour transformer  $Z$  en un  
 ‡ autre  $Z$  telle que  $ZZ' = 1$

$$Z = apply(Z, 2, reduction)$$

‡ les combinaisons linéaires  $A_i$  seront alors:

$$A1 = t(Z[, 1]) \% * \%X1.p \% * \%s1$$

$$A2 = t(Z[, 1]) \% * \%X2.p \% * \%s2$$

$$A3 = t(Z[, 1]) \% * \%X3.p \% * \%s3$$

‡ Si on veut chercher les combinaisons du deuxième ordre, on à qu'à remplacer  
 ‡  $Z[, 1]$  dans l'expression de  $A_i$  par  $Z[, 2]$  et ainsi de suite .

### 3.7. PROGRAMME SPLUS POUR LA DÉTERMINATION DE LA SOLUTION DE LAZRAQ, CLÉROUX ET KIERS

L'objectif de ce programme est de déterminer la matrice  $Z : n \times r$  qui permet de maximiser  $\sum_{k=1}^K \text{tr}(Z' S_k Z)^2$ . Ce programme est une fonction, appelée  $rvfun$ , codé en Splus dont les arguments sont les matrices  $s1, s2, s3$  qui dans le cas de

l'exemple des mouches sont les matrices de projections  $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$  puisque les poids  $w_1, w_2$  et  $w_3$  sont tous égales à 1. On commence par donner une fonction Splus, appelé tr, qui calcule la trace d'une matrice puisque la fonction rvfun lui fera appelle.

Programme 3 :

=====

# fonction qui calcule la trace d'une matrice :

#=====

```
tr=function(mat)
{
  sum(diag(as.matrix(mat)))
}
```

#=====

#La fonction qui donne la matrice Z:

#=====

```
rvfun=function(s1,s2,s3,r,conv)
{
  # S est une matrice qui regroupe s1, s2 et s3
  # r est la dimensionalite de la solution
  # conv est le critere de convergence
  # strat=0 default; start=1 valeur actuelle
  # Les poids des vecteurs X(1),X(2) et X(3)
  S=cbind(s1,s2,s3)
  n=dim(S)[1]
  mn=dim(S)[2]
  m=mn/n
  #initialisation de X
  U=matrix(0,nrow=n,ncol=n)
  for(k in 1:m)
```

```

        {
            c1=(k-1)*n+1
            c2=k*n
            sk=S[,c1:c2]
            U=U+sk
        }
        X=svd(U)$u[,1:r]
        tX=t(X)
sumU2=sum(U^2)
f=1000*sum(U^2)
fold=f+2*conv*f
iter=0
while(fold-f>conv*f || iter<3)
{
    fold=f
    iter=iter+1
    # actualisation de X
    BB=matrix(0,nrow=n, ncol=r)
    for(k in 1:m)
    {
        c1=(k-1)*n+1
        c2=k*n
        sk=S[,c1:c2]
        #tX=t(X)
        BB=BB+sk%*%X%*%tX%*%sk%*%X
    }
    u=svd(BB)$u
    v=svd(BB)$v
    tv=t(v)

```

```
X=u%*%tv
tX=t(X)
f=0
for(k in 1:m)
{
  c1=(k-1)*n+1
  c2=k*n
  sk=S[,c1:c2]
  mat1=tX%*%sk
  mat2=mat1%*%X
  f=f+sum(sk^2)-sum(mat2^2)
}
f}
tX=t(X)
tra=0
for(k in 1:m)
{
  c1=(k-1)*n+1
  c2=k*n
  sk=S[,c1:c2]
  mat3=tX%*%sk
  mat4=mat3%*%X
  mat5=mat4%*%mat4
  tra=tra+tr(mat5)
}
list(X)
}
```

Le programme suivant permet de calculer les matrices  $A_i$  dans le cas de la méthode  $RV - GCA$ . La matrice  $Z$  est  $r \times n$ . Elle est la transposée de la matrice  $Z$  calculé par la fonction `rvfun` du programme passé.

Programme 4 :

```
=====
tX1=t(X1)
tX2=t(X2)
tX3=t(X3)
mat1=solve(X1%*%tX1)
mat2=solve(X2%*%tX2)
mat3=solve(X3%*%tX3)
Z=t(Z)
A1=Z%*%tX1%*%mat1
A2=Z%*%tX2%*%mat2
A3=Z%*%tX3%*%mat3
```

### 3.8. PROGRAMME SPLUS POUR CALCULER LA SOLUTION DE GUPTA.

Le programme 5 permet de calculer la solution de Gupta.

```
# Programme 5:
#=====
# X1 est 2x35
# X2 est 3x35
# X3 est 2x35

# Construction de la matrice sigma_od.
# =====

zero11=matrix(0, nrow=2, ncol=2)
zero22=matrix(0, nrow=3, ncol=3)
```

```
zero33=matrix(0, nrow=2, ncol=2)

tX1=t(X1)
tX2=t(X2)
tX3=t(X3)
S11=var(tX1)
S12=var(tX1,tX2)
S13=var(tX1,tX3)
S21=t(S12)
S22=var(tX2)
S23=var(tX2,tX3)
S31=t(S13)
S32=t(S23)
S33=var(tX3)

sigma.od1=cbind(zero11, S12, S13)
sigma.od2=cbind(S21, zero22, S23)
sigma.od3=cbind(S31, S32, zero33)
sigma.od=rbind(sigma.od1, sigma.od2,sigma.od3)

# Construction de la matrice sigma_d
# =====

zero12=matrix(0, nrow=2, ncol=3)
zero13=zero11
zero21=t(zero12)
zero23=t(zero12)
zero31=zero11
zero32=zero12
```

```

sigma.d1=cbind(S11, zero12, zero13)
sigma.d2=cbind(zero21, S22, zero23)
sigma.d3=cbind(zero31, zero32,S33)
sigma.d=rbind(sigma.d1, sigma.d2, sigma.d3)

# calcul du produit sigma.d^(-1)*sigma_od
# =====

sol.d=solve(sigma.d)
prod.dod=sol.d%*%sigma.od

# calcul de valeurs et vecteurs propres de la matrice
# sigma.d^(-1)*sigma_od
# =====

val.dod=as.matrix(eigen.default(prod.dod,only.value=F)$value)
vec.dod=eigen.default(prod.dod,only.value=F)$vector

# Construction de la fonction qui calcule ke determinant de sigma_rho
# =====

det.sig=function(rho, k)
{
  ((1 - rho)^(k-1)) * (rho * (k - 1) + 1)
}

# calcul du determinant de sigma_rho pour chacune des valeurs de rho
# et determination du rho qui donne le minimum de sigma_rho

```

```
# =====
```

```
det.rho=apply(val.dod, 1, det.sig, 3)
```

```
rho.min=min(det.rho)
```



## ANNEXE

---

Les données CONSCRITS qui ont servis pour l'exemple 1.5 :

<i>IND</i>	<i>TAI</i>	<i>PDS</i>	<i>PTU</i>	<i>TAP</i>	<i>PDP</i>	<i>PTP</i>	<i>TAM</i>	<i>PDM</i>	<i>PTM</i>
1	184	73	42	180	85	45	164	55	37
2	175	62	43	170	70	41	159	54	39
3	189	76	45	171	65	42	170	68	41
4	173	73	44	167	67	39	156	63	37
5	169	62	41	166	80	44	161	58	39
6	175	68	42	170	64	41	165	63	39
7	164	64	40	163	58	39	162	54	37
8	190	85	43	180	82	44	168	62	37
9	174	55	41	171	71	44	154	50	38
10	184	74	43	182	85	44	169	68	40
11	168	56	41	167	70	41	155	45	36
12	179	64	41	171	62	40	164	67	37
13	174	65	40	155	57	39	156	60	38
14	163	58	41	163	65	41	160	60	38
15	172	53	41	160	65	39	153	53	36
16	176	61	42	174	72	42	154	53	36
17	190	85	45	174	65	41	163	62	39
18	172	56	41	169	70	42	150	43	36
19	178	68	42	180	70	42	157	40	36
20	173	65	42	182	78	44	166	58	37
21	175	67	42	170	75	43	163	50	37

22	168	64	41	170	78	41	162	58	38
23	180	66	44	170	78	42	165	65	39
24	175	65	40	162	64	40	163	60	38
25	184	75	43	165	65	41	162	55	38
26	181	75	43	173	76	43	159	49	38
27	169	60	40	161	64	40	155	55	37
28	174	65	43	175	80	44	166	72	38
29	175	66	44	172	70	42	160	55	39
30	175	60	42	176	80	44	160	55	37
31	180	60	41	176	80	41	158	45	36
32	172	58	41	168	66	41	164	50	38
33	179	68	43	177	75	42	175	70	38
34	183	90	44	190	110	45	160	60	37
35	170	62	40	165	74	40	165	55	38
36	178	75	41	171	73	41	154	50	37
37	168	50	40	164	65	40	158	51	36
38	188	70	44	166	65	40	167	67	38
39	177	68	43	170	85	41	163	56	40
40	165	55	40	160	75	39	150	50	37
41	172	55	40	170	70	41	160	55	38
42	173	56	42	172	65	42	165	49	39
43	176	66	43	178	80	43	158	65	38
44	177	70	42	168	85	41	161	60	38
45	180	62	42	178	90	46	158	70	39
46	170	52	42	168	64	42	160	50	37
47	184	70	44	178	76	42	168	60	38
48	175	67	41	173	75	42	158	56	36
49	181	67	40	175	78	41	161	55	37
50	175	61	42	178	69	41	157	65	39

51	162	63	40	165	62	42	160	57	38
52	176	60	40	172	62	41	156	52	37
53	187	68	45	170	70	42	161	62	39
54	180	69	43	170	70	42	165	55	38
55	173	64	42	170	80	42	160	55	36
56	172	68	41	164	67	40	155	50	37
57	171	60	41	172	75	42	156	54	38
58	170	56	41	169	68	42	162	53	36
59	172	70	44	160	72	42	164	64	40
60	180	65	41	178	72	41	167	55	38
61	172	70	40	169	75	43	150	50	37
62	178	63	42	168	68	41	162	50	38
63	173	62	40	169	80	42	159	50	36

## Bibliographie

---

- ANDERSON, T. (1984). *An introduction to multivariate statistical analysis*. John Wiley New York., 2nd ed.
- BARTLETT, M. (1939). The standar errors of discriminant function coefficients. *Journal of the Royal Statistic Soc .* **6**, 169–173.
- BUTLER, R., DAVIES, P. & JHUN, M. (1993). Asymptotics for the minimum covariance determinant estimator. *The Annals of Statistics* **21**, 1385–1400.
- CARROLL, D. (1968). Generalization of canonical correlation analysis to tree or more sets of variables. *Reprinted from the Proceedings, 76th Annual Convention, APA* , 227–228.
- CARROLL, J. & CHANG, J. (1970). Analysis of individual differences in multi-dimensional scaling via an n-way generalisation of ecart-young decomposition. *Psychometrika* , 283–319.
- CARROLL, J. & CHANG, J. (1972). Idioscal: A generalisation of indiscal allowing idiosyncratic reference systems as well as an analytic approximation of indiscal. *Paper presented at the Spring Meeting of the Psychometric Society, Princeton, NJ, Martch 30-31* , 227–228.
- DEHON, C., FILZMOSER, P. & CROUX, C. (2000). Robust methods for canonical correlation analysis. *Data Analysis, Classification, and Related Methods* , 322–326.
- FOUCART, T. (1984). *Analyse factorielle de tableau multiples*. Masson.
- GREEN, B. (1969). Best linear composites with a specified structure. *Psychometrika* **34**, 301–318.

- GUPTA, S. (1981). On the problems of construction and statistical inference associated with a generalisation of canonical variables. *Tech.Rept.under ONR contract N))14-75-C-0442,Dept of Stat, Stanford Univ .*
- HOTELLING, H. (1935). The most predictable criterion. *Journal of Educational Psychology* **26**, 139–142.
- HOTELLING, H. (1936). Relation between two sets of variates. *Biometrika* **28**, 321–377.
- JOHNSON, R. & WICHERN, D. (1982). *Applied Multivariate Statistical; Analysis*. Prentice-Hall Inc., New Jersey.
- KIERS, H., CLÉROUX, R. & TEN BERGE, J. (1994). Generalized canonical analysis based on optimizing matrix correlations and a relation with idioscal. *Computational Statistics and Data Analysis* **18**, 331–340.
- LAZRAQ, A. & CLÉROUX, R. (1988A). Étude comparative de différentes mesures de liaison entre deux vecteurs aléatoires. *Stat. et anal. des données* **13**, 15–38.
- LAZRAQ, A. & CLÉROUX, R. (1988B). Un algorithme pas á pas de selection de variables en régression lineaire multivariée. *Stat. et anal. des données* **13**, 39–58.
- LAZRAQ, A., CLÉROUX, R. & KIERS, H. (1992). Mesure de liaison vectorielle et généralisation de l'analyse canonique. *Revue statistique appliquée* **1**, 23–35.
- MARDIA, K., KENT, J. & BIBBY, J.M (1979). *Multivariate Analysis*. Z.W. Birnhaum and E.Lukacs.
- MARONNA, R, A. (1976). Robust m-estimators of multivariate location and scatter. *The Annals of Statistics* **4**, 51–67.
- MCKEON, J. (1965). Canonical analysis : some relations between canonical correlations factor analysis, discriminant function analysis and scaking theory. *Psychometric Monograph* .
- RAJA, R. (1969). Partial canonical correlation. *Trabajo de Estadistica y de investigation operativa* **20**, 211–219.

- ROBERT, P. & ESCOUFIER, Y. (1976). A unifying tool for linear multivariate statistical methods: the rv coefficient. *Applied Stat.* **25**, 257–265.
- ROMANAZZI, M. (1989). Derivatives of eigenvalue and eigenvectors in the generalised eigenproblem dependent on parameters. *Technical Repport Venice, Italy University of Venice Ca Foscari Statistical laboratory* .
- ROMANAZZI, M. (1992). Influence in canonical correlation analysis. *Psychometrika* **57**, 237–259.
- ROUSSEEUW, P. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Application, Vol B, eds. W.Grossmann,G* , 283–297.
- ROUSSEEUW, P. & VAN DRIESSEN, V. (1999). A fast algorithm for the covariance determinant estimator. *Technometrics* **V41**, 212–223.
- ROUSSEEUW, P. & YOHAI, V. (1984). Robust regression by means of s-estimators. *Lecture Notes in Statistics* , 256–272.
- ROY, S. (1957). *Some aspects of multivariates analysis*. John Wiley, New York.
- VISURI, S., KOIVVNEN, V. & OJA, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planing and Inference* **91**, 557–575.
- WOLD, H. (1966). Nonlinear estimation by iterative least squares procedures. *In: F.N.Davis(Ed): A Festschrift for J.Neyman Wiley and Sons, New York* , 411–444.