



**HAL**  
open science

# L'analyse multidimensionnelle des données de dissimilarité

Gérard Drouet d'Aubigny

► **To cite this version:**

Gérard Drouet d'Aubigny. L'analyse multidimensionnelle des données de dissimilarité. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1989. tel-00332393

**HAL Id: tel-00332393**

**<https://theses.hal.science/tel-00332393>**

Submitted on 20 Oct 2008

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

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

# THESE

*présentée à*

**l'Université Joseph Fourier - Grenoble I**

*pour obtenir le grade de*  
**DOCTEUR ES SCIENCES**  
**"Mathématiques"**

*par*

**Gérard DROUET d'AUBIGNY**

---

## **L'ANALYSE MULTIDIMENSIONNELLE DES DONNÉES DE DISSIMILARITÉ**

---

**Thèse soutenue le 27 Janvier 1989 devant la Commission d'Examen**

**B. VAN CUTSEM**      **Président**

**J.R. BARRA**  
**F. CRITCHLEY**  
**Y. ESCOUFIER**      **Examineurs**  
**W. HEISER**  
**G. LE CALVE**  
**G. ROMIER**



# UNIVERSITE Joseph FOURIER (GRENOBLE I)

Président de l'Université :  
M. PAYAN Jean Jacques

Année Universitaire 1987 - 1988

## MEMBRES DU CORPS ENSEIGNANT DE SCIENCES ET DE GEOGRAPHIE

### PROFESSEURS DE 1ère Classe

|                         |                                       |
|-------------------------|---------------------------------------|
| ARNAUD Paul             | Chimie Organique                      |
| ARVIEU ROBERT           | Physique Nucléaire I.S.N.             |
| AUBERT Guy              | Physique C.N.R.S                      |
| AURIAULT Jean-Louis     | Mécanique                             |
| AYANT Yves              | Physique Approfondie                  |
| BARBIER Marie-Jeanne    | Electrochimie                         |
| BARJON Robert           | Physique Nucléaire ISN                |
| BARNOUD Fernand         | Biochimie Macromoléculaire Végétale   |
| BARRA Jean-René         | Statistiques-Mathématiques Appliquées |
| BECKER Pierre           | Physique                              |
| BEGUIN Claude           | Chimie Organique                      |
| BELORISKY Elie          | Physique                              |
| BENZAKEN Claude         | Mathématiques Pures                   |
| BERARD Pierre           | Mathématiques Pures                   |
| BERNARD Alain           | Mathématiques Pures                   |
| BERTRANDIAS Françoise   | Mathématiques Pures                   |
| BERTRANDIAS Jean-Paul   | Mathématiques Pures                   |
| BILLET Jean             | Géographie                            |
| BOELHER Jean-Paul       | Mécanique                             |
| BONNIER Jane Marie      | Chimie Générale                       |
| BOUCHEZ Robert          | Physique Nucléaire ISN                |
| BRAVARD Yves            | Géographie                            |
| CARLIER Georges         | Biologie Végétale                     |
| CAUQUIS Georges         | Chimie Organique                      |
| CHARDON Michel          | Géographie                            |
| CHIBON Pierre           | Biologie Animale                      |
| COHEN ADDAD Jean-Pierre | Physique                              |
| COLIN DE VERDIERE Yves  | Mathématiques Pures                   |
| CYROT Michel            | Physique du Solide                    |
| DEBELMAS Jacques        | Géologie Générale                     |
| DEGRANGE Charles        | Zoologie                              |
| DEMAILLY Jean-Pierre    | Mathématiques Pures                   |
| DENEUVILLE Alain        | Physique                              |
| DEPORTES Charles        | Chimie Minérale                       |
| DOLIQUE Jean-Michel     | Physique des Plasmas                  |
| DOUCE Roland            | Physiologie Végétale                  |
| DUCROS Pierre           | Cristallographie                      |
| FONTAINE Jean-Marc      | Mathématiques Pures                   |
| GAGNAIRE Didier         | Chimie Physique                       |
| GERMAIN Jean-Pierre     | Mécanique,                            |
| GIRAUD Pierre           | Géologie                              |
| HICTER Pierre           | Chimie                                |
| IDELMAN Simon           | Physiologie Animale                   |
| JANIN Bernard           | Géographie                            |
| JOLY Jean-René          | Mathématiques Pures                   |
| KAHANE André, détaché   | Physique                              |
| KAHANE Josette          | Physique                              |
| KRAKOWIAK Sacha         | Mathématiques Appliquées              |

LAJZEROWICZ Jeanine  
 LAJZEROWICZ Joseph  
 LAURENT Pierre-Jean  
 LEBRETON Alain  
 DE LEIRIS Joël  
 LHOMME Jean  
 LLIBOUTRY Louis  
 LOISEAUX Jean-Marie  
 LUNA Domingo  
 MACHE Régis  
 MASCLE Georges  
 MAYNARD Roger  
 OMONT Alain  
 OZENDA Paul  
 PAYAN Jean-Jacques  
 PEBAY-PEYROULA Jean-Claude  
 PERRIER Guy  
 PIERRARD Jean-Marie  
 PIERRE Jean-Louis  
 RENARD Michel  
 RINAUDO Marguerite  
 ROSSI André  
 SAXOD Raymond  
 SENGEL Philippe  
 SERGERAERT Francis  
 SOUCHIER Bernard  
 SOUTIF Michel  
 STUTZ Pierre  
 TRILLING Laurent  
 VALENTIN Jacques  
 VAN CUTSEM Bernard  
 VIALON Pierre

Physique  
 Physique  
 Mathématiques Appliquées  
 Mathématiques Appliquées  
 Biologie  
 Chimie  
 Géophysique  
 Sciences Nucléaires I.S.N.  
 Mathématiques Pures  
 Physiologie Végétale  
 Géologie  
 Physique du Solide  
 Astrophysique  
 Botanique (Biologie Végétale)  
 Mathématiques Pures  
 Physique  
 Géophysique  
 Mécanique  
 Chimie Organique  
 Thermodynamique  
 Chimie CERMAV  
 Biologie  
 Biologie Animale  
 Biologie Animale  
 Mathématiques Pures  
 Biologie  
 Physique  
 Mécanique  
 Mathématiques Appliquées  
 Physique Nucléaire I.S.N.  
 Mathématiques Appliquées  
 Géologie

#### PROFESSEURS de 2<sup>ème</sup> Classe

ADIBA Michel  
 ANTOINE Pierre  
 ARMAND Gilbert  
 BARET Paul  
 BLANCHI J.Pierre  
 BLUM Jacques  
 BOITET Christian  
 BORNAREL Jean  
 BRUANDET J.François  
 BRUGAL Gérard  
 BRUN Gilbert  
 CASTAING Bernard  
 CERFF Rudiger  
 CHIARAMELLA Yves  
 COURT Jean  
 DUFRESNOY Alain  
 GASPARD François  
 GAUTRON René  
 GENIES Eugène  
 GIDON Maurice  
 GIGNOUX Claude  
 GILLARD Roland  
 GIORNI Alain  
 GONZALEZ SPRINBERG Gérardo  
 GUIGO Maryse  
 GUMUCHAIN Hervé  
 GUITTON Jacques

Mathématiques Pures  
 Géologie  
 Géographie  
 Chimie  
 STAPS  
 Mathématiques Appliquées  
 Mathématiques Appliquées  
 Physique  
 Physique  
 Biologie  
 Biologie  
 Physique  
 Biologie  
 Mathématiques Appliquées  
 Chimie  
 Mathématiques Pures  
 Physique  
 Chimie  
 Chimie  
 Géologie  
 Sciences Nucléaires  
 Mathématiques Pures  
 Sciences Nucléaires  
 Mathématiques Pures  
 Géographie  
 Géographie  
 Chimie

**HACQUES Gérard**  
**HERBIN Jacky**  
**HERAULT Jeanny**  
**JARDON Pierre**  
**JOSELEAU Jean-Paul**  
**KERCKHOVE Claude**  
**LONGEQUEUE Nicole**  
**LUCAS Robert**  
**MANDARON Paul**  
**MARTINEZ Francis**  
**NEMOZ Alain**  
**OUDET Bruno**  
**PECHER Arnaud**  
**PELMONT Jean**  
**PERRIN Claude**  
**PFISTER Jean-Claude**  
**PIBOULE Michel**  
**RAYNAUD Hervé**  
**RICHARD Jean-Marc**  
**RIEDTMANN Christine**  
**ROBERT Gilles**  
**ROBERT Jean-Bernard**  
**SARROT-REYNAULD Jean**  
**SAYETAT Françoise**  
**SERVE Denis**  
**STOECKEL Frédéric**  
**SCHOLL Pierre-Claude**  
**SUBRA Robert**  
**VALLADE Marcel**  
**VIDAL Michel**  
**VIVIAN Robert**  
**VOTTERO Philippe**

**Mathématiques Appliquées**  
**Géographie**  
**Physique**  
**Chimie**  
**Biochimie**  
**Géologie**  
**Sciences Nucléaires I.S.N.**  
**Physique**  
**Biologie**  
**Mathématiques Appliquées**  
**Thermodynamique CNRS - CRTBT**  
**Mathématiques Appliquées**  
**Géologie**  
**Biochimie**  
**Sciences Nucléaires I.S.N.**  
**Physique du Solide**  
**Géologie**  
**Mathématiques Appliquées**  
**Physique**  
**Mathématiques Pures**  
**Mathématiques Pures**  
**Chimie Physique**  
**Géologie**  
**Physique**  
**Chimie**  
**Physique**  
**Mathématiques Appliquées**  
**Chimie**  
**Physique**  
**Chimie Organique**  
**Géographie**  
**Chimie**

## **MEMBRES DU CORPS ENSEIGNANT DE L' IUT 1**

### **PROFESSEURS de 1ère Classe**

**BUISSON Roger**  
**DODU Jacques**  
**NEGRE Robert**  
**NOUGARET Marcel**  
**PERARD Jacques**

**Physique IUT 1**  
**Mécanique Appliquée IUT 1**  
**Génie Civil IUT 1**  
**Automatique IUT 1**  
**EEA. IUT 1**

### **PROFESSEURS de 2ème classe**

**BOUTHINON Michel**  
**CHAMBON René**  
**CHEHIKIAN Alain**  
**CHENAVAS Jean**  
**CHOUTEAU Gérard**  
**CONTE René**  
**GOSSE Jean-Pierre**  
**GROS Yves**  
**KUHN Gérard, (Détaché)**  
**MAZUER Jean**  
**MICHOULIER Jean**  
**MONLLOR Christian**  
**PEFFEN René**  
**PERRAUD Robert**  
**PIERRE Gérard**  
**TERRIEZ Jean-Michel**  
**TOUZAIN Philippe**  
**VINCENDON Marc**

**EEA. IUT 1**  
**Génie Mécanique IUT 1**  
**EEA. IUT 1**  
**Physique IUT 1**  
**Physique IUT 1**  
**Physique IUT 1**  
**EEA.IUT 1**  
**Physique IUT 1**  
**Physique IUT 1**  
**Physique IUT 1**  
**EEA.IUT 1**  
**Métallurgie IUT 1**  
**Chimie IUT 1**  
**Chimie IUT 1**  
**Génie Mécanique IUT 1**  
**Chimie IUT 1**  
**Chimie IUT 1**



**PROFESSEURS 2ème CLASSE**

|                           |   |                  |
|---------------------------|---|------------------|
| BACHELOT Yvan             | Endocrinologie                              | C.H.R.G.         |
| BARGE Michel              | Neurochirurgie                              | C.H.R.G.         |
| BENABID Alim Louis        | Biophysique                                 | Faculté La Merci |
| BENSA Jean-Claude         | Immunologie                                 | Hopital Sud      |
| BERNARD Pierre            | Gynécologie-Obstétrique                     | C.H.R.G.         |
| BESSARD Germain           | Pharmacologie                               | ABIDJAN          |
| BOLLA Michel              | Radiothérapie                               | C.H.R.G.         |
| BOST Michel               | Pédiatrie                                   | C.H.R.G.         |
| BOUCHARLAT Jacques        | Psychiatrie Adultes                         | Hopital Sud      |
| BRAMBILLA Christian       | Pneumologie                                 | C.H.R.G.         |
| CHIROSSSEL Jean-Paul      | Anatomie-Neurochirurgie                     | C.H.R.G.         |
| COMET Michel              | Biophysique                                 | Faculté La Merci |
| CONTAMIN Charles          | Chirurgie Thoracique et<br>Cardiovasculaire | C.H.R.G.         |
| CORDONNIER Daniel         | Néphrologie                                 | C.H.R.G.         |
| COULOMB Max               | Radiologie                                  | C.H.R.G.         |
| CROUZET Guy               | Radiologie                                  | C.H.R.G.         |
| DEBRU Jean-Luc            | Médecine Interne et Toxicologie             | C.H.R.G.         |
| DEMONGEOT Jacques         | Biostatistiques et Informatique<br>Médicale | Faculté La Merci |
| DUPRE Alain               | Chirurgie Générale                          | C.H.R.G.         |
| DYON Jean-François        | Chirurgie Infantile                         | C.H.R.G.         |
| ETERRADOSSI Jacqueline    | Physiologie                                 | Faculté La Merci |
| FAURE Claude              | Anatomie et Organogénèse                    | C.H.R.G.         |
| FAURE Gilbert             | Urologie                                    | C.H.R.G.         |
| FOURNET Jacques           | Hépatogastro-Entérologie                    | C.H.R.G.         |
| FRANCO Alain              | Médecine Interne                            | C.H.R.G.         |
| GIRARDET Pierre           | Anesthésiologie                             | C.H.R.G.         |
| GUIDICELLI Henri          | Chirurgie Générale et Vasculaire            | C.H.R.G.         |
| GUIGNIER Michel           | Thérapeutique et Réanimation<br>Médicale    | C.H.R.G.         |
| HADJIAN Arthur            | Biochimie                                   | Faculté La Merci |
| HALIMI Serge              | Endocrinologie et Maladies<br>Métaboliques  | C.H.R.G.         |
| HOSTEIN Jean              | Hépatogastro-Entérologie                    | C.H.R.G.         |
| HUGONOT Robert            | Médecine Interne                            | C.H.R.G.         |
| JALBERT Pierre            | Histologie-Cytogénétique                    | C.H.R.G.         |
| JUNIEN-LAVILLAULOY Claude | O.R.L.                                      | C.H.R.G.         |
| KOLODIE Lucien            | Hématologie Biologique                      | C.H.R.G.         |
| LETOUBLON Christian       | Chirurgie Générale                          | C.H.R.G.         |
| MACHECOURT Jacques        | Cardiologie et Maladies<br>Vasculaires      | C.H.R.G.         |
| MAGNIN Robert             | Hygiène                                     | C.H.R.G.         |
| MASSOT Christian          | Médecine Interne                            | C.H.R.G.         |
| MOUILLON Michel           | Ophtalmologie                               | C.H.R.G.         |
| PELLAT Jacques            | Neurologie                                  | C.H.R.G.         |
| PHELIP Xavier             | Rhumatologie                                | C.H.R.G.         |
| RACINET Claude            | Gynécologie-Obstétrique                     | Hopital Sud      |
| RAMBAUD Pierre            | Pédiatrie                                   | C.H.R.G.         |
| RAPHAEL Bernard           | Stomatologie                                | C.H.R.G.         |
| SCHAERER René             | Cancérologie                                | C.H.R.G.         |
| SEIGNEURIN Jean-Marie     | Bactériologie-Virologie                     | Faculté La Merci |
| SELE Bernard              | Cytogénétique                               | Faculté La Merci |
| SOTTO Jean-Jacques        | Hématologie                                 | C.H.R.G.         |
| STOEBNER Pierre           | Anatomie Pathologique                       | C.H.R.G.         |
| VROUSOS Constantin        | Radiothérapie                               | C.H.R.G.         |



*à Catherine,*



*Qu'il me soit permis de remercier tout d'abord Bernard VAN CUTSEM, Professeur à l'Université J. Fourier de Grenoble qui me fait l'honneur de présider ce Jury. Plus encore que son intérêt pour un sujet proche de ses préoccupations de recherche, j'y vois une marque d'amitié à laquelle je suis très sensible. Il est vrai que gentillesse et ouverture d'esprit ne sont pas ses moindres qualités.*

*J'attache beaucoup d'importance à la présence dans ce jury de Jean René BARRA, Professeur à l'Université J. Fourier de Grenoble. Fort des acquis méthodologiques de ses enseignements de Statistique Inférentielle, j'ai pu entreprendre des recherches en Analyse des Données sans sentiment de rupture et bien au contraire avec la certitude d'une complémentarité et d'une synergie. Depuis, ma conviction sur les dangers d'une pratique de l'Analyse des Données sans une base solide et rigoureuse en Statistique Inférentielle n'a cessé d'être renforcée par l'expérience.*

*La présence de Franck. CRITCHLEY, Professeur à l'Université de Warwick, n'étonnera pas les spécialistes, compte tenu de la valeur de ses contributions dans le domaine de recherche abordé dans cette thèse. Derrière le rire sonore et jovial se cache un esprit fin et chaleureux, auquel j'exprime ma gratitude pour l'intérêt qu'il me témoigne. Il y a cependant plus, et je voudrais saluer à travers Franck, la formidable capacité du système éducatif anglais à détecter et promouvoir de grands talents de statisticiens.*

*Yves ESCOUFIER, Professeur à l'USTL, me témoigne à nouveau son intérêt et son amitié, en acceptant de participer à ce jury. Je lui en suis d'autant plus reconnaissant que je sais la charge et les responsabilités prises par cet infatigable promoteur et propagandiste de la statistique française pour le bien de notre communauté.*

*Mes remerciements chaleureux vont à Jan de LEEUW, désormais professeur à UCLA qui a bien voulu accepter la tâche ingrate de rapporteur, alors qu'il savait ne pouvoir être présent au jury et à Willem HEISER, Professeur à l'Université de Leiden qui m'honore de sa présence. Comment séparer ces animateurs du "Department of Data Theory" de l'Université de Leiden, qui restera le laboratoire le plus créatif et le plus productif de la décennie dans le domaine. Leurs contributions séparées ou conjointes à la méthodologie du Codage Multidimensionnel furent pour moi un aiguillon continu au cours des années. Que chacun d'eux trouve ici l'expression de ma sincère gratitude.*

*Georges. LE CALVE, Professeur à l'Université de Rennes II m'a fait lui aussi l'honneur et l'amitié de rapporter sur mon travail, malgré la somme d'activité qu'il lui est actuellement imposée. Le dynamisme avec lequel il anime le groupe "Similarités" montre cependant l'étendue de ses ressources. Je lui exprime toute ma gratitude pour son aide. Nostalgie bretonne oblige, j'aime à penser que peut-être me suis-je une nouvelle fois enivré, par manuscrit interposé, des parfums de Groix face à "la mer calme où des mouettes éparses flottaient comme des corolles blanches".*

*Il me plaît de clore ces remerciements en soulignant le rôle essentiel de Guy ROMIER, Professeur à l'Université des Sciences Sociales de Grenoble, qui m'initia à l'Analyse des Données, sut attirer mon attention sur les spécificités et l'intérêt des pratiques statistiques en Sciences Sociales et prit la responsabilité de diriger mes recherches. Ses marques de confiance répétées ont été pour moi, un stimulant permanent et m'ont donné l'énergie nécessaire pour mener à bien ce travail malgré les vents contraires d'un quotidien accaparant. Nous nous connaissons trop pour faire un long discours. Alors, je voudrais simplement lui exprimer ici mes plus sincères remerciements pour toutes les aides qu'il m'a apportées pendant ces années.*

*Cette liste de remerciements ne saurait être complète sans une attention particulière*

- *à tous mes amis et collègues, membres de l'Equipe de Statistique de l'Université des Sciences Sociales de Grenoble, qui m'ont si souvent permis d'exposer les résultats de mes recherches dans un climat chaleureux.*
- *aux étudiants de 3ème cycle qui m'ont fait l'honneur de me demander de diriger ou superviser leurs recherches,*
- *à tous ceux qui sont venus me soumettre leurs problèmes statistiques. Ils ont toujours su capter mon intérêt, qu'ils fussent informés après une courte entrevue, ou que l'ambition du projet nécessita un travail long et justifia une publication en commun des résultats.*

*Puissent ils avoir retiré de moi, autant qu'il m'ont donné et appris.*

*La réalisation matérielle de ce document est due au Service de Reprographie de l'Institut IMAG animé avec compétence et efficacité par D. IGLESIAS dont on doit une fois encore saluer le travail.*

*La frappe éclairée du texte est due à ma femme CATHERINE, qui a su cumuler pendant cette période plus de charges familiales qu'il n'est normal. On comprendra donc que mes remerciements les plus tendres et les plus chaleureux lui soient réservés.*

## TABLE DES MATIERES

|  | <i>page</i> |
|--|-------------|
| INTRODUCTION GENERALE  | xvii        |
| <b>PARTIE A : UN MODELE ALGEBRIQUE POUR L'ANALYSE<br/>DES TABLEAUX DE DISSIMILARITE SANS CONTRAINTES</b> |             |
| CHAPITRE I INTRODUCTION  | 3           |
| CHAPITRE II PROBLEMES D'EXISTENCE D'UN PLONGEMENT<br>EUCLIDIEN   | 13          |
| II.1. RAPPELS SUR LES FONCTIONS DE DISSIMILARITÉ   | 13          |
| II.2. RAPPELS SUR LES ESPACES AFFINES EUCLIDIENS   | 23          |
| II.3. NOTION DE DISSIMILARITÉ EUCLIDIENNE  | 27          |
| II.4. EXISTENCE D'UN PLONGEMENT EUCLIDIEN D'UN ESPACE QUASI-<br>MÉTRIQUE $(I, \delta)$                   |             |
| CHAPITRE III STRUCTURES AFFINES ET ANALYSE D'UNE<br>DISSIMILARITE EUCLIDIENNE                            | 55          |
| III.1. STATISTIQUES AFFINES INVARIANTES ET PLONGEMENT EUCLIDIEN  | 55          |
| III.2. CHOIX D'UNE ORIGINE DANS E ET D'UNE PONDÉRATION DE I  | 64          |
| III.3. CHOIX PARTICULIERS D'UNE ORIGINE DE E   | 69          |
| III.3.1. Le Centre de l'Hypersphère Circonscrite au Simplexe $(P_i ; i \in [n])$<br>ou Circumcentre      | 70          |
| III.3.2. Le Centre de l'Hypersphère Inscrite dans le Simplexe $(P_i ; i \in [n])$<br>ou Centre Inscrit.  | 75          |
| CHAPITRE IV STRUCTURES LINEAIRES ET ANALYSE<br>DU TRIPLE $(I, \delta, P_I)$                              | 81          |
| IV.1. RAPPELS ET NOTATIONS   | 81          |
| IV.2. LES SOUS-ESPACES $\langle E', F' \rangle$ DE $F^* \otimes E^*$                                     | 86          |
| IV.3. RECHERCHE DU SOUS-ESPACE $\langle E', F' \rangle$ LE PLUS PROCHE DE $\underline{x}$                | 90          |
| IV.4. LES STRUCTURES LINÉAIRES EN ANALYSE DU TRIPLE $(I, \delta, P_I)$                                   | 95          |
| IV.5. REPRÉSENTATION MATRICIELLE DE $\alpha$ ET $\beta$ DANS $F \otimes F$                               | 108         |
| IV.6. CARACTÉRISATION DES OPÉRATEURS ADJOINTS DE $\alpha_1$ ET $\beta_1$                                 | 113         |
| IV.7. ANALYSE DES RELATIONS ALGÈBRIQUES ENTRE F ET $F \wedge F$  | 115         |

|              |   |     |
|--------------|---|-----|
| CHAPITRE V   | LES METHODES D'ANALYSE DU TRIPLE $(I, \delta, P_1)$   | 123 |
| V.1.         | L'ANALYSE D'UN TRIPLE $(I, \delta(Z), P_1)$ DÉFINI PAR UNE DISSIMILARITÉ EUCLIDIENNE        | 124 |
| V.1.1.       | L'Analyse en Coordonnées Principales  | 124 |
| V.1.2.       | Interprétation de l'Analyse en Coordonnées Principales en termes de contrastes $2 \times 2$ | 135 |
| V.1.3.       | Les méthodes concurrentes de l'Analyse en Coordonnées Principales                           | 150 |
| V.2.         | L'ANALYSE D'UN TRIPLE $(I, \delta, P_1)$ DÉFINI PAR UNE DISSIMILARITÉ QUELCONQUE            | 166 |
| V.2.1.       | L'Analyse en Coordonnées Principales  | 167 |
| V.2.2.       | Les problèmes de constante additive et leurs généralisations                                | 173 |
| CHAPITRE VI  | L'ANALYSE DU TRIPLE $(I, \delta, Q_{ x })$  | 191 |
| VI.1.        | RAPPELS ET DÉFINITIONS  | 193 |
| VI.2.        | LE CODAGE UNIDIMENSIONNEL DU TRIPLE $(I, \delta, Q_{ x })$                                  | 198 |
| VI.3.        | LE CODAGE MULTIDIMENSIONNEL DU TRIPLE $(I, \delta, Q_{ x })$                                | 199 |
| VI.3.1.      | Outils algébriques d'Analyse du triple $(I, \delta, Q_{ x })$                               | 200 |
| VI.3.2.      | Le cas particulier $Q_{ x } = P_1 \otimes P_1$  | 206 |
| VI.4.        | LES MÉTHODES DE CODAGE MULTIDIMENSIONNEL DU TRIPLE $(I, \delta, Q_{ x })$                   | 207 |
| VI.4.1.      | La méthode IV de quantification   | 209 |
| VI.4.2.      | La méthode MINRES   | 213 |
| VI.4.3.      | La méthode MINRES 2   | 228 |
| VI.4.4.      | La méthode ELEGANT  | 239 |
| VI.4.5.      | Introduction d'une constante additive   | 241 |
| CHAPITRE VII | ANALYSE ORDINALE DU TRIPLE $(I, \delta, Q_{ x })$   | 248 |
| VII.1.       | LES MÉTHODES DE CODAGE FONCTIONNEL DES RANGS  | 248 |
| VII.1.1.     | La méthode MINRES   | 252 |
| VII.1.2.     | La méthode MINRES 2   | 261 |
| VII.2.       | LES METHODES RELATIONNELLES   | 261 |
| VII.2.1.     | Analyse Ordinale du triple $(I, \delta, Q_{ x })$   | 263 |
| VII.2.2.     | Analyse Ordinale du triple $(I, \{\delta^k, k \in [K]\}, Q_{ x })$                          | 265 |
| REFERENCES   |   | 273 |

**PARTIE B : UN MODELE ALGEBRIQUE POUR L'ANALYSE  
DES TABLEAUX DE DISSIMILARITE SOUS CONTRAINTES**

|              |   |            |
|--------------|---|------------|
| CHAPITRE I   | INTRODUCTION  | <b>295</b> |
| CHAPITRE II  | L'ANALYSE DU TRIPLE $(I, \delta_I, Q_{ x })$<br>SANS CONTRAINTES                  | <b>309</b> |
| II.1.        | RAPPELS ET NOTATIONS  | 309        |
| II.2.        | L'ANALYSE EN COORDONNÉES PRINCIPALES DU TRIPLE $(I, \delta_I, P_I)$               | 312        |
| II.3.        | LE CODAGE MULTIDIMENSIONNEL AFFINE DU TRIPLE $(I, \delta_I, Q_{ x })$             | 318        |
| II.3.1.      | Outils Algébriques d'Analyse du Triple $(I, \delta_I, Q_{ x })$                   | 319        |
| II.3.2.      | Présentation d'un Algorithme de Codage Multidimensionnel                          | 325        |
| CHAPITRE III | L'ANALYSE DU TRIPLE $(I, \delta_I, Q_{ x })$<br>SOUS CONTRAINTES DE CONFIGURATION | <b>333</b> |
| III.1.       | CONTRAINTES LINEAIRES PORTANT SUR LA CONFIGURATION X                              | 334        |
| III.1.1.     | L'approche de CARROLL et al.: une formulation en<br>termes de régression.         | 334        |
| III.1.2.     | Résolution du problème en termes d'opérateurs                                     | 347        |
| III.1.3.     | Nouvelle approche en termes de rotations procrustes                               | 354        |
| III.1.4.     | Extension de ces modèles au cas de<br>contraintes structurelles                   | 360        |
| III.2.       | CONTRAINTES ORDINALES PORTANT SUR LA CONFIGURATION                                | 365        |
| III.3.       | OPTIMISATION DES CRITERES D'ÉCART ENTRE DISSIMILARITÉS                            | 371        |
| CHAPITRE IV  | L'ANALYSE DU TRIPLE $(I, \delta_I, Q_{ x })$ SOUS<br>CONTRAINTES RELATIONNELLES   | <b>375</b> |
| IV.1.        | ANALYSE EN COORDONNÉES PRINCIPALES  | 375        |
| IV.1.1.      | Le cas des contraintes de graphe  |            |
| IV.1.2.      | Extension au cas de relations valuées   |            |
| IV.2.        | CODAGE MULTIDIMENSIONNEL  | 393        |
| IV.2.1.      | Cas d'une contrainte définie par une matrice de dissimilarité                     | 393        |
| IV.2.2.      | Cas d'une contrainte définie par plusieurs matrices de dissimilarité              | 402        |
| IV.2.3.      | Dépouillement des résultats   | 408        |
| REFERENCES   |   | <b>413</b> |

## **PARTIE C : A TOOLKIT FOR MDS OF DISSIMILARITY DATA**

|                     |   |            |
|---------------------|---|------------|
| <b>CHAPITRE I</b>   | <b>INTRODUCTION</b>   | <b>429</b> |
| <b>CHAPITRE II</b>  | <b>EUCLIDEAN EMBEDDING OF DISSIMILARITY DATA</b>                              | <b>430</b> |
| <b>CHAPITRE III</b> | <b>EVALUATION OF THE DIFFERENTIAL QUALITY OF THE REPRESENTATION OF OBJETS</b> | <b>435</b> |
|                     | <b>III.1. EVALUATION OF THE ADEQUACY OF THE REPRESENTATION OF OBJETS</b>      | <b>435</b> |
|                     | <b>III.2. EVALUATION OF THE ADEQUACY OF THE RETAINED DIMENSIONALITY</b>       | <b>439</b> |
|                     | <b>III.3. INTERPRETATION OF MAPPINGS</b>                                      | <b>440</b> |
|                     | <b>III.4. INTERPRETATION OF THE COORDINATE AXES</b>                           | <b>442</b> |
| <b>CHAPITRE IV</b>  | <b>THE CASE OF ORDINAL DATA</b>   | <b>443</b> |
| <b>REFERENCES</b>   |   | <b>449</b> |

## **PARTIE D : THE ADDITIVE DECOMPOSITION OF SOME ENTROPY FUNCTIONS AND (CONSTRAINED-) ORDINATION METHODS**

|                     |   |            |
|---------------------|---|------------|
| <b>CHAPITRE I</b>   | <b>INTRODUCTION</b>   | <b>456</b> |
| <b>CHAPITRE II</b>  | <b>THE ANALYSIS OF DIVERSITY</b>  | <b>458</b> |
|                     | <b>II.1. THE ONE-WAY ANALYSIS OF VARIANCE</b>   | <b>458</b> |
|                     | <b>II.2. ONE EXTENSION OF THIS METHODOLOGY TO GENERAL DISSIMILARITY FUNCTIONS</b>             | <b>461</b> |
| <b>CHAPITRE III</b> | <b>THE EUCLIDEAN EMBEDDING OF DISSIMILARITY DATA</b>  | <b>464</b> |
| <b>CHAPITRE IV</b>  | <b>ORDINATION METHODS WITH INSTRUMENTAL VARIABLES</b>   | <b>471</b> |
|                     | <b>IV.1. THE DISCRIMINANT COORDINATES ANALYSIS</b>  | <b>472</b> |
|                     | <b>IV.2. THE REPRESENTATION OF THE STATISTICAL UNITS IN DISCRIMINANT COORDINATES ANALYSIS</b> | <b>477</b> |
| <b>REFERENCES</b>   |   | <b>483</b> |

## INTRODUCTION GENERALE

Les travaux présentés dans ce mémoire ont pour thème commun l'exploration des outils métriques utilisables en statistique multidimensionnelle. Du point de vue formel, les méthodes d'Analyse Multidimensionnelle des Données associent à un système relationnel empirique  $\mathcal{R} = \{\mathcal{R}^j, j \in J\}$  où  $\mathcal{R}^j$  est une relation  $k^j$ -aire définie sur un ensemble  $I$ , un système relationnel métrique  $\mathcal{Y} = \{\mathcal{Y}^j, j \in J\}$  où la relation  $\mathcal{Y}^j$  est à valeurs dans un espace métrique de dimension fixée par l'analyste. Leur emploi fonde l'essentiel de son attrait et de son efficacité sur l'emploi de représentations graphiques, dont la lecture obéit aux règles de la géométrie

Les méthodes d'Analyse des Données de Dissimilarité fournissent un cas particulier de telles méthodes, où  $\mathcal{R}^j$  est définie par une relation binaire. Dans le cas le plus simple, on dispose d'un couple  $(I, \delta)$  formé d'un ensemble fini  $I$  d'unités statistiques et d'une fonction réelle positive  $\delta$ , définie sur  $I \times I$ , mesurant le degré de dissemblance (ou d'écart) entre les unités statistiques. On se propose alors de construire un plongement de  $I$  dans un espace métrique  $(X, \rho)$  de dimension fixée a priori, c'est à dire une application

$$\begin{aligned} \varphi &: I \rightarrow (X, \rho) \\ i &\mapsto \varphi(i) = x_i \in X \end{aligned}$$

telle que

$$d_{ij} = \rho(\varphi(i), \varphi(j)) = \rho(x_i, x_j)$$

Le plus souvent,  $X$  est un espace affine et une géométrie Euclidienne est retenue. Ce choix présente l'avantage de la simplicité, bien que de nombreuses techniques aient été développées pour construire des représentations dans des espaces métriques de structure plus complexe. Ainsi la collaboration entre statisticiens et psychologues grenoblois pour l'étude de la perception tactile, dirige actuellement nos recherches vers le développement de méthodes de description de données de dissimilarité dans le cadre des géométries de Cayley-Klein.

Notre intérêt pour les méthodes d'analyse des données de dissimilarité est né de trois observations:

i) Tout d'abord, le dépouillement des graphiques résultant de l'application de techniques d'Analyse Multidimensionnelle des Données utilise, en général de façon indifférenciée, deux structures mathématiques complémentaires: la géométrie des angles formés par les vecteurs et celle des distances inter-points. Ces deux outils mathématiques peuvent d'ailleurs être retenus pour construire une axiomatique de la structure d'espace Euclidien, qui guide notre perception des représentations graphiques. De plus, toute méthode d'analyse qui construit une représentation des données de façon à minimiser un critère de distorsion des distances inter-points construit des graphiques d'aspect identique aux configurations obtenues par une méthode destinée à reconstruire au mieux les angles entre vecteurs. Mais dans ce cas la distorsion des distances peut être importante et inversement, aucune rationalité ne justifie une lecture angulaire, lorsque la fonction de dissimilarité analysée n'est pas Euclidienne. L'identification des structures affines et vectorielles des espaces de représentation ajoute alors à la confusion.

Du point de vue mathématique, étant donné un espace vectoriel de dimension finie, la notion de biplot (cf. GABRIEL (1976), GREENACRE (1978)) est la forme la plus avérée des lectures angulaires des représentations graphiques en analyse multivariée, puisqu'elle utilise les relations de dualité entre représentants des lignes et des colonnes du tableau de données. L'extension de cette approche au cas des espaces vectoriels de dimension infinie repose sur l'existence d'une structure d'espace de Hilbert, et son emploi en Analyse des Données, étudié en détail par DAUXOIS et POUSSE (1976), accorde un rôle primordial aux notions de produit scalaire défini sur un espace vectoriel et fonde l'ensemble des méthodes statistiques d'analyse des structures de covariance ou de corrélation entre variables. Par suite, de nombreuses fonction de similarité entre objets sont construites en termes de cosinus d'angles et de nombreuses techniques statistiques d'analyse multivariée tirent leur légitimité de l'utilisation de modèles de décomposition additive des matrices de produits scalaires et de la construction d'approximations fortes des produits scalaires, précisées par la démonstration de théorèmes d'optimalité. On peut consulter à ce sujet RAO (1964, 1980), OKAMOTO et KANAZAWA (1968) et OKAMOTO (1969) dans le cas de l'Analyse en Composantes Principales ou POLIT (1986) et d'AUBIGNY et POLIT (1988) pour l'extension de ces résultats au cas d'un  $n$ -hypercube de données ( $n \geq 3$ ). Ces propriétés d'optimalité sont alors une conséquence de l'adaptation des outils algébriques mis en oeuvre, à la description d'un espace vectoriel muni d'un produit scalaire.

Nous devons alors parler d'approximation faible des dissimilarités du fait de la pauvreté des résultats d'optimalité concernant l'approximation des distances initiales par des distances Euclidiennes évaluées sur la représentation résumée, cf. GREENACRE (1978). Cette situation résulte de deux causes. Tout d'abord, pour un critère d'approximation donné, reconstruire au mieux des angles entre vecteurs ou des distances entre les extrémités de ces vecteurs conduit à deux problèmes en général non équivalents. D'autre part, la caractérisation métrique des espaces Euclidiens pose un problème difficile qui explique le faible nombre de travaux mathématiques utilisables en statistique pour l'analyse intrinsèque des espaces métriques, c'est à dire sans référence à la structure

d'espace vectoriel normé sous-jacente ou à un système de coordonnées. Notre intuition est de plus souvent mise en défaut lors de l'analyse des structures métriques (cf. LE CALVE (1985)).

ii) Malgré les risques importants d'erreur d'interprétation qui en résultent, les utilisateurs de méthodes d'analyse multivariée marquent une propension marquée à dépouiller les résultats d'analyse en s'appuyant sur une lecture des graphiques en termes de distances inter-points. Ils sont encouragés en cela par certaines présentations approximatives des méthodes employées. Ainsi le recours au principe barycentrique fonde certaines présentations de l'analyse des correspondances (cf. LEBART et al. (1986)) sur l'attrait pour l'esprit humain de telles lectures naïves. Bien que ce principe ne soit justifié que dans des situations exceptionnelles, il est souvent considéré à tort par l'utilisateur non averti, comme la règle générale.

D'autre part, le recours à une lecture distanciée peut paraître paradoxale dans le cas de données de profil, lorsqu'on compare la richesse des outils (règles de lecture, aides à l'interprétation) guidant une lecture angulaire et qui résultent des relations de dualité entre unités statistiques  $\{\omega_i, i \in I\}$  et variables  $\{X^j, j \in J\}$  induites par le tableau de données  $X = \{x_i^j = X^j(\omega_i); i \in I, j \in J\}$ , au faible nombre d'outils équivalents permettant de contrôler la lecture des graphiques en termes de distances inter-points. L'analyse des tableaux de données de dissimilarité est plus complexe puisqu'un grand nombre d'outils algébriques disparaissent précisément avec la donnée a priori de variables explicites et la perte induite de ces relations de dualité.

iii) Du fait des difficultés théoriques rencontrées, l'approximation métrique de données de dissimilarité pose des problèmes numériques importants nécessitant le recours à des moyens de calcul, très longtemps non accessibles. On traite alors un problème non linéaire, dont la résolution nécessite la construction d'algorithmes. Par suite l'étude des propriétés statistiques des résumés ainsi construits reste en grande partie un problème ouvert. Quelles qualités doit on exiger d'estimateurs construits par cette voie? Quelle modélisation probabiliste nécessaire au développement d'inférences statistiques ou d'expérimentations sur données simulées est possible?

Les travaux présentés dans ce mémoire poursuivent quatre objectifs:

- (A) Définir un cadre algébrique unique, permettant d'étudier et de comparer les termes et les moyens des méthodes existantes,
- (B) Mener une étude détaillée des classes de contraintes qu'il est possible de traiter dans ce cadre algébrique d'analyse de données de dissimilarité,
- (C) Utiliser ce formalisme pour évaluer la qualité des solutions et pour dégager des outils d'aide à l'interprétation des résultats de l'analyse,
- (D) Fournir un cadre méthodologique permettant de traiter des données de dissimilarité collectées dans le cadre d'une expérience contrôlée.

Nous espérons contribuer ainsi à définir les éléments de base d'une méthodologie d'Analyse des Données de Dissimilarité rendue nécessaire par le caractère original des problèmes statistiques posés et par le développement insuffisant d'outils de diagnostique dans ce domaine par comparaison aux apports de l'école américaine en Régression Multiple (cf. BELSEY et al. (1980)) ou de l'école française en Analyse Multidimensionnelle Linéaire des Données (cf. LEBART et al. (1986)) et en Classification Automatique (cf. JAMBU (1978)).

Notre étude des méthodes d'analyse des données de dissimilarité ne prétend pas à l'exhaustivité. Elle laisse pendants de nombreux problèmes théoriques et pratiques. En particulier, nous nous sommes limités à de rares exceptions près, au plongement Euclidien de données de dissimilarité. De plus nous ne présentons pas de traitements de données. Le lecteur intéressé peut en particulier trouver des informations complémentaires sur le sujet dans les livres de TORGERSON (1958), SHEPARD et al. (1972), KRUSKAL et WISH (1978), GREENACRE (1978), HARTMAN (1979), LINGEOS et al. (1979), BORG (1981), HEISER (1981), SCHIFFMAN et al. (1981), GOWER (1982), DAVISON (1983) et MEULMAN (1986).

Dans un cadre exploratoire, la méthodologie usuelle de l'Analyse Linéaire des Données tire pour beaucoup sa crédibilité d'une communauté de formalisme mathématique utilisé avec l'Analyse Multivariée qui interprète le tableau traité comme l'observation d'un vecteur aléatoire sur un échantillon de taille suffisante et se propose d'estimer les paramètres décrivant le support de la loi de répartition observée. Le traitement de structures statistiques multivariées particulières offre alors une palette de jauges permettant d'étalonner les méthodes sur des données de caractéristiques connues, par l'application de résultats mathématiques ou à l'aide de simulations (voire GIFI (1980) pour une discussion plus complète de ce point de vue). D'autre part, la compréhension détaillée des fondements mathématiques de ces méthodes et le développement d'outils d'aides à l'interprétation permettent une lecture raisonnée des résultats d'analyse, de nature à limiter les risques de conclusions hâtives ou résultant de considérations subjectives sans grand rapport avec les données traitées. De ces deux points de vue l'analyse des données de dissimilarité nous a semblé manquer actuellement de connaissances suffisantes des propriétés des méthodes, en particulier dans le cas le plus courant de l'analyse d'une fonction de dissimilarité non Euclidienne.

De plus, même dans le cas simple d'une description des positions relatives des unités statistiques dans un espace métrique Euclidien, la modélisation de la loi de probabilité d'une matrice aléatoire de distances Euclidiennes constitue un problème difficile (cf. WALKER (1976)), et l'on dispose actuellement de peu de jauges utilisables pour étalonner les méthodes. C'est pourquoi il nous a semblé utile de développer les outils

algébriques permettant de proposer de nouvelles formulations du problème et permettre ainsi le développement futur d'études probabilistes.

Le cas de référence, fourni par l'étude d'une fonction de dissimilarité Euclidienne, c'est à dire satisfaisant aux axiomes d'une distance Euclidienne, permet de décrire la plupart des méthodes d'Analyse Linéaire des Données en termes de distance, cf. GIFI (1980), MEULMAN (1986) Cependant, alors que les techniques statistiques accordent alors une place prépondérante à la détermination des sous-espaces décrits par l'espérance mathématique de l'élément aléatoire étudié, notre présentation accorde une place prépondérante à la notion de dissimilarité Euclidienne, c'est à dire d'écart entre unités statistiques. Ce changement de perspective se traduit dans le cas unidimensionnel, par l'identité

$$\frac{1}{2n(n-1)} \sum_{i,j=1}^n \delta_{ij}^2 = \frac{1}{2n(n-1)} \sum_{i,j=1}^n (y_i - y_j)^2 = \frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2 = \text{Var}(Y)$$

L'analyse Euclidienne de la dissimilarité  $\delta_{ij} = \sqrt{(y_i - y_j)^2}$  fournit donc une description du domaine de variation de la variable observée, et de façon indirecte seulement du support de la variable centrée  $\check{Y}$ . Les méthodes d'analyse des tableaux de dissimilarité fournissent en fait un cadre unificateur de présentation des méthodes d'analyse de la dispersion, plutôt qu'un outil de description du support d'un indicateur de position centrale de la loi de probabilité étudiée.

La partie A de ce mémoire reprend en le complétant sensiblement, notre exposé aux Journées Franco-Soviétiques de Statistique de Paris (1984). Nous y définissons un CADRE ALGÈBRE D'ANALYSE DES TABLEAUX DE DISSIMILARITÉ suffisamment général pour montrer les relations existant entre les méthodes qui privilégient la structure hilbertienne de l'espace de représentation, dont l'Analyse en Coordonnées Principales, et les méthodes de Codage Multidimensionnel, attachées au respect des structures métriques. Nous nous sommes attachés à expliciter les structures algébriques et fonctionnelles mises en œuvre, dans un souci d'homogénéité aux présentations françaises des méthodes statistiques d'analyse des données de profil. Les aspects algorithmiques sont alors traités de façon séparée, afin de distinguer les aspects méthodologiques des problèmes de mise en œuvre.

Nous montrons que les méthodes Hilbertiennes sont essentiellement adaptées à l'étude des dissimilarités Euclidiennes, c'est à dire représentables en tant que mesures de la distance entre vecteurs représentatifs des unités statistiques dans un espace affine Euclidien. On étudie alors le triple  $(I, \delta, P_I)$ , où  $\delta$  est une dissimilarité propre définie sur  $I \times I$  et  $P_I$  est une pondération positive des éléments de  $I$ . Ces méthodes tirent leur

spécificité de l'utilisation de la relation classique entre produits scalaires et normes hilbertiennes:

$$\|x_i - x_j\|^2 = \|x_i\|^2 + \|x_j\|^2 - 2 \langle x_i, x_j \rangle$$

souvent présentée sous le nom de "transformation de Torgerson", bien qu'elle soit connue depuis les travaux de GAUSS (1831), DIRICHLET (1850) et HERMITE (1850). L'application de cette formule suppose que l'on dispose d'une mesure de dissimilarité Euclidienne et que l'on ait choisi une origine, alors que la distance inter-points utilise uniquement la structure d'espace affine de l'espace ambiant.

En Analyse en Coordonnées Principales, la matrice  $W$  déduite des données par la transformation de Torgerson

$$w_{ij} = -\frac{1}{2} (\|x_i - x_j\|^2 - \|x_i\|^2 - \|x_j\|^2) \quad \forall (i, j), (i, j) \in I \times I$$

est décomposée à l'aide d'une analyse spectrale qui fournit la configuration de rang  $r$  désirée, selon une démarche analogue à l'Analyse en Composantes Principales. Lorsque  $\delta$  n'est pas Euclidienne, par construction ou par suite d'erreurs de mesures, la matrice  $W$  n'est plus Gramienne. Les difficultés qui en découlent limitent alors l'intérêt d'une telle formulation du problème, et nous conduisent à poser une définition plus générale de l'Analyse en Coordonnées Principales.

La deuxième classe de méthodes, née des travaux de HAYASHI (1952), SHEPARD (1962), KRUSKAL (1964) et GUTTMAN (1968) et regroupée sous l'intitulé de CODAGE MULTIDIMENSIONNEL, étudie le triple  $(I, \delta, Q_{I \times I})$  où la pondération  $Q$  définie sur  $I \times I$ , fournit un cadre d'analyse plus général que la classe précédente. La falsifiabilité du modèle de description porte alors directement sur les dissimilarités et l'on ajuste aux données en un sens à préciser, des distances déduites d'une configuration optimale  $X$  de rang faible. La résolution de ce problème conduit alors à une approximation forte des dissimilarités observées par les distances reconstruites.

Le cadre algébrique proposé permet de comparer ces deux classes de méthodes. Son élaboration s'inscrit dans une problématique statistique fondée sur la construction de l'application  $t$  qui associe à un vecteur  $z$  la liste de ses contrastes simples:

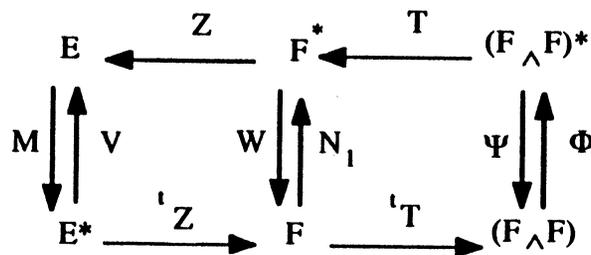
$$t: z \rightarrow y = \{(z_i - z_j); i, j \in [n]\} \text{ si } z = \{z_i; i \in [n]\} \text{ et } [n] = \{1, \dots, n\}$$

Nous montrons que  $y$  décrit sous forme vectorielle les éléments d'une matrice antisymétrique  $Y$  de terme général  $y_{ij} = (z_i - z_j)$ . Pour un choix adapté de bases, la matrice de  $t$  est  $T$  et le vecteur  $z$  de  $i$ -ème ligne  $z_i$ , a pour image par  $t$  le vecteur  $y = {}^tTz$  dont

la ligne  $b=(i,j)$  est l'élément  $y_b = (z_i - z_j)$  qui intervient dans l'évaluation de la distance Euclidienne induite par une forme quadratique  $m$  de matrice  $M$

$$\delta_{ij}^2(Z) = \|z_i - z_j\|^2 = {}^t(z_i - z_j) M (z_i - z_j) = m(z_i - z_j)$$

$\delta$  est une distance particulière déterminée par une norme  $m$  regardée comme une fonctionnelle opérant sur l'ensemble des opérateurs antisymétriques agissant sur un espace vectoriel  $F$ . Nous formulons en conséquence le problème du Codage Multidimensionnel dans l'espace vectoriel  $F \wedge F$ , produit extérieur de  $F$  par lui même et décrit par les opérateurs antisymétriques définis sur  $F$ .  $t$  définit alors une isométrie partielle entre  $F$  et  $F \wedge F$ , traduite par le schéma



Nous étudions chacun des constituants de cette structure algébrique et nous explicitons en particulier les liens entre l'Analyse en Coordonnées Principales, décrite de façon classique dans  $F$ , et les méthodes de Codage Multidimensionnel, dont nous rendons compte dans  $F \wedge F$ . Une attention particulière est portée à l'étude des propriétés de la matrice des contrastes  $2 \times 2$  formés à partir du tableau à double entrée  $\Delta$ ,  $\Psi = {}^t T W T$  de terme général

$$\Psi_{ij,kl} = -\frac{1}{2} (\delta_{ik}^2 + \delta_{jl}^2 - \delta_{il}^2 - \delta_{jk}^2)$$

Le plan de cet exposé est le suivant:

Après un rappel des définitions et résultats portant sur les notions de dissimilarité, les préordonnances, et les espaces affines Euclidiens, le chapitre II est consacré à l'exposé des théorèmes d'existence d'un plongement Euclidien d'un espace semi-métrique dans un espace affine de dimension donnée. Puis nous rappelons brièvement les difficultés rencontrées dans l'établissement et la validation empirique de résultats portant sur l'existence d'une congruence entre un espace (semi-) métrique et un espace Euclidien.

Le chapitre III est consacré à l'étude des structures affines intervenant lors de l'analyse d'une fonction de dissimilarité Euclidienne, mises en évidence par d'AUBIGNY (1975) et détaillées par FICHET (1983). Nous explicitons dans ce cadre les relations entre les travaux d'origine française, qui insistent sur le rôle de la pondération des unités statistiques et le choix induit de métrique et les travaux d'origine anglaise, cf. GOWER

(1966, 1982 a, b, 1984, 1985), qui portent sur la détermination d'une origine de l'espace de représentation.

Le chapitre IV suppose ces problèmes résolus et expose tout d'abord une formulation tensorielle de l'Analyse en Composantes Principales qui nous permet d'étudier les conditions suffisantes d'existence d'une propriété d'emboîtement des sous-espaces de représentation résumée:

$$E_r \subset E_{r'} \text{ si } \dim(E_r) < \dim(E_{r'})$$

Le formalisme tensoriel nous permet de plus de présenter une analyse fine des opérateurs utilisés sous forme matricielle dans le traitement classique de L'Analyse en Coordonnées Principales qui intègre les travaux de CRITCHLEY (1985 a, b, c) et les complète par l'introduction d'une dualité algébrique entre les espaces de description des unités statistiques et des paires d'unités statistiques respectivement. Cette dualité joue selon nous un rôle essentiel dans la compréhension des méthodes d'analyse des tableaux de dissimilarité, en termes d'Analyse en Coordonnées Principales, de Codage Multidimensionnel (métrique ou ordinal), de Classification Hiérarchique et d'Arbres Additifs.

La démarche classique d'Analyse du Triple  $(I, \delta, P_1)$  est présentée au chapitre V. Le cas d'une dissimilarité Euclidienne est tout d'abord étudié, et nous replaçons cette méthode dans une perspective beaucoup plus générale fondée sur l'analogie au cas des données de profil. Nous adoptons ici la distinction fondamentale faite par McDONALD (1968) entre méthodes tautologiques, telles que l'Analyse en Coordonnées Principales, et méthodes modélisantes. Nous introduisons alors une classe nouvelle de méthodes dites d'Analyse Factorielle du Triple  $(I, \delta, P_1)$  et nous présentons des méthodes de construction de la solution de ce problème. Nous accordons ensuite une attention particulière à une méthode d'analyse qui adapte au cadre des données de dissimilarité la notion d'Analyse d'Images Partielles développée par GUTTMAN (1953) pour l'analyse des données de profil. C'est une analyse en Coordonnées, donc directement concurrente de l'Analyse en Coordonnées Principales, pour laquelle nous montrons de plus que les résultats de GOWER (1982) permettent de fixer une origine de l'espace de représentation cohérente avec notre approche. Notre apport à ce sujet réside moins dans la volonté de proposer des méthodes nouvelles d'analyse, que dans l'explicitation de la symétrie totale des situations rencontrées lors de l'analyse des tableaux de covariance et de l'analyse des tableaux de dissimilarité Euclidienne. Cette symétrie permet d'enrichir notablement la méthodologie d'Analyse des Tableaux de Dissimilarité par analogie aux résultats établis en Analyse Factorielle Classique.

La deuxième partie de ce chapitre est consacrée à l'étude du rôle des contrastes  $2 \times 2$  en Analyse en Coordonnées Principales. Cette conséquence de la dualité entre unités statistiques et paires définie plus haut, ne semble pas avoir été notée auparavant et nous permet de donner un statut statistique à cette méthode. Elle fournit des règles de lecture des résultats qui faisaient défaut à notre avis. Nous montrons de plus l'intérêt des contrastes  $2 \times 2$  pour d'autres problèmes d'analyse des tableaux de dissimilarité, tels que les arbres additifs. Nous démontrons enfin un théorème d'existence d'un plongement Euclidien d'un espace quasi-métrique  $(I, \delta)$  en terme de positivité des contrastes  $2 \times 2$ , qui montre en particulier les liens de la méthode avec l'analyse des indices d'association appelés odds ratio et proposés par YULE (1912) pour l'étude des tables de contingence binaire.

La dernière partie de ce chapitre traite le cas des fonctions de dissimilarité non nécessairement Euclidiennes. Nous donnons alors une interprétation, nouvelle semble-t-il, des problèmes de constante additive en termes de perturbation, fondée sur l'analogie aux méthodes de Ridge Régression. Puis nous montrons que la généralisation de ces méthodes au cas d'une perturbation diagonale quelconque, analogue de la Régression Bayésienne, permet de retrouver l'Analyse Factorielle du Triple dans un cas, et les modèles d'Analyse de données de dissimilarité avec effets de contexte (ou de densité spatiale) dans un autre cas. Nous généralisons alors les travaux de CAILLEZ (1983) et proposons un algorithme résolutif, homogène à la logique de l'Analyse en Coordonnées Principales, et donc complémentaire des travaux de SAITO (1983, 1985, 1986) en Codage Multidimensionnel. Ce modèle permet de plus de construire une décomposition additive d'une dissimilarité comme somme d'une distance Euclidienne et d'une distance à centre, dont l'existence est prouvée par LE CALVE (1988) et qui est optimale au sens du critère adopté par Caillez.

Le chapitre VI est consacré aux méthodes d'Analyse du Triple  $(I, \delta, Q_{I \times I})$ , où  $Q_{I \times I}$  est une pondération des paires d'unités statistiques. La prise en compte d'une pondération  $Q_{I \times I}$  opérant sur  $I \times I$  nécessite des outils mathématiques particuliers. Nous montrons tout d'abord l'adaptation du cadre algébrique proposé à l'étude du triple  $(I, \delta, Q_{I \times I})$  et nous l'utilisons pour rendre compte des algorithmes de Codage Multidimensionnel et les comparer. Nous retrouvons en particulier les équations de l'Analyse en Coordonnées Principales dans le cas d'une pondération produit, du type  $Q_{I \times I} = P_I \otimes P_I$ . Cette formulation du Codage Multidimensionnel nous permet, de façon plus importante à nos yeux, de fournir une interprétation nouvelle du problème posé en termes d'énergie dissipée dans un réseau électrique, complémentaire de celle qui découle de l'utilisation des contrastes  $2 \times 2$  et s'exprime en termes plus classiques pour le statisticien, d'analyse approchée du potentiel d'interaction décrit par un tableau à double entrée.

Le chapitre VII traite le problème complémentaire posé par l'utilisation de transformations des données en Codage Multidimensionnel. Celles ci sont en général imposées par des contraintes de mesurage induites par le processus de collecte des données. On retrouve ainsi les origines des recherches de SHEPARD (1962), KRUSKAL (1964) et BENZECRI (1965) qui appliquent les méthodes de Codage Multidimensionnel dans un cadre ordinal et sont donc conduits à prendre en compte une transformation monotone des dissimilarités observées. Nous distinguons deux classes de critères d'approximation proposés dans la littérature pour l'analyse du cas ordinal. La première classe utilise les recodages fonctionnels de l'indice de rang associé à une préordonnance observée. Nous montrons comment s'intègre cette logique dans la méthodologie du Codage Multidimensionnel et dans le cadre algébrique défini plus haut. La régression isotone est alors resituée dans ce cadre en termes de dualité Lagrangienne et la méthode alternative proposée par GUTTMAN (1968) en termes d'images de rangs reçoit ici une interprétation nouvelle en termes d'Analyse Canonique. La deuxième classe de critères est elle aussi connue et a fait l'objet de propositions d'algorithmes (cf. en particulier de LEEUW (1970)). Elle utilise les propriétés relationnelles de la préordonnance induite par la dissimilarité observée. Son analyse nécessite de compléter le formalisme algébrique mis en place précédemment. et nous montrons les liens existant entre le formalisme résultant et certaines méthodes classiques de la statistique nonparamétrique et de la théorie des votes. L'extension des méthodes précédentes au cas d'un échantillon de K tableaux de dissimilarité mesurées sur une échelle ordinale est alors éclairée par ces liens. Nous décrivons donc brièvement le principe de construction d'un algorithme résolutif pour ce problème.

Le texte consacré à l'étude des méthodes d'ANALYSE DES TABLEAUX DE DISSIMILARITÉ SOUS CONTRAINTES (PARTIE B), fournit une version complétée de notre exposé aux rencontres Franco-Soviétique de Statistique (SOUZDAL 1987). La classe des contraintes de structure introduite a pour but de prendre en compte dans l'analyse, des informations externes portant soit sur la configuration fixant les coordonnées des points représentatifs des unités statistiques, soit sur les distances entre points représentatifs. Ces informations externes résultent en général de connaissances a priori, issues par exemple d'une théorie sur les données ou bien d'hypothèses de travail formulées lors de l'analyse des résultats d'une analyse antécédente du même jeu de données.

Notre étude ne porte pas sur ce qu'il est convenu d'appeler des contraintes de régularité. Celles-ci consistent par exemple à autoriser une transformation des données observées, déterminée dans une classe de transformations admissibles telles que les fonctions monotones utilisées en Codage Multidimensionnel Ordinal, ou encore les fonctions polynomiales (cf. KRUSKAL et WISH (1978)), les fonctions splines (monotones ou non) utilisées par WINSBERG et RAMSAY (1983). et les lisseurs empiriques étudiés par

EL HADRI (1988) L'imposition de contraintes de régularité réduit le nombre de paramètres définissant la transformation des données et améliore ainsi la stabilité des résultats. Elle relève donc de préoccupations distinctes mais complémentaires de la recherche de modèles de description des données.

Nous nous intéressons au cas des données de simple observation. de LEEUW et HEISER (1980) distinguent alors trois types d'applications, en fonction des connaissances a priori de l'analyste. La situation la plus simple à traiter est rencontrée principalement dans le domaine des sciences de la nature. CRIPPEN (1977) en fournit un exemple en biochimie, à propos du positionnement spatial des molécules. On sait alors qu'il existe une structure spatiale Euclidienne tridimensionnelle des molécules que l'on se propose de la comparer à la configuration "optimale" obtenue par un algorithme de Codage Multidimensionnel appliqué à des données observées parfois incomplètes et le plus souvent entâchées d'erreurs de mesure.

Ce type de situation est distinct du cadre d'application courant en géographie (GILBERT (1974), KENDALL (1971)) et dans certaines expériences de psychologie perceptive (EKMAN (1954), INDOVS (1975, 1979)) pour lequel le caractère Euclidien de la structure spatiale sous-jacente n'est pas assuré et l'on construit une représentation Euclidienne ou plus généralement dans un espace de Minkoski de type  $l_p$ , approchant au mieux la structure géométrique de référence. SHEPARD (1974) précise que son objectif initial était de mettre à l'épreuve de l'expérience une théorie et de dégager par une représentation graphique une structure spatiale latente qui en résultait.

Le mode d'utilisation des méthodes de Codage Multidimensionnel le plus répandu, s'inscrit dans un cadre exploratoire. On traite alors l'information contenue dans le triple  $(I, \delta; Q_{|X|})$ . La recherche d'une interprétation plausible de la représentation des unités statistiques ainsi construite repose alors sur l'inspection des positions relatives des unités statistiques. Cependant, trois difficultés techniques sont rencontrées lors de l'utilisation des algorithmes de Codage Multidimensionnel.

Tout d'abord, à l'exception de l'Analyse en Coordonnées Principales, le problème posé n'est pas réductible à une technique d'analyse linéaire des données et n'admet donc pas de solution analytique explicite. Les algorithmes itératifs mis en œuvre fournissent toujours une solution dont la qualité et la pertinence sont parfois difficiles à évaluer. Les algorithmes de sous-gradient mis en œuvre sont de plus sensibles au choix de la configuration de départ. Ils peuvent converger vers un optimum local ou un point stationnaire. Leur vitesse de convergence est en général lente et influencée par la courbure du graphe de la fonction objectif au voisinage de l'optimum. Leur mise en œuvre conduit donc en général à itérer les analyses d'un même tableau de données. On recherche de plus des représentations stables des données, présentant un degré d'unicité suffisant. Enfin,

dans le cas ordinal, on est parfois amené à se prémunir contre l'obtention de solutions dégénérées (cf. KRUSKAL 1964).

En deuxième lieu, l'unicité de la configuration réalisant le minimum de la fonction objectif est d'autant moins assurée que le nombre d'unités statistiques représentées est faible relativement à la dimension de l'espace de représentation (SHERMAN (1972)). De plus des configurations très distinctes peuvent correspondre à une même valeur minimale de la fonction objectif. Une telle situation traduit par exemple un déplacement d'un sous-groupe des unités statistiques par symétrie autour d'un axe. (cf. CAELLI (1981)). Le problème s'aggrave lorsque l'on recherche des représentations non Euclidiennes, par exemple dans un espace de Minkowski.

La troisième difficulté résulte de la nécessité d'itérer les constructions de représentations spatiales des données dans la phase d'identification d'un modèle descriptif, afin de mener une analyse approfondie des données, permettant de comprendre la structure de la relation de ressemblance étudiée. Chaque analyse joue un rôle provocateur et éveille des hypothèses de travail que l'analyste souhaite mettre à l'épreuve à l'aide d'un programme d'Analyse de Tableaux de Dissimilarités capable de prendre en compte des contraintes traduisant ces hypothèses de travail lors de la construction de nouvelles configurations. Cette démarche exploratoire doit être très nettement distinguée d'une méthodologie relevant de la statistique inférentielle, puisqu'un seul jeu de données est soumis à plusieurs analyses itérées, sous différentes hypothèses descriptives traduites par des contraintes spécifiques. La validation statistique éventuelle de la structure ainsi explicitée nécessite une étude confirmatoire complémentaire, menée dans une optique inférentielle et fondée sur un nouveau corpus de données, collecté selon une méthodologie d'échantillonnage rigoureusement contrôlée. Dans cette phase confirmatoire, les hypothèses testées sont décrites de façon classique comme en Statistique multivariée, par la donnée d'un ensemble de contraintes de structure. La partie D de ce mémoire illustre cette démarche dans le cas de contraintes de structures déterminées par la donnée d'un plan d'expérience. L'élaboration récente de modèles probabilistes exploitables (RAMSAY (1969, 1977, 1978, 1980, 1982), TAKANE (1977, 1981), TAKANE et SERGENT (1983), ZINNES et al. (1974, 1977, 1981)) permettent désormais d'estimer les paramètres de modèles d'Analyse de Données de Dissimilarité dans le cadre d'une structure statistique adaptée à la résolution du problème. Bien que prometteuse, cette approche confirmatoire a cependant rarement été mise en application jusqu'à présent du fait de l'effort expérimental qu'impose la construction d'un échantillon de matrices de dissimilarité, et de l'absence d'outils de tests statistiques adaptés à de telles données.

Les problèmes numériques posés par la prise en compte de contraintes lors de l'analyse de données de dissimilarité ont retenu très tôt l'attention des promoteurs des méthodes de Codage Multidimensionnel. Le plus souvent l'introduction de contraintes fut justifiée par

des arguments numériques plutôt que méthodologiques. Nous étudions ici ces modèles comme des cas particuliers d'une classe de modèles décrite dans le cadre algébrique présenté plus haut. L'aspect algorithmique du problème intervient ensuite, dans la mise en oeuvre des méthodes présentées.

Comme tout modèle descriptif, les modèles utilisés dans l'analyse d'un tableau de dissimilarité sont définis par des contraintes, qui doivent être distinguées de celles qui définissent un sous-modèle du modèle de référence. Cette situation est classique en Statistique. Ainsi la régression linéaire multiple impose la contrainte d'appartenance de l'espérance mathématique de la variable réponse au sous-espace engendré par les variables explicatives. Cet impératif définit le modèle et doit être distingué de ce que l'on reconnaît usuellement comme des contraintes. Ainsi l'étude d'un modèle du type

$$y = 'Xb + \varepsilon \quad (1)$$

$$'Ab \leq c \quad (2)$$

considère que seule la relation linéaire  $'Ab \leq c$  définit les contraintes puisqu'elle caractérise une classe de sous modèles du modèle de référence défini par l'équation (1), cf. MONLEZUN et SPEED (1980), DENT (1980), LAWSON et HANSON (1974). De façon analogue, la pratique de la classification automatique impose des contraintes fortes aux données, traduites par la structure métrique de la configuration reconstruite. Ainsi les méthodes de classification hiérarchique construisent une matrice de distances ultramétriques approchant la dissimilarité de départ. Cette contrainte de modèle doit être distinguée des contraintes de contiguïté récemment introduites afin d'enrichir cette méthodologie (cf. de SARBO et MAHAJAN (1984), FERLIGOJ et BATAGELJ (1982,1983), GORDON (1973a, 1973b), LEBART(1978), LECHEVALIER (1980), LEFKOVITCH (1980)).

Nous retrouvons donc dans le cas de l'analyse d'un tableau de données de dissimilarité une ambiguïté de termes inhérente à toute démarche exploratoire soucieuse d'identifier le sous-modèle le plus adéquat d'un modèle de description de données, dans un souci légitime de description fidèle stable et parsimonieuse des données. Tout modèle non saturé peut en effet être regardé comme un sous modèle d'un modèle de référence plus général. Cette démarche progressive caractérise à notre point de vue une analyse exploratoire et l'oppose à la simple analyse descriptive.

Le chapitre I de la partie B est consacré à un historique du Codage Multidimensionnel Sous Contraintes, illustré par des exemples. Le chapitre II rappelle les outils algébriques nécessaires à notre exposé et les chapitres suivants constituent la partie essentielle de notre exposé. Nous distinguons deux classes de contraintes. Elles correspondent à deux formulations distinctes des informations a priori et réfèrent en pratique à deux situations différentes. La classe des contraintes de configuration s'applique à la matrice des

coordonnées des unités statistiques  $X$ , alors que la classe des contraintes relationnelles réfère aux distances interpoints  $D(X)$ . La correspondance non linéaire liant  $X$  à  $D(X)$  rend ces deux classes de contraintes non équivalentes en général. Nous décrivons cependant un certain nombre de situations dans lesquelles les deux approches sont équivalentes.

Les contraintes de configuration font référence à la structure d'espace vectoriel normé de l'espace de représentation. Dans un cadre linéaire, nous montrons que la classe des contraintes la plus générale est fournie par Carroll et al. (1976). Nous en donnons une présentation nouvelle en section III.1.1, à l'aide du formalisme tensoriel qui permet de caractériser cette approche comme un problème de régression linéaire particulier dont nous étudions les liens avec la méthodologie de l'étude des courbes de croissance (ou growth curves) (cf. GEISSER 1980). Ce formalisme nous permet de simplifier de nombreuses démonstrations et de mieux situer le statut statistique des méthodes étudiées. De plus nous formulons nos résultats dans un cadre Euclidien général, supposant la donnée de métriques quelconques. Ce parti pris est nouveau dans ce domaine de recherche et ouvre la voie à l'introduction de modélisations probabilistes.

L'Analyse en Coordonnées Principales sur Variables Instrumentales, exprimée en termes d'opérateurs dans la section III.1.2, ne prend en compte qu'une sous classe de ces contraintes et fournit une méthodologie d'analyse analogue à l'Analyse en Composantes Principales sur Variables Instrumentales (cf. RAO (1964)). Elle permet cependant d'exprimer en termes de dissimilarités Euclidiennes toutes les méthodes d'Analyse Linéaires des Données de profil, cf. par exemple GIFI (1980) et MEULMAN (1986).

Nous étendons de plus notre étude à des contraintes liées aux propriétés d'invariance dans les espaces Euclidiens en III.1.3. Nous montrons que la méthode des rotations procrustes (cf. par exemple SIBSON (1980)), étendue par la prise en compte de métriques non triviales s'applique aussi au cas de données de dissimilarité permettant ainsi de traduire dans ce cadre d'étude, des hypothèses de travail sous forme de configurations de référence. Notre solution utilise cependant la transformation de Torgerson et n'est donc pas applicable aux dissimilarités non Euclidiennes.

Le cas des contraintes non linéaires, en particulier ordinales portant sur la configuration  $X$  est ensuite étudié (section III.2). Leur généralité en interdit le traitement par Analyse en Coordonnées Principales. Nous justifions alors l'approche heuristique développée par NOMA et JOHNSON (1979) dans le cadre des méthodes de Codage Multidimensionnel. Nous montrons en effet que les contraintes définies par des variables instrumentales forment une sous-classe de l'ensemble des contraintes qu'il est possible de traiter simplement par un algorithme de Codage Multidimensionnel tel que celui qui est proposé par de LEEUW et HEISER (1980).

La classe des contraintes portant sur les distances inter points présente un enrichissement de l'arsenal de méthodes d'Analyse des données, puisqu'elle fait référence à la structure d'espace métrique de l'espace de représentation. Le cadre le plus général d'Analyse des Données de Dissimilarité sous contraintes nous semble être fourni par l'étude des contraintes relationnelles, exprimées à l'aide d'une ou de plusieurs matrices de dissimilarité. Cette classe très générale de contraintes est tout d'abord étudiée au chapitre IV, dans le cadre de l'Analyse en Coordonnées Principales. Nous traitons en détail le cas d'une contrainte relationnelle décrite par un graphe, afin d'en montrer les liens avec certaines pratiques plus classiques en l'analyse linéaire des données de profil. Ces contraintes de graphe sont associées à une relation binaire, non valuée et sont étudiées en section IV.1.1. La section IV.1.2 considère alors l'extension de l'approche proposée au cas d'une relation valuée. Le cas de relations floues est d'abord traité, puis nous présentons le cas général, dans lequel la valuation du graphe des contraintes est interprétée comme un outil de description de contraintes métriques, c'est à dire portant sur la structure topologique de la surface réponse. Les méthodes de Codage Multidimensionnel permettent de traiter des contraintes plus générales et font l'objet de la section IV.2. Ces méthodes intègrent par exemple les problèmes de Codage Multidimensionnel Ordinal, sous des contraintes déterminées par une ou plusieurs matrices de dissimilarité, considérées du point de vue numérique puis du point de vue ordinal. Nous résolvons ce problème dans le cas où chaque matrice de contraintes définit dans l'ensemble des matrices de configuration  $X$ , un sous-ensemble de contraintes caractérisé comme un convexe fermé pour la topologie usuelle de l'espace ambiant. La méthode proposée est alors caractérisée en termes d'analyse canonique généralisée entre convexes fermés et nous proposons des algorithmes résolutoires dont les propriétés de convergence sont connues. L'extension de ces résultats au cas de contraintes définissant des ensembles proximaux est immédiate. Nous n'avons cependant pas recherché la plus grande généralité possible à ce propos, dans un souci d'adéquation de notre étude au cadre actuel d'application des méthodes étudiées et de faisabilité numérique.

La partie C de ce mémoire est consacrée aux AIDES À L'INTERPRÉTATION EN ANALYSE DES TABLEAUX DE DISSIMILARITÉ. L'absence relative d'outils d'aide à l'interprétation et d'évaluation globale de la solution fournie par un algorithme particulier, nous semble en effet constituer l'un des freins essentiels à l'utilisation des méthodes d'analyse des tableaux de dissimilarités. Par comparaison, en analyse de données de profil ou en classification automatique, les pratiques actuelles s'appuient sur un ensemble d'outils permettant une exploitation raisonnée des résultats d'une analyse. Ceux-ci nous semblaient insuffisamment développés dans le cas des méthodes de Codage Multidimensionnel.

Ceci nous semble expliquer le fait qu'en France ces méthodes se soient heurtées à un scepticisme certain. Nous étudions en conséquence ce problème dans le texte de notre exposé aux Journées Analyse des Données et Informatique, Versailles (1987) et nous proposons un ensemble d'outils destinés à évaluer la qualité globale de la solution, à rechercher les points et les paires de points qui jouent un rôle particulier dans l'élaboration de la solution. L'interprétation du problème en termes de graphes fournit alors de puissants outils de dépouillement des résultats d'analyse. Enfin la technique des points supplémentaires est sensiblement étendue pour les deux classes de problèmes, grâce encore aux informations fournies par le schéma algébrique cité plus haut. Ces outils ont pour but de compléter les méthodes d'analyse de la stabilité des résultats proposés par l'école Hollandaise (voir en particulier HEISER (1981) et MEULMAN (1986)), qui met à profit des méthodes de re-échantillonnage, en particulier de bootstrap, et qui ne sont pas étudiées ici.

L'élaboration d'outils d'aide à l'interprétation est traitée de façon distincte en Analyse en Coordonnées Principales et en Codage Multidimensionnel, puisque les deux classes de méthodes optimisent des critères distincts. Tout d'abord nous montrons que les aides à l'interprétation développées en Analyse en Composantes Principales ne sont utilisables en Analyse en Coordonnées Principales que dans le cas de dissimilarités Euclidiennes. Dans ce cas, la transformation  $Y = TZ$  fournit pour les arêtes du graphe des outils comparables à ceux dont on dispose pour les unités statistiques (sommets du graphe). Leur utilisation permet alors d'évaluer la contribution de chaque axe ou de chaque sous-espace à la reconstitution d'une dissimilarité particulière, ainsi que la contribution de chaque dissimilarité à la détermination de chaque axe puisque ces aides à l'interprétation sont calculées en fonction des normes des vecteurs représentatifs et que l'on a

$$\delta_{ij}(Z) = \|z_i - z_j\| = \|y_i\|$$

Dans le cas d'une dissimilarité quelconque, l'existence d'axes complexes, même s'ils sont ignorés lors de la construction d'un résumé de dimension  $r$ , introduisent des effets de compensation dans l'évaluation des normes qui interviennent dans le calcul de tous les indices proposés dans le cadre de l'Analyse en Composantes Principales. L'analyse spectrale de l'opérateur  $WNWN$  de préférence à celle de  $WN$  qui fonde l'Analyse en Coordonnées Principales, offre une solution au problème de détermination de la dimension du sous-espace résumé. Cependant, l'évaluation de la qualité de la représentation des unités statistiques n'est plus possible, puisque l'emboîtement des sous-espaces propres de cette matrice n'implique plus le propriété d'approximations par défaut des dissimilarités.

Les méthodes de Codage Multidimensionnel fournissent elles aussi des approximations bilatérales des dissimilarités et laissent donc peu d'espoir de dégager des indices de qualité

de représentation d'un point. L'évaluation globale d'une solution construite par un algorithme de Codage Multidimensionnel s'appuie en général sur une évaluation de la fonction critère et une consultation du diagramme de Shepard.

Nous montrons de plus que la pente de la droite ajustée au diagramme de Shepard fournit un indicateur du biais résultant d'une sous-estimation de la dimension de l'espace de représentation. Nous préconisons enfin le recours aux mesures d'association locale proposée par FRIEDMAN et RAFSKI (1983) pour évaluer la qualité de reconstitution des propriétés locales de la relation induite par les données, à l'aide de la configuration reconstruite.

Nous proposons donc de nouveaux indices qui décrivent les contributions de chaque sommet  $i$  et de chaque arête  $(i,j)$  du graphe sous-jacent, à la valeur de la fonction d'écart à l'optimum. On peut d'autre part interpréter ces indices en termes de décomposition additive d'une quantité d'information, en appliquant les résultats exposés en II-4.

La partie D de ce travail est consacrée à l'étude de l'application des méthodes d'ANALYSE DES DONNÉES DE DISSIMILARITÉ observées DANS LE CADRE D'UNE EXPÉRIENCE PLANIFIÉE. Rédigée à l'occasion d'une conférence invitée à la XVI-th International Biométrie Conference (Namur 1988), elle s'adresse plus particulièrement aux spécialistes de l'écologie quantitative ou de biologie. On parle alors de Méthodes d'Ordination pour désigner les méthodes désormais classiques en biométrie (GOWER (1984)) d'analyse des tableaux d'incidence ou d'abondance et plus généralement de mesures de diversité interne des populations et de dissimilarité (ou de similarité) entre populations

Considérons par exemple l'information caractérisant la performance d'espèces végétales, la composition ou l'équilibre écologique d'un éco-système. Elle est souvent communiquée sous la forme d'un tableau  $\Delta = (\delta_{ij}; i, j = 1, 2, \dots, n)$  mesurant le degré de dissemblance entre espèces. Dans ce cas les méthodes de Codage Multidimensionnel sont en général utilisées pour explorer les structures non linéaires de la surface réponse décrite en termes topologiques par la variable réponse  $\Delta$ . L'analyste consulte les graphiques obtenus dans l'espoir de dégager une information sous jacente à l'établissement des mesures de dissimilarité. L'arbitraire qui résulte du choix d'une dimension de représentation, dicté par un impératif mathématique de minimisation du critère d'approximation de  $\Delta$  par un tableau de distances, conduit à la construction de représentations résumée non nécessairement optimales ou même pertinentes du point de vue sémantique. Les problèmes méthodologiques posés par une telle démarche sont connus depuis longtemps, comme en atteste la critique de BEALS (1973) et ne sont pas discutés ici. Nous avons retenu de cette critique la nécessité de définir un cadre inférentiel adapté permettant de valider les résultats par une étude de liaisons potentielles, a priori non linéaires et/ou discontinues avec ce qui

est appelé dans ce domaine de recherche des gradients d'influence de variables d'environnement.

Nous proposons donc des méthodes statistiques d'analyse des données de dissimilarité expérimentales. On dispose alors d'une matrice de dissimilarité et de la mesure d'un ensemble de variables facteurs définissant la structure du plan d'expérience. Chaque facteur ou combinaison de facteurs induit alors une partition des éléments de I et fournit une décomposition en blocs des éléments de la matrice de dissimilarités observée. Nous nous plaçons dans un cadre inférentiel et nous montrons la cohérence de notre démarche avec la méthodologie classique de l'Analyse de La Variance. Nous espérons ainsi motiver un intérêt nouveau pour l'application des techniques de Codage Multidimensionnel aux problèmes d'analyse linéaire ou non des Plans d'Expérience et des Surfaces Réponses. Notre approche est fondée sur une interprétation statistique nouvelle des composantes des méthodes de Codage Multidimensionnel qui résulte directement du formalisme algébrique mis en place dans les parties A et B, ainsi que sur la démonstration faite de la complémentarité des outils exploratoires présentés à la méthodologie de test que nous proposons.

Notre approche utilise le constat, fait par Rao (1982), que dissimilarité et entropie sont en fait deux facettes du traitement statistique de la notion de diversité entre individus de la population étudiée. Le premier point de vue conduit en pratique à choisir une fonction de dissimilarité à cause de sa signification biologique afin de refléter une notion intrinsèque de dissimilarité entre individus pertinente pour les objectifs de l'étude. Ce choix induit la définition implicite, mise en lumière ici, d'une fonction d'entropie qui est le plus souvent négligée ou ignorée. Le second point de vue considère directement une mesure d'entropie, regardée comme une fonction définie sur l'espace des fonctions de répartition définies sur la population étudiée dans le but de comparer leur écart à la distribution uniforme ou bien leurs différences par référence à la distribution uniforme. Cette entropie doit satisfaire à un ensemble classique d'axiomes qui ne font pas référence explicitement aux individus constituant la population d'intérêt et ignore leurs dissimilarités. Ce point de vue est illustré par la généralisation du modèle linéaire aux structures exponentielles fondé sur une mesure de déviance et le recours à des méthodes d'estimation au sens de la quasi vraisemblance, étudiée par exemple par McCULLAGH and NELDER (1983). Dans ce cas la détermination d'une fonction de dissimilarité associée à la déviance est le plus souvent très difficile.

Tout d'abord, nous interprétons en Section 1 l'Analyse de la Variance à un facteur en termes de dissimilarité, afin d'illustrer la méthode proposée dans le cas particulier bien connu de l'analyse de la variance univariée, qui reçoit une description complète à l'aide de la matrice des dissimilarités Euclidiennes  $\delta_{ij} = \sqrt{(y_i - y_j)^2}$  calculées à partir de la variable réponse Y. On en déduit de façon élémentaire les décompositions de la variance

utilisées dans l'élaboration des statistiques usuelles de tests. L'extension de cette formulation à la classe des fonctions de dissimilarité qui engendrent une fonction d'entropie quadratique (Rao (1982)) est traitée en Section 2. Les problèmes inférentiels posés dans ce cadre par l'analyse de l'influence des facteurs définissant le plan d'expérience, ne sont pas classiques puisque la variable réponse mesure ici la dissimilarité entre unités statistiques, et la définition d'une structure statistique précisant la loi de probabilité de la variable réponse reste un problème ouvert. De plus les résidus  $\epsilon_{ij}$  sont généralement des variables associées donc non indépendantes en loi.

Nous proposons donc d'utiliser les décompositions de la quantité d'information définie par le triple  $(I, \delta, Q_{I \times I})$  dans le cadre d'une inférence de plan. Les unités statistiques ne sont pas alors interprétées comme des éléments aléatoires issus d'une procédure d'échantillonnage et les résultats des tests portent alors sur un effet observé dans la population observée sans référence à une super population. On recourt alors à une méthodologie de tests de permutation (ou "randomization tests", cf. EDGINGTON (1980)) afin d'évaluer le degré de signification des hypothèses mises à l'épreuve. Nous utilisons pour ce faire une procédure proposée par HUBERT (1978) et nous en interprétons les termes dans  $F$  et dans  $F \wedge F$ .

Nous reprenons alors les résultats de RAO (1982, 1984) afin de caractériser la classe des mesures de dissimilarité qui permettent de généraliser la méthodologie de l'analyse de la variance au cas de variables réponse mesurant les différences entre unités statistiques. De telles dissimilarités fournissent une décomposition additive d'une quantité d'information définie à partir de  $(\delta_I, P_I)$  ou de  $(\delta_I, Q_{I \times I})$ . Nous montrons de plus que les dissimilarités Euclidiennes jouent là encore un rôle privilégié du fait de leurs propriétés de convexité. Le modèle d'erreur sous-jacent aux méthodes de Codage Multidimensionnel

$$\delta_{ij} = d_{ij}(X) + \epsilon_{ij}$$

se prête alors simplement à la description des propriétés métriques de la surface réponse et l'on peut interpréter cette équation de façon classique à l'aide d'un développement en série de Taylor. Le degré de régularité imposé aux fonctions de dissimilarité afin de permettre la mise en œuvre de la méthode dépend directement de la complexité du plan d'expérience traité

En Section 3, nous interprétons cette méthodologie pour décrire le potentiel d'interaction évalué par la donnée d'un tableau à double entrée symétrique. La notion de contraste  $2 \times 2$  joue un rôle prééminent dans cette interprétation et fournit une interprétation statistique des variables latentes associées à chaque axe de la configuration construite par Analyse en Coordonnées Principales en termes de contrastes optimaux. De plus, nous montrons en section II-4, que la complémentarité entre l'Analyse Multidimensionnelle de la Variance et l'Analyse Discriminante Linéaire, si importante pour l'étude des données de profil, est

préservée lorsque l'on traite des données de dissimilarité. Nous en décrivons les termes en détail. et une extension de notre approche, intégrant des variables instrumentales au cadre d'étude de l'Analyse des Tableaux de Dissimilarité, est brièvement évoquée en section II-4, par référence à l'étude complète qui en est faite dans un cadre exploratoire, dans la partie B. Ces variables externes résultent classiquement soit de connaissances a priori sur le problème traité, comme dans le cas de la description d'un plan d'expérience, soit de la mesure de variables numériques ou qualitatives d'environnement, par exemple des variables pédologiques en végétation.

**PARTIE A**

**UN MODELE ALGEBRIQUE**

**POUR**

**L'ANALYSE DES DONNEES DE DISSIMILARITE**

**SANS CONTRAINTES**

**Chapitre I INTRODUCTION**

**Chapitre II PROBLEMES D'EXISTENCE D'UN PLONGEMENT  
EUCLIDIEN**

**Chapitre III STRUCTURES AFFINES ET ANALYSE D'UNE  
DISSIMILARITE EUCLIDIENNE**

**Chapitre IV STRUCTURES LINEAIRES ET ANALYSE  
DU TRIPLE  $(I, \delta, P_I)$**

**Chapitre V LES METHODES D'ANALYSE DU TRIPLE  $(I, \delta, P_I)$**

**Chapitre VI L'ANALYSE DU TRIPLE  $(I, \delta, Q_{|x|})$**

**Chapitre VII ANALYSE ORDINALE DU TRIPLE  $(I, \delta, Q_{|x|})$**



## CHAPITRE I

### INTRODUCTION

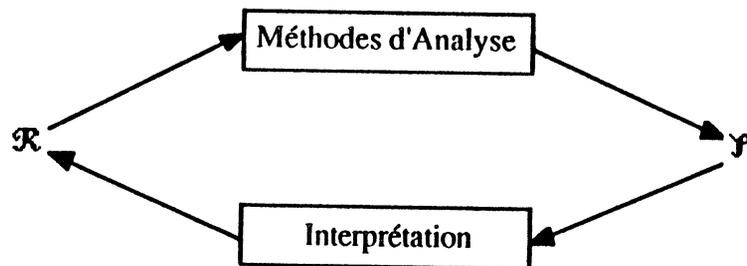
La pratique de l'analyse multidimensionnelle d'un tableau de données fonde l'essentiel de son attrait et de son efficacité sur l'emploi de représentations graphiques, dont la lecture obéit aux règles de la géométrie.

Du point de vue formel, les méthodes d'analyse des données apparaissent comme des outils destinés à associer à un système relationnel empirique  $\mathcal{R} = \{R^j, j \in J\}$  où  $R^j$  est une relation  $k^j$ -aire définie sur un ensemble  $I$ , un système relationnel métrique  $\mathcal{Y} = \{S^j, j \in J\}$  où la relation  $S^j$  est à valeurs dans un espace métrique de dimension fixée par l'analyste. Le plus souvent, une géométrie Euclidienne est retenue. Ce choix présente l'avantage de la simplicité, bien que de nombreuses techniques aient été développées pour construire des représentations dans des espaces métriques de structure plus complexe. Ainsi ATTNEAVE (1950) justifie t'il de façon empirique le recours à une métrique associée à la fonction  $l \times l^p$  pour traiter ses données. Lorsque  $p=2$ , on retrouve le carré d'une norme Euclidienne et l'attention de nombreux auteurs a été retenue par le cas  $p=1$ . On parle alors souvent de métrique  $l_n^1$  ou encore de city-bloc metric. Par exemple, RESTLE (1959) tente de caractériser les mesures de similarité induisant une structure métrique de type  $l_n^1$  sur  $I$ . L'emploi de géométries de Minkowski  $l_n^p$  plus générales a été souvent préconisé, en particulier en psychophysique. Une étude axiomatique des propriétés qualitatives auxquelles doit satisfaire une fonction de dissimilarité pour justifier l'existence de telles représentations est menée par TVERSKY et KRANTZ (1970). Ces auteurs étendent leur étude à la classe des structures métriques caractérisées par une fonction de distance dont l'expression analytique combine de façon additive certaines fonctions des différences entre coordonnées.

L'intérêt pratique de ces études se heurte cependant à l'absence d'outils permettant d'évaluer le degré de non respect de ces axiomes sur un jeu de données particulier. D'autre part, le traitement de problèmes concrets a éveillé l'intérêt des psychologues pour des géométries Riemaniennes à courbure constante ou non (PIESZKO (1975), LINDMAN et CAELLI (1978)). Le même cadre d'applications en psychologie perceptive conduit de plus à étendre l'étude des propriétés projectives des géométries Euclidiennes au cadre des métriques de Cayley-Klein (DROSLER (1981a)), adaptées en particulier à

l'étude de certains problèmes spécifiques tels que l'étude de la perception visuelle binoculaire (cf. LUNEBURG (1947, 1950), INDOVS (1975,1979), DROSLER (1981a)) faisant appel à des géométries hyperboliques.

La lecture de graphiques résultant de telles analyses suppose une compétence qui nous semble partagée par bien peu d'utilisateurs potentiels. Pourtant, le bon sens voudrait que l'apport du système relationnel métrique, à des fins d'analyse ou de communication, reposât sur un contrôle réel des règles de lecture des graphiques. Celles-ci fondent en effet l'interprétation des résultats de l'analyse, c'est à dire de traduction dans le cadre du système  $\mathfrak{R}$ , des relations décrivant les résultats de l'analyse selon le schéma suivant :



L'étude des règles de lecture métrique des résultats de l'analyse dans le cas le plus simple, fournis par les représentations Euclidiennes, constitue l'objet principal de ce mémoire. Constatons en effet après GOWER (1983), que même les représentations Euclidiennes bidimensionnelles sont moins simples qu'il ne peut paraître. Considérons en effet deux points  $P_1$  et  $P_2$  représentant dans le plan Euclidien deux unités statistiques dont l'écart observé est  $\delta_{12}$ . Notons alors  $d_1(X)$  et  $d_2(X)$  les distances respectives de  $P_1$  et  $P_2$  à l'origine  $O$ . Ainsi

$$d_1^2(X) = \sum_{s=1}^2 [x_1^s]^2$$

Supposons de plus qu'un tel graphique est issu d'une méthode linéaire, telle que l'Analyse en Composantes Principales. Il exprime la relation d'approximation

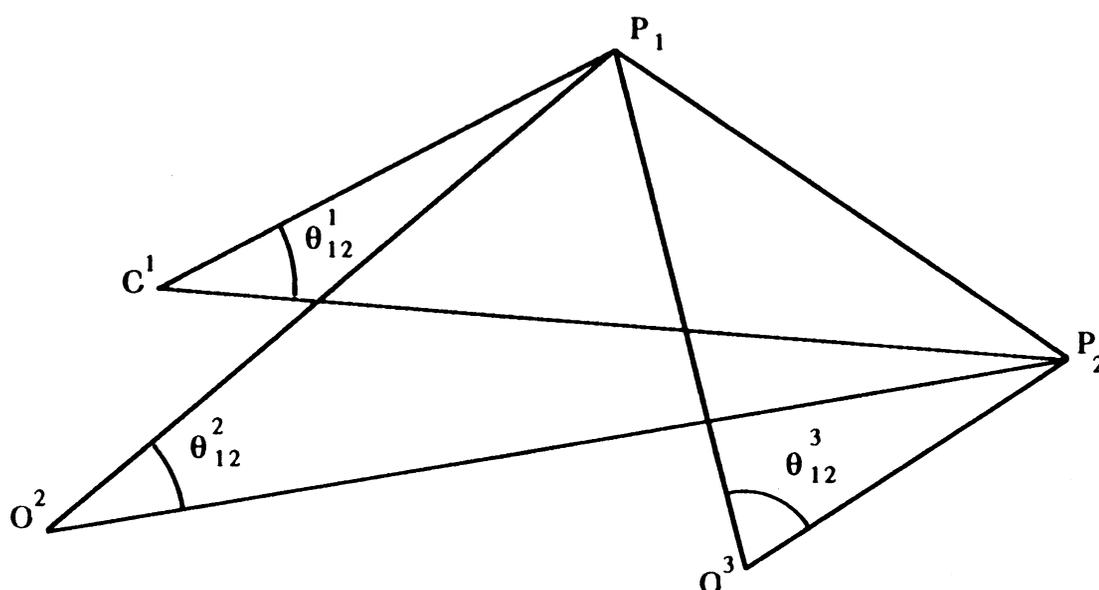
$$\delta_{12} \approx d_1(X)d_2(X)\cos(\theta_{12}) \quad (1)$$

où  $\delta_{12}$  mesure un degré de similarité et  $\theta_{12}$  est l'angle formé par les vecteurs  $\overrightarrow{OP_1}$  et  $\overrightarrow{OP_2}$ . Cette relation utilise les propriétés du produit scalaire associé à la géométrie Euclidienne ambiante. Elle utilise la structure hilbertienne de l'espace vectoriel Euclidien dans lequel est décrit le problème. En particulier, elle impose le choix d'une origine qui détermine les valeurs de  $d_1(X)$ ,  $d_2(X)$  et  $\theta_{12}$ .

Le même graphique, s'il est issu d'une méthode de Codage Multidimensionnel, exprime la relation d'approximation

$$\delta_{12} \approx [d_1^2(X) + d_2^2(X) - 2d_1(X)d_2(X)\cos(\theta_{12})]^{1/2} \quad (2)$$

c'est à dire que  $\delta_{12}$  est approchée par la distance Euclidienne  $d_{12}(X)$  dont l'évaluation repose sur les propriétés métriques de l'espace affine Euclidien ambiant. Elle est en particulier indépendante du choix d'une origine et d'une base de coordonnées. Dans les deux cas, nous lisons sur le graphique une distance Euclidienne  $d_{12}(X)$ . Mais dans le premier cas,  $d_{12}(X)$  a peu de rapport avec  $\delta_{12}$ , puisque la même valeur de la distance peut être associée avec une infinité d'angles  $\theta_{12}^k$  ( $k=1,2,3$  sont représentés ci-dessous), déterminés par autant de choix de l'origine  $O^k$ . GREENACRE (1979) utilise dans ce cas l'euphémisme d'"approximation faible" des distances.



Les méthodes d'analyse des données peuvent donc être distinguées selon qu'elles conduisent à des approximations faibles ou fortes des distances interpoints. Dans le premier cas, toute lecture distancée des graphiques obtenus devrait être abandonnée compte tenu des risques encourus d'erreurs et d'interprétations non fondées de l'information portée par les données, en particulier lorsque l'on doit recourir à une représentation résumée des données dans un espace de faible dimension. Dans le cas de données de profil, le recours à une lecture distancée ne trouve d'ailleurs aucune justification stricte en termes d'optimalité et repose le plus souvent sur une connaissance insuffisante des fondements algébriques des méthodes. Cette situation peut de plus paraître paradoxale par comparaison avec la richesse des outils (règles de lecture, aides à l'interprétation) qui résultent des relations de dualité entre unités statistiques  $\{\omega_i, i \in I\}$  et variables  $\{X^j, j \in J\}$ , induites par le tableau de données  $X = \{x_i^j = X^j(\omega_i); i \in I, j \in J\}$ .

L'analyse des tableaux de données de dissimilarité est plus complexe de ce point de vue puisqu'un grand nombre d'outils algébriques disparaissent avec la donnée a priori de variables explicites et la perte induite de relations de dualité. On dispose alors d'un couple

$(I, \delta)$  formé d'un ensemble fini  $I$  d'unités statistiques et d'une fonction réelle positive  $\delta$ , définie sur  $I \times I$ , mesurant le degré de dissemblance (d'écart) entre les unités statistiques. On se propose alors de construire un plongement métrique (le plus souvent Euclidien) de  $I$  dans un espace métrique  $(\mathcal{X}, \rho)$  de dimension fixée à priori, c'est à dire une application

$$\begin{aligned} \varphi : I &\rightarrow (\mathcal{X}, \rho) \\ i &\mapsto \varphi(i) = x_i \in \mathcal{X} \end{aligned}$$

telle que

$$\delta_{ij} = \rho(\varphi(i), \varphi(j)) = \rho(x_i, x_j)$$

de LEEUW et HEISER (1982) attribuent les premiers travaux sur les données de dissimilarité à STUMPF (1880). Cet auteur réfléchit à la mesure du degré de dissimilarité, fit la relation avec la notion de distance (psychologique) et discuta des propriétés de fidélité des mesures de dissimilarité. Il ne proposa cependant pas de méthode de Codage Multidimensionnel, c'est à dire de construction d'un plongement métrique de  $I$ . Selon de LEEUW et HEISER (1982), les premiers algorithmes furent proposés par FISHER (1922) pour résoudre un problème de représentation des chromosomes à partir de l'information portée par leur fréquence de recouvrement et ces auteurs citent Kruskal qui fait référence aux travaux de Boyden vers 1930 en zoologie.

Bien que tardives par rapport aux travaux de PEARSON (1901) à l'origine de l'Analyse en Composantes Principales et aux premiers travaux en Analyse Factorielle Classique, les origines des méthodes modernes de Codage Multidimensionnel doivent être recherchées dans les travaux des factorialistes en psychométrie, en particulier de l'école de THURSTONE. Les écrits les plus marquants sont dus à RICHARDSON (1938) et à KLINGSBERG (1941) qui utilisèrent les constituants essentiels de la pratique actuelle de l'Analyse en Coordonnées Principales, à savoir d'une part l'estimation d'une constante additive destinée à associer à un indice de "distance psychologique" une fonction de dissimilarité vérifiant les axiomes classiques d'une distance, et d'autre part la construction d'une représentation des éléments de  $I$  à l'aide des résultats de YOUNG (1938). Cette approche du problème culmina avec la thèse de TORGERSON (1951) et surtout la publication de son livre en 1958.

Les premiers travaux mathématiques directement utilisables dans une optique d'Analyse des Données portent sur le plongement des espaces métriques dans un espace distancié. Ils sont dus à FRECHET (1935) et complétés par SCHOENBERG (1935, 1937, 1938) ils fournirent les premiers résultats d'existence d'une solution. On peut cependant comme de LEEUW et HEISER (1982) ou BARTHELEMY et GUENOCHÉ (1988), faire référence aux travaux de GAUSS (1831) qui caractérisa les distances Euclidiennes dans le cas d'un espace tridimensionnel à l'aide des propriétés de définie positivité de la forme

quadratique représentée par la matrice  $W^n$  étudiée plus bas. Cette caractérisation fut plus complètement établie par DIRICHLET (1850) et HERMITE (1850) étendit ce résultat au cas général, c'est à dire non limité à la dimension trois. Ses travaux furent repris par MINKOWSKI (1891), mais c'est FRECHET (1935) qui montra l'intérêt de ces résultats pour l'étude de la géométrie des espaces distanciés.

Ces caractérisations sont comparées ici aux travaux fondamentaux de CAYLEY (1841) repris et développés par MENGER (1928). A l'opposé de Fréchet et de Schoenberg qui utilisent les propriétés de formes quadratiques particulières, ces auteurs utilisent les propriétés affines de l'espace de représentation et utilisent les propriétés des déterminants pour étudier l'existence d'un plongement Euclidien d'un espace semi-métrique. Bien que d'un intérêt moindre pour le traitement des données, les outils utilisés par ces auteurs permettent de traiter le problème plus difficile d'existence d'une congruence entre un espace semi-métrique et un espace Euclidien. Un certain nombre de résultats sur ce sujet sont présentés ici, qui s'appuient en particulier sur les travaux de synthèse publiés par BLUMENTHAL (1938, 1953) et BLUMENTHAL et MENGER (1970).

Nous montrons plus bas que l'ensemble des méthodes fondées sur les résultats de YOUNG et HOUSEHOLDER (1938), qui fondent la tradition psychométrique, privilégient la qualité de l'approximation des angles entre vecteurs et conduisent à des approximations faibles (cf. infra pour une définition précise de ce terme) des dissimilarités observées, lorsqu'on se propose de construire une représentation résumée de  $I$  dans un souci de réduction de dimension. La recherche d'approximations fortes des dissimilarités fit l'objet de publications plus tardives. En effet, bien que contemporaine des travaux de Torgerson, la méthode de quantification IV proposée par HAYASHI (1952), ne semble pas avoir retenu l'attention de l'école psychométrique américaine. Il fallut donc attendre les travaux de SHEPARD (1957, 1958a, b) et surtout la publication d'une heuristique par cet auteur (SHEPARD (1962a, b)) pour observer un intérêt réel pour ce problème. Shepard parle de Codage Multidimensionnel (Multidimensional Scaling en anglais) et nous retenons ce terme de préférence aux intitulés d'Analyse de Proximités proposé par BENZECRI (1965) et de Positionnement Multidimensionnel proposé par ESCOUFIER (1975).

KRUSKAL (1964a, b) formula le problème en termes d'ajustement au sens des moindres carrés et le résolut à l'aide d'un algorithme de gradient gérant de façon sophistiquée et très efficace le pas de l'itération. Cette date marqua en fait la naissance d'une véritable alternative à l'Analyse en Coordonnées Principales, nom donné par GOWER (1966) à la méthode proposée par TORGERSON (1952). De très nombreux travaux suivirent, qui sont cités en référence au cours de notre exposé. Le problème posé par le Codage Multidimensionnel est en effet de nature très différente de l'Analyse en Coordonnées

Principales. Le recours à une méthodologie d'approximation directe des dissimilarités par des distances est un changement de perspective très important qui oblige à une lecture prudente des théorèmes consacrés au plongement Euclidien d'un espace semi-métrique (cf. CRITCHLEY (1986)). L'importance de la notion de dimension de l'espace de représentation doit en effet être relativisée lorsqu'on s'intéresse à l'approximation des dissimilarités elles-mêmes. La dispersion (ou l'inertie) des nuages projetés sur chaque axe de coordonnées de l'espace de représentation détermine en effet les propriétés de séparation des valeurs propres de l'opérateur soumis à décomposition spectrale en Analyse en Coordonnées Principales, et pas uniquement leur nullité éventuelle. BENZECRI (1965) complété par GAUD et FICHET (1984) et CRITCHLEY (1986) illustrent de façon détaillée les conséquences importantes de cette particularité du problème sur les problèmes de détermination d'un plongement Euclidien approché à  $\epsilon$  près, en dimension donnée. En fait, on retrouve ainsi en Codage Multidimensionnel, une difficulté reconnue des méthodes d'Analyse Factorielle Classique, analysée en détail par HAKIM et al. (1973 a, b). Ces auteurs montrent qu'elle résulte de la structure particulière de l'ensemble des matrices semi-définies positives, mise en évidence lors de l'étude de sa décomposition en sous-ensembles de matrices semi-définies positives de rang donné.

La différence de nature des problèmes traités par les deux écoles de pensée explique sans doute que peu de travaux aient été consacrés à l'élaboration d'un formalisme unificateur susceptible de rendre compte d'un point de vue théorique, des liens et des particularités de chaque classe de problème. Un certain nombre de comparaisons numériques de méthodes a été publié, en particulier au Japon. Ainsi, par exemple, SAITO et al. (1972), OKAMOTO et TODA (1973), HAYASHI (1973), SAITO (1982) comparent l'approche de Hayashi à l'Analyse en Coordonnées Principales du point de vue de la reconstruction des données et TAKANE (1977) fournit une première comparaison des matrices impliquées dans la description de quatre algorithmes publiés. Nous en interprétons les termes dans le cadre du formalisme unificateur proposé dans ce mémoire.

Après un rappel des définitions et résultats portant sur les notions de dissimilarité, les préordonnances, et les espaces affines Euclidiens, nécessaires à une présentation précise des méthodes d'analyse des données de dissimilarité, le chapitre II est consacré à l'exposé des théorèmes d'existence d'un plongement Euclidien d'un espace semi-métrique dans un espace affine de dimension donnée. Nous présentons tout d'abord les travaux issus des recherches de Fréchet, Schoenberg et Menger, puis nous rappelons brièvement les difficultés rencontrées dans l'établissement et la validation empirique de résultats portant sur l'existence d'une congruence entre un espace (semi-) métrique et un espace Euclidien.

Le chapitre III est consacré à l'étude des structures affines intervenant lors de l'analyse d'une fonction de dissimilarité Euclidienne. d'AUBIGNY (1975) indique que l'on traite

en fait un problème de géométrie affine, et les relations entre cette formulation et l'écriture classique de Torgerson dans l'espace vectoriel sous-jacent ont été détaillées par FICHET (1983) dont nous reprenons et complétons les résultats nécessaires à notre exposé. Nous montrons alors que le cadre affine est nécessaire à l'explicitation des relations existant entre les travaux d'origine française d'une part, qui insistent sur le rôle de la pondération des unités statistiques et le choix induit de métrique et les travaux d'origine anglaise d'autre part, en particulier dus à GOWER (1966, 1982 a, b, 1984, 1985), qui portent sur la détermination d'une origine de l'espace de représentation.

Le chapitre IV suppose ces problèmes résolus et traite tout d'abord le problème d'emboîtement des espaces de représentation des données. Nous rappelons tout d'abord les outils algébriques mis en œuvre pour l'analyse linéaire d'un tableau de données de profils. Puis nous exposons une formulation tensorielle de l'Analyse en Composantes Principales qui nous semble utile pour étudier les conditions mathématiques qui fondent la propriété recherchée d'emboîtement des sous-espaces de représentation résumée :

$$E_r \subset E_{r'} \text{ si } \dim(E_r) < \dim(E_{r'})$$

Nous montrons alors que le cadre tensoriel permet une analyse fine des opérateurs utilisés sous forme matricielle dans le traitement classique de L'Analyse en Coordonnées Principales et nous montrons les liens de la méthode avec les théorèmes d'existence. Cet exposé intègre les travaux de CRITCHLEY (1985 a, b, c) que nous complétons par l'introduction d'une dualité algébrique entre espaces de description des unités statistiques et des paires d'unités statistiques, qui joue selon nous un rôle essentiel dans la compréhension des méthodes d'analyse des tableaux de dissimilarité, en termes d'Analyse en Coordonnées Principales, de Codage Multidimensionnel (métrique ou ordinal), de Classification Hiérarchique et d'Arbres Additifs, des liens qui unissent ces méthodes et de la spécificité de chacune d'elles.

La démarche classique de l'Analyse du Triple  $(I, \delta, P_I)$  où  $P_I$  est une pondération des unités statistiques, opère par analogie avec la méthodologie de l'Analyse en Composantes Principales. Nous en étudions les termes au chapitre V. Trois objectifs sont poursuivis dans cet exposé. Le cas d'une dissimilarité Euclidienne est tout d'abord étudié. Nous montrons à ce propos que l'on peut étendre de façon simple l'arsenal existant de méthodes d'Analyse du Triple  $(I, \delta, P_I)$ , à l'aide de modèles descriptifs établis par analogie avec le problème traité en Analyse Factorielle Classique. Nous adoptons ici la distinction fondamentale introduite par McDONALD (1968) entre les méthodes tautologiques, telles que l'Analyse en Coordonnées Principales, qui ont pour but de construire des résumés dans une optique de réduction de dimension, et les méthodes modélisantes auxquelles nous réservons le nom d'Analyse Factorielle du triple  $(I, \delta, P_I)$ . Nous montrons alors comment résoudre le problème posé par l'ajustement de tels modèles, puis nous accordons une

attention particulière à une méthode d'analyse qui adapte au cadre des données de dissimilarité la notion d'Analyse d'Images Partielles développée par GUTTMAN (1953) pour l'analyse des données de profil. C'est une analyse en Coordonnées et donc directement concurrente de l'Analyse en Coordonnées Principales. Nous montrons de plus que les résultats de GOWER (1982) permettent de fixer une origine de l'espace de représentation cohérente avec notre approche.

La deuxième partie de ce chapitre est consacrée à l'étude des liens existant entre l'Analyse en Coordonnées Principales et la notion statistique de contrastes  $2 \times 2$ . C'est là une conséquence de la dualité entre unités statistiques et paires définie dans ce mémoire, qui nous paraît utile pour définir un statut statistique à cette méthode, et donner les règles de lecture des résultats qui lui faisaient défaut. Nous montrons de plus l'intérêt des contrastes  $2 \times 2$  pour d'autres problèmes d'analyse des tableaux de dissimilarité, tels que les arbres additifs. Nous démontrons enfin un théorème d'existence d'un plongement Euclidien d'un espace quasi-métrique  $(I, \delta)$  en terme de positivité des contrastes  $2 \times 2$ , qui complète les résultats de Schoenberg et montre en particulier les liens existant entre les méthodes d'analyse Euclidienne d'un tableau de dissimilarité et l'étude de l'association entre variables aléatoires à l'aide des indices utilisant la notion mathématique de birapport, appelés odds ratio et proposés par YULE (1912) pour l'étude des tables de contingence binaire.

La dernière partie de ce chapitre traite le cas des fonctions de dissimilarité non nécessairement Euclidiennes. Nous donnons une définition non classique de l'Analyse en Coordonnées Principales complétée par une interprétation nouvelle semble-t-il, des problèmes de constante additive en termes de perturbation, dont nous montrons l'analogie aux méthodes de Ridge Regression. Puis nous montrons que la généralisation de ces méthodes au cas d'une perturbation diagonale quelconque, analogue de la Régression Bayésienne, permet de retrouver l'Analyse Factorielle du Triple dans un cas, et les modèles d'Analyse de données de dissimilarité avec effets de contexte (ou de densité spatiale) dans un autre cas. Nous généralisons alors des travaux de CAILLEZ (1983) et proposons un algorithme résolutif, homogène à la logique de l'Analyse en Coordonnées Principales, et donc complémentaire des travaux de SAITO (1983, 1985, 1986) en Codage Multidimensionnel. Ce modèle nous permet de plus de proposer un algorithme permettant de construire une décomposition additive d'une dissimilarité en somme d'une distance Euclidienne et d'une distance à centre, dont l'existence a été prouvée par LE CALVE (1988) et qui est optimale au sens du critère adopté par Caillez.

Le chapitre VI est consacré aux méthodes d'Analyse du Triple  $(I, \delta, Q_{|I|})$ , où  $Q_{|I|}$  est une pondération des paires d'unités statistiques. Nous montrons que la prise en compte d'une pondération  $Q_{|I|}$  opérant sur  $|I|$  change la nature des problèmes rencontrés et nécessite

l'utilisation d'outils mathématiques particuliers. Nous établissons donc tout d'abord un cadre algébrique adapté à l'étude du triple  $(I, \delta, Q_{I \times I})$  que nous utilisons pour rendre compte des algorithmes de Codage Multidimensionnel et les comparer. Nous montrons en particulier que le traitement algébrique du cas d'une pondération produit, du type  $Q_{I \times I} = P_I \otimes P_I$  permet de retrouver les équations de l'Analyse en Coordonnées Principales. Cette formulation du Codage Multidimensionnel nous permet de façon plus importante à nos yeux de fournir une interprétation nouvelle du problème posé en termes d'énergie dissipée dans un réseau électrique, complémentaire de celle en termes plus classiques pour le statisticien, d'analyse approchée du potentiel d'interaction décrit par un tableau à double entrée, qui découle de l'utilisation des contrastes  $2 \times 2$ .

Le chapitre VII traite le problème complémentaire des transformations des données en Codage Multidimensionnel. Celles ci sont en général imposées par des contraintes de mesurage induites par le processus de collecte des données. On retrouve ainsi les origines des recherches de SHEPARD (1962), KRUSKAL (1964), BENZECRI (1965) et GUTTMAN (1968) qui traitent le problème du Codage Multidimensionnel dans un cadre ordinal et sont donc conduits à prendre en compte une transformation monotone des dissimilarités observées. Nous distinguons deux classes de critères d'approximation proposés dans la littérature pour analyser le cas ordinal. La première classe utilise les codages fonctionnels de l'indice de rang associé à une préordonnance observée. Nous montrons alors comment s'intègre cette logique dans la méthodologie du Codage Multidimensionnel et dans le cadre algébrique défini plus haut. La régression isotone est alors resituée dans ce cadre en termes de dualité Lagrangienne et la méthode alternative proposée par GUTTMAN (1968) en termes d'images de rangs reçoit ici une interprétation nouvelle en termes d'Analyse Canonique. La deuxième classe de critères est elle aussi connue et a fait l'objet de propositions d'algorithmes (cf. en particulier de LEEUW (1970)). Elle utilise les propriétés relationnelles de la préordonnance induite par la dissimilarité observée. Son analyse nécessite de compléter le formalisme algébrique mis en place précédemment. L'extension est en fait naturelle quand on a compris la logique suivie par notre construction au paragraphe IV.1.3.7 et nous montrons les liens existant entre le formalisme résultant et certaines méthodes classiques de la statistique nonparamétrique ainsi que de la théorie des votes. L'extension des méthodes précédentes au cas d'un échantillon de  $K$  tableaux de dissimilarité mesurée sur une échelle ordinale est alors rendue possible par ces liens. Nous décrivons donc brièvement le principe de construction d'un algorithme de Codage Multidimensionnel Ordinal Multiple dans ce cadre. Aucune méthode de résolution directe du problème ne semble en effet avoir été proposée à ce jour.



## CHAPITRE II

### PROBLEMES D'EXISTENCE D'UN PLONGEMENT EUCLIDIEN

Soit  $I$  un ensemble d'unités statistiques (des individus, des objets, des stimuli,...). Supposons  $I$  de cardinal fini  $n$ . La variable étudiée est une mesure du degré de dissemblance ou d'écart entre les unités statistiques

$$\delta : I \times I \rightarrow \mathbb{R}$$

résultant soit de l'observation, soit d'une procédure de calcul à partir de données de profils observées sur  $I$ . Les notions duales de proximité et de dissimilarité recouvrent notre expérience quotidienne, qui modélise la réalité en identifiant les éléments d'un ensemble abstrait comme des points d'un continuum, que l'on juge plus ou moins proches, ou plus ou moins éloignés.

Ces notions intuitives sont à la base du développement de la topologie, qui étudie les concepts mathématiques nécessaires à leur formalisation. La notion de base est alors celle d'espace abstrait défini par la donnée d'un couple formé par un ensemble abstrait, disons ici  $I$ , et une topologie sur  $I$ . Trois méthodes sont usuellement retenues pour définir une topologie sur  $I$ . La première est due à FRECHET, qui montra dès 1908 l'intérêt d'une formalisation de la notion d'espace de type  $L$ , pour lequel la notion de limite séquentielle de points de  $I$  est fondamentale. Les présentations modernes préfèrent une axiomatique fondée sur la notion d'ensembles ouverts, et les concepts liés de bases de voisinages, etc. Nous adoptons ici une troisième approche, imposée par certains processus de collecte des données. La notion fondamentale est alors celle de dissimilarité que nous formalisons ci-dessous.

#### II.1. RAPPELS SUR LES FONCTIONS DE DISSIMILARITÉ

**Définition 1 :** *Une application  $\delta$  définie sur  $I \times I$  et à valeurs réelles est une fonction de dissimilarité si*

- (i)  $\delta_{ij} \geq 0 \quad \forall (i,j) \in I \times I$
- (ii)  $\delta_{ii} = 0 \quad \forall i \in I$
- (iii)  $\delta_{ij} = \delta_{ji} \quad \forall (i,j) \in I \times I$

où l'on note  $\delta_{ij} = \delta(i,j)$  pour tout  $(i,j)$  dans  $I \times I$ .

**Définition 2 :** Le couple  $(I, \delta)$  formé par un ensemble non vide  $I$  (fini ou non) et une fonction de dissimilarité  $\delta$  définie sur  $I \times I$  est appelé un espace quasi-métrique.

La définition 1 introduit un ensemble de propriétés raisonnables décrivant la notion intuitive d'écart entre unités statistiques. Elle apparaît en pratique comme très exigeante et la littérature psychométrique témoigne d'un très large débat portant sur la pertinence de chacun des axiomes. L'axiome de positivité semble être le plus souvent accepté, seules des théories de l'erreur amènent parfois à sa remise en cause. L'axiome le plus fortement débattu est celui de symétrie. Les critiques les plus motivées sont dues à TVERSKY (1977). Elles reposent sur notre perception psychologique des notions de dissimilarité et montre le caractère très (ou trop) exigeant de cet axiome dans la modélisation des données perceptives. La dissymétrie observée entre jugements de dissimilarité est en général porteuse d'informations riches que l'on peut analyser à l'aide de modèles mathématiques proposés en particulier par GOWER (1977), GOWER et CONSTANTINE (1978). Une critique plus fondamentale du point de vue statistique nous semble liée à l'utilisation pratique de cet axiome. La symétrie d'une dissimilarité  $\delta$  ne peut en effet être observée qu'a posteriori, une fois l'expérience faite, à la consultation des données. Elle est le plus souvent imposée de façon artificielle par l'utilisation d'un plan d'expérience incomplet n'autorisant que l'une des comparaisons  $(i,j)$  ou  $(j,i)$ . L'axiome ii) semble lui aussi difficile à justifier dans de nombreuses expériences de psychologie perceptive ou cognitive. Du point de vue mathématique cependant, nous montrons dans la suite que le non respect de cet axiome n'oblitére pas tout traitement des données.

Sauf mention particulière nous nous limitons au cadre défini par la définition 1 dans la suite de ce travail. La définition de topologies réalistes se heurte à deux problèmes centraux. D'une part, on doit imposer une propriété de continuité à la fonction de dissimilarité, dont l'étude suppose des propriétés de densité de l'espace. Ce problème échappe au cadre d'étude du statisticien, puisque la densité de l'espace ne peut être mise à l'épreuve par l'observation d'un échantillon de taille finie. D'autre part, on doit imposer des contraintes plus fortes de séparation aux fonctions ou indices de dissimilarité.

**Définition 3 :** Une fonction de dissimilarité  $\delta$  définie sur  $I$  est dite semi-propre si elle vérifie pour tout couple  $(i,j)$  de  $I \times I$

$$(iv) (\delta_{ij} = 0) \Rightarrow (\delta_{ik} = \delta_{jk}) \quad \forall k \in I$$

Elle est propre si elle vérifie

$$(v) (\delta_{ij} = 0) \Rightarrow (i = j)$$

Ainsi toute fonction propre est semi-propre.

Les définitions 1 et 3 donnent un sens mathématique aux notions subjectives de dissemblance ou d'écart entre unités statistiques et donc de façon duale de proximité ou de ressemblance ou encore de similarité que les techniques étudiées dans ce mémoire se proposent de quantifier. Les axiomes proposés sont cependant plus faibles que ceux qu'il est usuel de retenir en géométrie métrique. Ceci est précisé par la propriété 1 ci-dessous, dont la démonstration est immédiate.

**Définition 4 :** Soit  $(\mathcal{X}, \rho)$  un couple formé d'un ensemble non vide  $\mathcal{X}$  et d'une application réelle  $\rho$  définie sur  $\mathcal{X} \times \mathcal{X}$ .  $(\mathcal{X}, \rho)$  est un espace semi-métrique si  $\rho$  vérifie pour tout triplet  $(x, y, z)$  d'éléments de  $\mathcal{X}$

$$d1 : \rho(x, y) \geq 0$$

$$d2 : \rho(x, y) = 0 \Leftrightarrow (x = y)$$

$$d3 : \rho(x, y) = \rho(y, x)$$

$(\mathcal{X}, \rho)$  est un espace métrique si  $\rho$  vérifie de plus l'inégalité triangulaire

$$d4 : \rho(x, y) + \rho(y, z) \geq \rho(x, z)$$

**Définition 5 :** Soit  $(\mathcal{X}, \rho)$  un couple formé d'un ensemble non vide  $\mathcal{X}$  et d'une application réelle  $\rho$  définie sur  $\mathcal{X} \times \mathcal{X}$ .  $(\mathcal{X}, \rho)$  est un espace pseudo-métrique si  $\rho$  vérifie pour tout triplet  $(x, y, z)$  d'éléments de  $\mathcal{X}$  les axiomes d1, d3 et d4.

Etant donnée une fonction de dissimilarité  $\delta$  observée sur  $I \times I$ , le plongement métrique (ou plus précisément le plongement métrique isométrique) de l'espace quasi-métrique  $(I, \delta)$  consiste en la construction d'une application

$$\begin{aligned} x : I &\rightarrow \mathcal{X} \\ i &\mapsto x(i) \end{aligned}$$

telle que  $\delta_{ij} = \rho(x(i), x(j))$  pour tout élément  $(i, j)$  de  $I \times I$ . La résolution du problème repose en particulier sur la vérification de l'inégalité triangulaire d4, rarement vérifiée de façon stricte et consistante par les jugements subjectifs de dissimilarité. La difficulté est accrue lorsqu'on spécifie la classe d'espaces métriques dans laquelle est recherchée la solution. C'est en particulier le cas des représentations Euclidiennes de  $(I, \delta)$ . On réalise alors un plongement métrique approché de  $(I, \delta)$  et l'on doit se donner les moyens d'évaluer la qualité de cette approximation.

**Propriété 1 :** Une fonction de dissimilarité  $\delta$  définit une pseudo-distance sur  $I$ , et alors  $(I, \delta)$  est un espace pseudo-métrique, si et seulement si  $\delta$  vérifie l'inégalité triangulaire.

WELLS et WILLIAMS (1972) remarquent à juste titre que la notion de plongement isométrique peut apparaître bien restrictif dans le cadre des problèmes de plongements

topologiques mais que c'est précisément cette caractéristique qui permet d'utiliser des méthodes analytiques. En particulier, BANACH (1932) montre que tout espace métrique séparable peut être plongé isométriquement dans l'ensemble  $C([0,1])$  des fonctions réelles continues, définies sur l'intervalle  $[0,1]$ . Mais cet espace de Banach est peu adapté aux objectifs de l'Analyse des Données puisqu'il conduit à des représentations de lecture difficile. Un cas particulier de telles représentations, dû à ANDREWS (1972), atteste par la rareté de son emploi, du peu d'engouement des utilisateurs pour une telle solution. De façon complémentaire, en corollaire au théorème d'Uryson, tout espace métrique séparable peut être plongé topologiquement dans un espace de Hilbert, mais l'exemple des espaces de Banach séparés  $\ell^p$  ( $1 \leq p < \infty$ ,  $p \neq 2$ ) montre (cf. infra) qu'un plongement isométrique est impossible. Ces résultats expliquent la difficulté d'établissement de résultats exploitables (cf. TVERSKY et KRANTZ (1970)) justifiant les méthodes de Codage Multidimensionnel Ordinal (Nonmetric Multidimensional Scaling en anglais), qui traitent le cas d'une fonction de dissimilarité mesurée sur une échelle ordinale. On ne retient alors que l'information ordinale induite sur  $I \times I$  par la fonction de dissimilarité observée.

Rappelons qu'une relation binaire  $\mathcal{R}$  définie sur un ensemble non vide  $\mathcal{X}$  est

- (i) réflexive si et seulement si  $(\forall x, x \in \mathcal{X}) \quad x \mathcal{R} x$
- (ii) transitive si et seulement si  $(\forall (x,y,z) \in \mathcal{X}^3) \quad (x \mathcal{R} y \text{ et } y \mathcal{R} z) \Rightarrow (x \mathcal{R} z)$
- (iii) antisymétrique si et seulement si  $(\forall (x,y) \in \mathcal{X}^2) \quad (x \mathcal{R} y \text{ et } y \mathcal{R} x) \Rightarrow (x=y)$
- (iv) symétrique si et seulement si  $(\forall (x,y) \in \mathcal{X}^2) \quad (x \mathcal{R} y) \Rightarrow (y \mathcal{R} x)$

La relation  $\mathcal{R}$  est dite totale ou complète si et seulement si pour tout  $(x,y) \in \mathcal{X}^2$ , on a soit  $x \mathcal{R} y$  soit  $y \mathcal{R} x$ . Une relation qui n'est pas totale est dite partielle.  $\mathcal{R}$  est une relation de préordre sur  $I$  si elle est réflexive et transitive. C'est une relation d'ordre si elle est de plus antisymétrique. Le couple  $(I, \mathcal{R})$  définit un ensemble (pré)ordonné si  $\mathcal{R}$  est un (pré)ordre défini sur  $I$ .

Nous utilisons dans la suite les conventions classiques en théorie des résiduations (cf. BLYTH et JANOWITZ (1972)) et nous supposons que les relations traitées sont complètes. On pose alors la définition

**Définition 6 :** Soit  $J$  une partie non vide de  $\mathcal{X}$ .

(i)  $J$  est un idéal (resp. un filtre) d'ordre de  $\mathcal{X}$  si et seulement si  $J$  vérifie

$$(x \in J \text{ et } y \mathcal{R} x) \Rightarrow (y \in J)$$

$$(\text{resp. } (x \in J \text{ et } x \mathcal{R} y) \Rightarrow (y \in J))$$

(ii) On appelle idéal (resp. filtre) principal d'ordre de  $\mathcal{X}$  tout idéal (resp. filtre) de  $\mathcal{X}$  de la forme

$$[\leftarrow, x] = \{y \in X ; y \mathcal{R} x\}$$

$$(\text{resp. } [x, \rightarrow] = \{y \in X ; x \mathcal{R} y\})$$

Ces parties de l' permettent de caractériser une classe de fonctions importante pour la suite de ce travail. Soit en effet  $I$  et  $J$  deux ensembles non vides et  $f$  une application de  $I$  dans  $J$

**Théorème 1 :** (BLYTH et JANOWITZ (1972)) : *Soient  $(I, \mathcal{R})$  et  $(J, \mathcal{F})$  deux ensembles ordonnés, et  $f$  une application de  $I$  dans  $J$ . Les conditions suivantes sont équivalentes :*

- (1)  $(x \mathcal{R} y) \Rightarrow (f(x) \mathcal{F} f(y))$
- (2) *l' image réciproque de tout filtre principal de  $J$  est soit vide, soit un filtre de  $I$ .*
- (3) *l' image réciproque de tout idéal principal de  $J$  est soit vide, soit un idéal de  $I$ .*

**Définition 7 :** *Soient  $(I, \mathcal{R})$  et  $(J, \mathcal{F})$  deux ensembles ordonnés. Une application  $f$  de  $I$  dans  $J$  est dite isotone si elle satisfait l'une des conditions équivalentes du théorème 1.*

Les filtres et idéaux principaux jouent un rôle privilégié dans l'étude des méthodes de régression isotone et de leurs applications à l'étude des fonctions de dissimilarité. En particulier, d'AUBIGNY (1975) exprime ces méthodes dans un espace de Hilbert  $L^2(\Omega, \mathcal{A}, P)$  et montre que les propriétés de mesurabilité par rapport à un  $\sigma$ -treillis qui les fondent, sont liées aux propriétés des fonctions indicatrices des filtres et des idéaux principaux. Posons alors la définition suivante :

**Définition 8 :** *Soit  $(I, \mathcal{R})$  un ensemble ordonné. On appelle indicatrice du filtre principal  $[x, \rightarrow]$  de  $I$  la fonction :*

$$\tau_x : I \rightarrow [0,1]$$

$$i \mapsto \tau_x(i) = \begin{cases} 1 & \text{si } i \in [x, \rightarrow] (\Leftrightarrow x \mathcal{R} i) \\ 0 & \text{sinon} \end{cases}$$

Construisons de plus l'indice de rang associé à la relation d'ordre  $\mathcal{R}$  et posons la définition classique (cf. HAJEK et SZIDAK (1967), LEHMAN(1975)) :

**Définition 9 :** *Soit  $(I, \mathcal{R})$  un ensemble ordonné où  $\text{card}(I) = n$ . On appelle indice de rang associé à la relation  $\mathcal{R}$ , la fonction réelle définie sur  $I$  par*

$$r : x \mapsto r(x) = (n + 1) - \text{card}([x, \rightarrow])$$

Appliquons maintenant les outils précédents aux relations définies sur  $I \times I$  par les fonctions de dissimilarité observées sur  $I$ .

**Définition 10 :** *Soit  $I$  un ensemble fini non vide et  $\mathcal{R}$  une relation binaire définie sur  $I \times I$ .  $\mathcal{R}$  est une préordonnance sur  $I$  si et seulement si elle est complète, réflexive et transitive et vérifie l'axiome de minimalité*

$$\forall (i,j,k) \in I^3 \quad (i,i) \mathcal{R} (j,k).$$

$\mathcal{R}$  est une ordonnance sur I si  $\mathcal{R}$  est de plus antisymétrique.

En pratique, une préordonnance  $\mathcal{R}$  sur I est induite par l'observation d'une fonction de dissimilarité sur I.

**Propriété 2 :** Soit  $(I,\delta)$  un espace quasi-métrique et  $\mathcal{R}$  la relation binaire définie sur  $I \times I$  par

$$\forall ((i,j),(k,l)) \in I^2 \times I^2 \quad (i,j) \mathcal{R} (k,l) \text{ si et seulement si } \delta_{ij} \leq \delta_{kl}$$

alors  $\mathcal{R}$  est une préordonnance sur I. On dit que  $\mathcal{R}$  est associée à  $\delta$  et on note  $\mathcal{R}_\delta$  lorsque cela est nécessaire.

Cependant, une même préordonnance peut être associée à plusieurs fonctions de dissimilarité et par suite on pose la définition

**Définition 11 :** Soient  $d$  et  $\delta$  deux fonctions de dissimilarité définies sur le même ensemble non vide I.  $d$  et  $\delta$  sont équivalentes si et seulement si pour tout élément  $((i,j),(k,l))$  de  $I^2 \times I^2$

$$(d_{ij} \leq d_{kl}) \Leftrightarrow (\delta_{ij} \leq \delta_{kl})$$

On note alors  $d \approx \delta$  et l'on vérifie sans peine que  $\approx$  est une relation d'équivalence.

En particulier nous utilisons dans la suite la fonction de dissimilarité suivante :

**Définition 12 :** Soit  $\mathcal{R}$  une préordonnance sur I. Supposons que  $\text{card}(I)=n$ . On appelle indice de rang de la préordonnance  $\mathcal{R}$ , l'application

$$\begin{aligned} r : I \times I &\rightarrow \mathbb{R}^+ \\ (i,j) &\mapsto r(i,j) = n^2 + 1 - \text{card}(\{(i,j), \rightarrow\}) \end{aligned}$$

On vérifie en effet simplement que  $r$  est une fonction de dissimilarité sur I. On déduit immédiatement des définitions précédentes la propriété suivante (FICHET (1983))

**Propriété 3 :** Soient  $d$  et  $\delta$  deux fonctions de dissimilarité équivalentes définies sur I. Alors  $d$  est propre (resp. semi-propre) si et seulement si  $\delta$  l'est.

De plus, l'introduction de contraintes de séparation permet de se ramener à l'étude des espaces semi-métriques. Nous considérons maintenant cette question en termes de préordonnances.

**Définition 13 :** (FICHET 1983) : Une préordonnance  $\mathcal{R}$  sur  $I$  est dite propre si et seulement si pour tout élément  $(i,j)$  de  $I^2$  :

$$(\forall (k,l) \in I^2 (i,j) \mathcal{R} (k,l)) \Rightarrow (i = j)$$

Elle est semi-propre si et seulement si pour tout  $(i,j)$  de  $I^2$  :

$$(\forall (k,l) \in I^2 (i,j) \mathcal{R} (k,l)) \Rightarrow (\forall k \in I (i,k) \mathcal{R} (j,k))$$

Il est alors clair qu'une préordonnance définie sur  $I$  est propre (resp. semi-propre) si et seulement si elle est associée à une fonction de dissimilarité propre (resp. semi-propre). Remarquons de plus qu'une préordonnance non propre sur  $I$  ne peut être semi-propre sur  $I$ , lorsque  $I$  est de cardinal  $n \geq 3$ .

D'autre part, bien que notre exposé soit centré sur les problèmes de plongement Euclidien d'un espace quasi-métrique, ce parti pris ne traduit pas de notre part une conviction d'optimalité du type de représentations qui en résultent. L'Analyse des Données utilise de façon au moins aussi fréquente la représentation  $(I, \delta)$  par des hiérarchies indicées ou encore de façon plus récente, par des arbres additifs. Dans le premier cas on utilise l'information contenue dans la préordonnance  $\mathcal{R}$  induite par  $\delta$  et l'on pose la définition

**Définition 14 :** Une fonction de dissimilarité  $\delta$  est ultramétrique sur  $I$  si pour triplet  $(i,j,k)$  d'éléments de  $I$  elle vérifie

$$d5 : \delta_{ij} \leq \text{Max} (\delta_{ik}, \delta_{jk})$$

Cet axiome est plus exigeant que l'inégalité triangulaire d4 et conduit à une représentation arborescente de  $I$ , puisque JOHNSON( 1974) montre qu'il existe une correspondance bijective entre l'ensemble des fonctions de dissimilarité ultramétriques et l'ensemble des hiérarchies indicées (cf. aussi BENZECRI (1973), JAMBU et LEBEAU (1978)).

Clairement, toute fonction de dissimilarité ultramétrique est une pseudo-distance.

**Définition 15 :** Une préordonnance  $\mathcal{R}$  sur  $I$  est dite ultramétrique si et seulement si pour tout triplet  $(i,j,k)$  d'éléments de  $I$  la proposition suivante est vraie

$$((i,j) \mathcal{R} (i,k)) \text{ ou } ((i,j) \mathcal{R} (j,k))$$

On vérifie alors (FICHET (1983)) que toute préordonnance ultramétrique sur  $I$  est semi-propre sur  $I$ . D'autre part si  $n \geq 3$ , une ordonnance sur  $I$  ne peut pas être ultramétrique. Enfin, notons  $\mathfrak{D}_I$ , l'ensemble des fonctions de dissimilarité définies sur  $I$  et  $[\delta]$  la classe d'équivalence de la fonction de dissimilarité  $\delta$  pour la relation d'équivalence  $\approx$ , on vérifie la proposition suivante (cf. FICHET (1983))

**Proposition 1 :** *Soit  $\mathcal{R}$  une préordonnance définie sur l'ensemble fini  $I$  et  $[r]$  la classe d'équivalence de son indice de rang  $r$ . Alors  $\mathcal{R}$  est propre (resp. semi-propre, ultramétrique) si et seulement si tout élément de  $[r]$  est propre (resp. semi-propre, ultramétrique).*

Deux unités statistiques  $i$  et  $j$  jouent le même rôle vis à vis des autres unités statistiques si elles rendent  $\mathcal{R}$  semi-propre. On peut alors définir sur  $I$  la relation

$$(i \sim j) \Leftrightarrow (\forall (k,l) \in I^2 (i,j) \mathcal{R} (k,l))$$

et l'on vérifie que  $\sim$  est une relation d'équivalence sur  $I$ , à laquelle on associe l'espace quotient  $I/\sim$  et la relation binaire  $[\mathcal{R}]$  sur  $I/\sim \times I/\sim$  définie par

$$((i],[j]) [\mathcal{R}] ((k],[l])) \Leftrightarrow (i,j) \mathcal{R} (k,l)$$

FICHET (1983) démontre alors la proposition suivante :

**Proposition 2 :**

- (i) Si  $\mathcal{R}$  est une préordonnance semi-propre sur  $I$ ,  $[\mathcal{R}]$  est une préordonnance sur  $I/\sim$ .
- (ii) La préordonnance semi-propre  $\mathcal{R}$  est ultramétrique si et seulement si  $[\mathcal{R}]$  est ultramétrique.

On peut donc limiter notre étude au cas des dissimilarités propres, que l'on obtient au besoin à l'aide d'un passage à l'espace quotient. Considérons en effet l'espace quotient  $I/\delta$  défini à partir de  $(I,\delta)$  et  $[\delta]$  la fonction de dissimilarité induite sur  $I/\delta$ . Alors (FICHET (1983))

**Proposition 3 :** *Soit  $(I,\delta)$  un espace quasi-métrique défini par une fonction de dissimilarité semi-propre sur  $I$ . Alors  $I/\sim$  et  $I/\delta$  sont égaux,  $[\mathcal{R}]$  est la préordonnance associée à  $[\delta]$  et  $(I/\delta, [\delta])$  est un espace semi-métrique.*

Dans cette construction d'une structure d'espace semi-métrique, la préordonnance triviale  $\mathcal{R}_0$  définie sur  $I$  par

$$\forall ((i,j),(k,l)) \in I^2 \times I^2 \quad (i,j) \mathcal{R}_0 (k,l)$$

joue un rôle particulier. On remarque en effet qu'elle est associée à la classe des fonctions de dissimilarité constantes sur  $I \times I$ . Si  $\mathfrak{R}$  est une préordonnance non triviale, c'est à dire si  $\mathfrak{R} \neq \mathfrak{R}_0$ , on pose la définition suivante :

**Définition 16 :** Si  $\mathfrak{R}$  est une préordonnance non triviale sur  $I$ , dont l'indice de rang prend  $(p + 1)$  valeurs distinctes  $(0 = r_0 < r_1 < \dots < r_p)$ , on dit que  $\mathfrak{R}$  est de degré  $p$ . Par extension, on dit que  $\mathfrak{R}_0$  est de degré 0.

Alors les indicatrices des filtres principaux de  $\mathfrak{R}$  s'écrivent de façon équivalente

$$\begin{aligned} \tau_l : I &\rightarrow [0,1] \\ (i,j) &\mapsto \tau_l(i,j) = \begin{cases} 1 & \text{si } r(i,j) \geq r_l \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

pour  $l = 1, 2, \dots, p$ . On vérifie alors la proposition suivante :

**Proposition 4 :** Soit  $\mathfrak{R}$  une préordonnance non triviale sur  $I$  ( $\text{card}(I) = n$ ). Si  $\mathfrak{R}$  est de degré  $p$ , une fonction de dissimilarité  $\delta$  est associée à  $\mathfrak{R}$  si et seulement si elle admet une décomposition unique de degré  $p$  de la forme

$$\delta = \sum_{j=1}^p y^j \tau_j$$

où les coefficients  $y^j$  sont réels, strictement positifs. De plus les  $p$  valeurs strictement positives  $\lambda_1 < \lambda_2 < \dots < \lambda_p$  prises par  $\delta$  sur  $I \times I$  vérifient pour tout  $l = 1, 2, \dots, p$

$$\lambda_l = \sum_{j=1}^l y^j$$

**Preuve (FICHET (1983)) :** Supposons tout d'abord que  $\delta$  vérifie l'équation (1) avec  $y^m > 0$ ,  $m=1,2,\dots,p$ . Alors pour tout  $(i,j)$  dans  $I \times I$

$$r(i,j) = r_l \Leftrightarrow \tau_k(i,j) = 1 \quad \forall k \leq l$$

Par suite

$$r(i,j) = 0 \Leftrightarrow \delta(i,j) = \sum_{j=1}^p y^j 0 = 0$$

$$r(i,j) = r_l \Leftrightarrow \delta(i,j) = \sum_{j=1}^l y^j 1 + \sum_{j=l+1}^p y^j 0 = \sum_{j=1}^l y^j$$

et donc  $\delta$  prend  $(p+1)$  valeurs distinctes  $\lambda_l$

$$\lambda_0 = 0, \quad \lambda_l = \sum_{j=1}^l y^j, \quad l = 1, \dots, p$$

qui sont strictement ordonnées si  $y^j > 0$  pour tout  $j$ . Réciproquement, si  $\delta$  est associée à  $\mathcal{R}$ , puisque  $r$  prend  $(p+1)$  valeurs distinctes, il en est de même de  $\tau$  qui appartient à  $[r]$ .

Notons ces valeurs

$$0 = \lambda_0 < \lambda_1 < \dots < \lambda_p$$

On a alors pour tout  $l = 0, 1, \dots, p$

$$(r(i,j) = r_l) \Leftrightarrow (\delta_{ij} = \lambda_l)$$

Posons  $y^1 = \lambda_1$  et  $y^j = \lambda_j - \lambda_{j-1}$  pour  $j = 2, 3, \dots, p$ . Alors la fonction de dissimilarité

$$d = \sum_{j=1}^p y^j \tau_j$$

vérifie pour tout  $l = 1, 2, \dots, p$

$$\delta_{ij} = \lambda_l \Leftrightarrow r(i,j) = r_l \Leftrightarrow d(i,j) = \sum_{j=1}^l y^j = \lambda_l$$

$$\delta_{ij} = 0 \Leftrightarrow r(i,j) = 0 \Leftrightarrow d(i,j) = 0$$

Par suite  $\delta = d$       ¶

**Remarque :** Nous donnons au chapitre VI une interprétation géométrique de ce résultat, associée aux propriétés de dualité Lagrangienne entre l'espace vectoriel  $F \wedge F$  décrit par  $\underline{y} = (y^j)_j$  et l'espace vectoriel  $G$  décrit par  $\underline{\lambda} = (\lambda_l)_l$

Une extension possible des hiérarchies indicées, à laquelle il est fait référence au chapitre VI, est fournie par les arbres additifs, dont une étude synthétique est présentée par BARTHELEMY et GUENOCHÉ (1988). Ce sont des graphes  $G = (I, U)$  particuliers caractérisés par le fait que chaque arête  $u_{ij}$  est pondérée de telle sorte que la distance séparant deux sommets de l'arbre est égale à la somme des poids affectés à chacune des arêtes formant le plus court chemin joignant ces deux sommets dans l'arbre. BUNEMAN (1974) montre qu'il existe une correspondance bijective entre les arbres additifs et les fonctions de dissimilarité quadripolaire

**Définition 17 :** Une fonction de dissimilarité  $\delta$  est quadripolaire sur  $I$  si elle vérifie pour tout  $(i, j, k, l)$  de  $I^4$

$$d6 : \delta_{ij} + \delta_{kl} \leq \text{Max} (\delta_{ij} + \delta_{jl}, \delta_{il} + \delta_{jk} )$$

Le problème mathématique traité par les méthodes de représentation des fonctions de dissimilarité quadripolaire est cependant de nature très différente du cas ultramétrique, parce que l'information ordinale induite sur  $I \times I$  par  $\mathfrak{R}_\delta$  ne suffit alors pas à la description de la propriété  $d_6$ , puisqu'on se propose d'évaluer des sommes de dissimilarités. Les valeurs de  $\delta$  doivent de plus être prises en compte.

La recherche d'une représentation Euclidienne suppose en particulier que l'inégalité triangulaire est vérifiée par la dissimilarité observée  $\delta$ . Nous plaçons donc ici notre étude dans le cadre de l'analyse des espaces pseudo-métriques.

SHEPARD et CUNNINGHAM (1974), d'AUBIGNY (1975) et CRITCHLEY (1978) montrent comment on construit une approximation d'une fonction de dissimilarité par une pseudo-distance au sens des moindres carrés. Cette approximation est cependant d'un intérêt limité pour l'Analyse des Données, lorsqu'on se propose de construire une représentation graphique résumée de données de dissimilarité. On étudie donc maintenant les espaces semi-métriques, définis par la donnée d'une dissimilarité semi-propre observée, et passage à l'espace quotient si nécessaire.

Le plongement Euclidien d'un espace semi-métrique  $(I, \delta)$  pose un double problème. On traite en effet tout d'abord un problème géométrique, puis on décrit la solution dans un cadre algébrique.

Tout d'abord, nous formalisons le problème de géométrie affine sous-jacent, après un bref rappel des notions classiques sur les espaces affines nécessaires à notre exposé. Le lecteur intéressé peut se reporter à LANG (1965) ou TISSERON (1983) pour une étude plus complète des espaces affines.

## II.2. RAPPELS SUR LES ESPACES AFFINES EUCLIDIENS

Soit  $E$  un espace affine, de direction  $\vec{E}$ . Supposons de plus que  $E$ , ou de façon équivalente  $\vec{E}$  est de dimension  $p$ . Alors par définition il existe une application

$$\begin{aligned} \psi : E \times E &\rightarrow \vec{E} \\ (P_1, P_2) &\mapsto \overrightarrow{P_1 P_2} \end{aligned}$$

telle que tout triplet  $(P_1, P_2, P_3)$  d'éléments de  $E$  vérifie la relation de Chasles

$$\overrightarrow{P_1 P_2} + \overrightarrow{P_2 P_3} = \overrightarrow{P_1 P_3}$$

et telle que pour tout élément  $P_0$  de  $E$ , l'application

$$\theta_0 : P \mapsto \overrightarrow{P_0 P}$$

est une bijection de  $E$  sur  $\vec{E}$ .

Supposons de plus que  $\vec{E}$  est muni d'une base  $\mathfrak{B} = \{\vec{e}_j, j = 1, 2, \dots, p\}$ , de sorte que tout choix d'origine  $P_0$  dans  $E$  permet de définir un référentiel (ou repère cartésien)  $\{O, (\vec{e}_j, j = 1, 2, \dots, p)\}$ . L'ensemble de points  $\{P_0, \{P_j; j \in [p]\}\}$  forment alors un repère affine de  $E$ , si

$$\vec{e}_j = \overrightarrow{P_0 P_j}$$

Alors tout point  $P$  de  $E$  est associé au vecteur  $\overrightarrow{P_0 P}$  de  $\vec{E}$  de coordonnées  $(t_j, j \in [p])$  dans  $\mathfrak{B}$ . Les  $t_j$  sont les uniques scalaires tels que

$$P = P_0 + \sum_{j=1}^p t_j \vec{e}_j = \sum_{j=0}^p t_j \vec{e}_j \quad \text{où} \quad t_0 = 1 - \sum_{j=1}^p t_j$$

$(t_0, t_1, \dots, t_p)$  fournissent alors les coordonnées barycentriques de  $P$  dans le repère affine  $\{P_0, P_1, \dots, P_p\}$ , puisque

$$\sum_{j=0}^p t_j = 1$$

Le choix de  $(E, \mathfrak{B})$  permet d'associer à tout élément  $i$  de  $I$  un point  $P_i$  de  $E$ . Pour tout choix d'une origine  $P_0$ , il lui correspond un vecteur  $\mathbf{z}_i$  de  $\vec{E}$  de coordonnées  $\mathbf{z}_i$  dans la base  $\mathfrak{B}$ . On note de plus  $\mathbf{t}_i$  le  $(p+1)$ -tuple de ses coordonnées barycentriques dans  $E$ , muni du référentiel  $(P_0, \mathfrak{B})$ . Puisque  $E$  est un espace affine de direction  $\vec{E}$ , il existe un espace vectoriel  $\hat{E}$  tel que  $E$  (resp.  $\vec{E}$ ) soit un sous-espace affine (resp. vectoriel) de  $\hat{E}$ . On vérifie de plus la propriété suivante (TISSERON (1981))

#### Propriété 4 :

(i)  $E$  est un hyperplan affine de  $\hat{E}$  de direction  $\vec{E}$

(ii) tout vecteur  $y$  de  $\vec{E}$  qui n'est pas élément de  $E$  s'écrit de façon unique

$$y = \lambda P \quad P \in E, \quad \lambda \neq 0$$

(iii)  $\forall P, P \in E$ , on a  $\hat{E} = \{\lambda P + \vec{E}; \lambda \in \mathbb{R}\}$

Alors tout ensemble de points massiques  $\{(P_j, m_j); j \in [p]\}$  où les poids  $m_j$  réels vérifient

$$\sum_j m_j = m, \quad m \neq 0$$

admet pour barycentre le point  $P$  de  $E$  tel que pour tout choix d'une origine  $O$  dans  $E$

$$m \vec{OP} = \sum_j m_j \vec{OP}_j$$

Cette relation s'écrit dans  $\hat{E}$

$$m (\vec{x} - \vec{0}) = \sum_j m_j (\vec{x}_j - \vec{0}) = \sum_j m_j \vec{x}_j$$

et donc en posant  $p_j = \frac{m_j}{m}$

$$\vec{x} = \sum_j p_j \vec{x}_j$$

On vérifie ainsi que les barycentres dans  $E$  sont exactement les combinaisons linéaires dans  $\hat{E}$ . Par suite, les familles affines libres de  $E$  sont exactement les familles libres de  $\hat{E}$ , formées d'éléments de  $E$ . De même, les repères affines de  $E$  sont exactement les bases de  $\hat{E}$  formées d'éléments de  $E$  et pour tout repère affine  $\{P_0, \{P_j; j \in [p]\}\}$  de  $E$ , les coordonnées barycentriques d'un point  $P$  dans ce repère sont exactement les coordonnées du vecteur d'extrémité  $P$  dans  $\hat{E}$ , dans cette base. Nous identifions donc dans la suite les notations  $t_i$  et  $z_i$ .

**Remarque :** Si  $\{\vec{e}_j; j \in [p]\} \subset \vec{E}$ , alors, puisque  $\hat{E} = \{\lambda P + \vec{E}; \lambda \in \mathbb{R}\}$ , le choix d'une origine  $P_0 \in E$  permet d'écrire tout élément  $P$  de  $E$  sous la forme

$$P = P_0 + \sum_{j=1}^p x_j \vec{e}_j$$

On lui associe donc ses coordonnées dans  $\{P_0, \{\vec{e}_j; j \in [p]\}\}$

$$\tilde{x} = \begin{pmatrix} 1 \\ \underline{x} \end{pmatrix}$$

où  $\underline{x}$  est le vecteur de ses coordonnées dans la base  $\{\vec{e}_j; j \in [p]\}$ .

Soit maintenant  $b$  une forme bilinéaire symétrique définie sur  $\hat{E}$ . On lui associe classiquement la forme quadratique  $Q$  définie par

$$Q(x) = b(x, x) \quad \forall x \in \hat{E}$$

et lorsqu'il n'y a pas de confusion possible, on identifie les notations  $q$  et  $b$ .  $b$  est la forme polaire de  $Q$  et l'on retrouve  $b$  à partir de  $Q$  par la relation classique

$$b(x, x) = \frac{1}{2} [Q(x+y) - Q(x) - Q(y)]$$

Munissons  $\hat{E}$  de la base  $\{e_j; j \in [p]\}$ . On associe alors à  $b$  et à  $Q$  la même matrice  $B = (b_{ij})$  où  $b_{ij} = b(e_i, e_j)$ . Dans la suite on interprète de façon équivalente  $B$  comme la matrice d'une application linéaire  $n : \hat{E} \rightarrow \hat{E}^*$  où  $\hat{E}^*$  est le dual algébrique de  $\hat{E}$ ,  $n$  a pour matrice  $B$  lorsque  $\hat{E}^*$  est muni de la base duale  $\mathfrak{E}^* = \{\vec{e}_j^*; j = 1, 2, \dots, p\}$  de  $\mathfrak{E}$ . On identifie donc par abus  $n$  et  $b$ .

Considérons maintenant la forme 2-affine  $a$  définie sur un espace affine  $E$  de dimension  $p$ , rapporté au référentiel  $\{P_0, \{\vec{e}_j; j \in [p]\}\}$ .  $a$  est définie par

$$a : E \rightarrow \mathbb{R}$$

$$P \mapsto a(P) = \sum_i \sum_j m_{ij} x_i x_j + 2 \sum_i b_i x_i + c$$

où  $m_{ij} = m_{ji}$  pour tout  $(i, j)$  et il existe  $(i, j)$  tel que  $m_{ij} \neq 0$  et

$$P = P_0 + \vec{x} \quad , \quad \vec{x} \in \vec{E} \quad \vec{x} = \sum_j x_j \vec{e}_j$$

On associe alors classiquement à l'application  $a$  les matrices

$$M = (m_{ij}) \quad A = \begin{pmatrix} c & \underline{b} \\ \underline{b} & M \end{pmatrix}$$

**Remarque :** Nous montrons dans la suite le rôle privilégié joué par de telles formes 2-affines pour l'élaboration de conditions d'existence de plongement Euclidien de  $(I, \delta)$ . On a alors  $c = 0$  et  $\underline{b} = \underline{1}$ , le vecteur constant de coordonnées égales à 1.

La forme quadratique  $Q$  définie sur  $\hat{E}$  par la matrice  $A$  dans  $\{P_0, \{\vec{e}_j; j \in [p]\}\}$  interprétée comme base de  $\hat{E}$ , vérifie

$$Q(\vec{y}) = \sum_i \sum_j m_{ij} x_i x_j + 2k \sum_i b_i x_i + k^2 c \quad (1)$$

pour tout élément  $\vec{y}$  de  $\hat{E}$  tel que

$$\vec{y} = \sum_j x_j \vec{e}_j + kP_0$$

$E$  est alors décrit par les vecteurs  $\underline{y}$  correspondant à la valeur  $k=1$  et donc  $a$  est la restriction de  $Q$  à  $E$  et  $Q$  est l'unique extension de  $a$  à  $\hat{E}$  puisque l'équation (1) détermine complètement la matrice associée à  $Q$  dans la base retenue. De plus  $\hat{E}$  est décrit par les vecteurs  $\underline{y}$  correspondant à la valeur  $k = 0$  et donc

$$q(\vec{x}) = \sum_i \sum_j m_{ij} x_i x_j \quad \vec{x} \in \vec{E}$$

est la restriction de  $Q$  à  $\vec{E}$ .  $q$  est non nulle s'il existe au moins un couple  $(i,j)$  tel que  $m_{ij} \neq 0$ . La matrice  $M$  de terme général  $m_{ij}$ , définit une structure Euclidienne sur  $\vec{E}$  et donc sur  $E$ . On dit souvent par abus de langage que la forme bilinéaire  $m$  polaire de  $q$  définit une métrique sur  $E$  et l'on note lorsque la précision est nécessaire  $E = (\mathbb{R}^p, m)$ .

Notons  $\overrightarrow{P_i P_j}$  le vecteur de  $\vec{E}$  d'origine  $P_i$  et d'extrémité  $P_j$ . Alors le plongement Euclidien isométrique de  $(I, \delta)$  consiste à associer à tout élément  $i$  de  $I$  un point  $P_i$  de  $E$  tel que

$$\|\overrightarrow{P_i P_j}\|_M = \delta_{ij}$$

où on note  $\|\overrightarrow{P_i P_j}\|_M$  la norme du vecteur  $\overrightarrow{P_i P_j}$  pour la métrique  $M$  :

$$\|\overrightarrow{P_i P_j}\|_M = m(\overrightarrow{P_i P_j}, \overrightarrow{P_i P_j})$$

### II.3. NOTION DE DISSIMILARITÉ EUCLIDIENNE

**Définition 18 :** Soit  $(I, \delta)$  un espace semi-métrique et  $E$  un espace affine Euclidien de dimension finie  $p$ . On appelle image Euclidienne de  $(I, \delta)$ , lorsqu'elle existe, la famille des points  $\{P_i, i \in I\}$  de  $E$  tels que

$$\|\overrightarrow{P_i P_j}\|_M = \delta_{ij}$$

Il est clair qu'une telle image n'existe pas toujours. Si les axiomes définissant une semi-métrique sont nécessaires, ils ne sont pas suffisants. De nombreux contre-exemples ont été publiés, nous citons ici celui proposé par FICHET (1983):

**Contre exemple :** Soit  $I = \{1,2,3,4\}$  et  $\delta^1$  et  $\delta^2$  deux fonctions de dissimilarité définies sur  $I$  par

$$\begin{aligned} \delta_{12}^1 = \delta_{13}^1 = \delta_{23}^1 = 0; \delta_{14}^1 = \delta_{24}^1 = \delta_{34}^1 = 1 \\ \delta_{12}^2 = \delta_{13}^2 = \delta_{23}^2 = \delta_{34}^2 = 0; \delta_{14}^2 = \delta_{24}^2 = 1 \end{aligned}$$

On vérifie que  $\delta^1$  admet une image Euclidienne dans  $E^1 = (\mathbb{R}, I)$  où  $I$  est la matrice identité, sous la forme des extrémités  $A$  et  $B$  d'un segment de longueur 1 où  $A = P_1 = P_2 = P_3$  et  $B = P_4$ . De même,  $\delta^2$  admet une image Euclidienne  $E^2 = (\mathbb{R}^2, I)$  sous la forme des sommets  $A, B$  et  $C$  d'un triangle équilatéral dont la longueur des cotés est égale à 1 et tel que  $A = P_1, B = P_2, C = P_3 = P_4$ . Considérons alors la combinaison convexe de  $\delta^1$  et  $\delta^2$

$$\delta = \frac{1}{2} (\delta^1 + \delta^2)$$

$\delta$  vérifie les axiomes d'une semi-distance car

$$\delta_{12} = \delta_{13} = \delta_{23} = \delta_{34} = \frac{1}{2} ; \delta_{14} = \delta_{24} = 1$$

mais il n'existe pas d'image Euclidienne de  $\delta$  (cf. FICHET (1983)). En effet dans le cas contraire, celle-ci serait possible dans un espace de dimension 3 ou plus et  $P_1, P_2, P_3$  formerait un triangle équilatéral de coté de longueur  $1/2$ . Puisque  $\delta_{14} = 1$  et  $\delta_{34} = 1/2$ , le point  $P_4$  appartiendrait alors à l'intersection de deux sphères de rayons respectifs 1 et  $1/2$  centrés respectivement en  $P_1$  et  $P_2$ . Mais  $\delta_{13} = 1/2$  de sorte que ces deux sphères sont tangentes intérieurement à  $P'_1$  symétrique de  $P_1$  par rapport à  $P_3$ . Le même argument, utilisant  $\delta_{24} = 1$  et  $\delta_{34} = 1/2$  montre que  $P_4$  doit être confondu avec  $P'_2$  symétrique de  $P_2$  par rapport à  $P_3$ . Mais puisque  $\delta_{12} = 1/2$ , ces deux conditions ne peuvent être réalisées simultanément.

Supposons maintenant donnée une bijection arbitraire de  $I$  dans  $[n] = \{1, 2, \dots, n\}$ . Dans la suite nous identifions  $I$  et  $[n]$ . Supposons de plus que  $(I, \delta)$  admette une image Euclidienne. Le point  $P_i$  est repéré dans  $\vec{E}$  par le vecteur de ses coordonnées  $z_i$  dans la base  $\mathfrak{Z}$ . On interprète le problème dans un cadre linéaire, en considérant l'espace vectoriel réel libre  $\vec{F}$  de dimension  $n$ , muni d'une base  $\mathfrak{F} = \{\vec{f}_i, i \in [n]\}$ . Nous notons  $\vec{F}^*$  son dual algébrique et  $\mathfrak{F}^* = \{\vec{f}_i^*, i \in [n]\}$  la base duale de  $\mathfrak{F}$ .

Nous identifions alors les unités statistiques  $i$  de  $I$  aux vecteurs de base  $\vec{f}_i^*$ , de sorte que  $z_i$  s'interprète comme l'image de  $\vec{f}_i^*$  par une application linéaire :

$$\begin{aligned} z : \vec{F}^* &\rightarrow \vec{E} \\ \vec{f}_i^* &\mapsto z(\vec{f}_i^*) = z_i \end{aligned}$$

$z$  est totalement décrite dans les bases  $\mathfrak{Z}$  et  $\mathfrak{F}^*$  par la donnée du tableau  $Z$  de  $i$ -ème colonne  $z_i$ .  $Z$  est alors un élément de l'espace vectoriel des matrices réelles à  $p$  lignes et  $n$  colonnes  $\mathcal{M}_{p,n}(\mathbb{R})$  et l'on note

$$Z = (z_i^j)_{i=1\dots n, j=1\dots p}$$

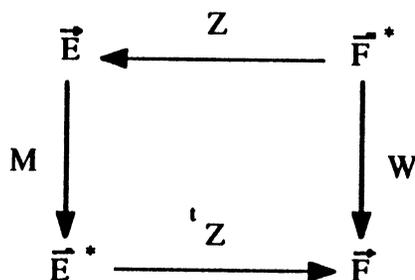
**Définition 19 :** Soit  $(I, \delta)$  un espace semi-métrique admettant une image Euclidienne dans un espace affine Euclidien de dimension  $p$ . La fonction de dissimilarité  $\delta$  est alors dite Euclidienne. On note alors  $\delta = \delta(Z)$ , où  $Z$  est un élément de  $\mathcal{M}_{p,n}(\mathbb{R})$  appelé la (matrice de) configuration de  $I$  dans  $E$ . On dit encore que  $Z$  est une représentation de  $I$ .

Le contre-exemple proposé par FICHET (1983) et présenté plus haut, permet de plus de montrer qu'une combinaison convexe de fonctions de dissimilarité euclidiennes n'est pas forcément euclidienne.

Si  $\delta$  est Euclidienne,  $P_i$  est représenté dans  $\vec{E}$  par le vecteur  $z_i$  et par suite

$$\delta_{ij}(Z) = \| \overrightarrow{P_i P_j} \|_M = \| z_i - z_j \|_M = [ {}^t(z_i - z_j) M (z_i - z_j) ]^{\frac{1}{2}}$$

Complétons de plus le schéma de dualité associé à l'interprétation linéaire du tableau Z associé à la représentation de I dans E



On interprète classiquement (cf. CAILLEZ et PAGES (1976)) Z comme la matrice d'une isométrie entre  $\vec{E}$  et  $\vec{F}^*$ , dans les bases  $\mathfrak{Z}$  et  $\mathfrak{F}^*$ , en considérant la forme quadratique w définie sur  $\vec{F}^*$  par

$$w(\vec{f}_i^* - \vec{f}_j^*, \vec{f}_i^* - \vec{f}_j^*) = m(z_i - z_j, z_i - z_j) = \delta_{ij}^2(Z) \quad \forall i, j \in [n] \quad (2)$$

c'est dire que w a pour matrice dans les bases  $\mathfrak{F}^*$  et  $\mathfrak{F}$

$$W = {}^t Z M Z \quad (3)$$

W dépend clairement de M et de Z. M est une donnée du problème fixée par l'utilisateur, alors que W est déduite des observations Z et dépend du modèle descriptif adopté. On note cette matrice  $W_Z$ , lorsqu'il est nécessaire.

La relation (2) a deux conséquences :

(a) dans  $\vec{F}^*$ , l'expression (3) montre que W est symétrique, positive semi-définie et l'on peut noter  $\langle \cdot, \cdot \rangle_{W_Z}$  la forme bilinéaire symétrique de matrice  $W_Z$  et  $\| \cdot \|_{W_Z}$  la semi-norme

$$\| \underline{x} \|_{W_Z}^2 = \langle \underline{x}, \underline{x} \rangle_{W_Z} \quad \underline{x} \in \vec{F}^*$$

On vérifie de plus à partir de (2), par linéarité

$$\begin{aligned}
 \delta_{ij}^2(Z) &= w(\vec{f}_i^*, \vec{f}_i^*) + w(\vec{f}_j^*, \vec{f}_j^*) - 2 w(\vec{f}_i^*, \vec{f}_j^*) \\
 &= w_{ii} + w_{jj} - 2w_{ij}
 \end{aligned} \quad (4)$$

où  $w_{ij}$  est le terme général de la matrice  $W_Z$ .

(b) dans  $\vec{E}$  l'expression (2) s'écrit de façon analogue

$$\delta_{ij}^2(Z) = \|z_i\|_M^2 + \|z_j\|_M^2 - 2\langle z_i, z_j \rangle_M$$

où par définition

$$\begin{aligned} \langle z_i, z_j \rangle_M &= m(z_i, z_j) = \sum_{s=1}^p \sum_{t=1}^p z_i^s z_j^t m(\vec{e}_s, \vec{e}_t) \\ &= \sum_{s=1}^p \sum_{t=1}^p z_i^s z_j^t m_{st} = {}^t z_i M z_j \end{aligned}$$

L'intérêt d'une telle construction repose sur les propriétés d'unicité des outils mis en œuvre pour la construction d'une image Euclidienne de  $(I, \delta)$ . Ce problème est étudié par FICHET (1983) qui démontre la proposition suivante

**Proposition 5 :** Soient  $E^1$  et  $E^2$  deux espaces affines Euclidiens de dimensions finies respectives  $p_1$  et  $p_2$ . On note  $\{P_i^k, i=1, 2, \dots, n\}$  les points représentatifs de  $I$  dans  $E^k$ ,  $k = 1, 2$ . Si pour tout couple  $(i, j)$  d'éléments de  $I$  on a

$$m_1(\overrightarrow{P_i^1 P_j^1}, \overrightarrow{P_i^1 P_j^1}) = m_2(\overrightarrow{P_i^2 P_j^2}, \overrightarrow{P_i^2 P_j^2})$$

alors

- i)  $p_1 = p_2$
- ii) Pour  $k = 1, 2$ , il existe dans  $E^k$  un référentiel  $(O^k, \mathcal{E}^k)$  tel que
  - (1)  $m_1$  et  $m_2$  ont pour représentation matricielle  $M = M_1 = M_2$
  - (2)  $P_i^1$  et  $P_i^2$  ont pour vecteur de coordonnées dans leur base respective le même vecteur  $z_i$ .

Par suite on pose la définition suivante, rappelant à nouveau que la dimension d'un espace affine Euclidien  $E$  est la dimension de l'espace vectoriel sous-jacent  $\vec{E}$ .

**Définition 20** (FICHET (1983)) : On appelle dimension d'un espace quasi-métrique  $(I, \delta)$ , où  $\delta$  est une fonction de dissimilarité Euclidienne, la dimension de l'espace affine Euclidien  $E$  support de l'une quelconque de ses images Euclidiennes.

De plus nous avons vu dans la section §II.1 que l'étude des fonctions de dissimilarité propres fournit en fait un cadre général d'étude d'un espace quasi-métrique, par passage éventuel à l'espace quotient, afin d'assurer le respect des axiomes de séparation. Cette construction est complétée par la proposition suivante due à FICHET (1983)

**Proposition 6 :** Soit  $(I, \delta)$  un espace quasi-métrique et  $[\delta]$  la dissimilarité induite par  $\delta$  sur l'espace quotient  $I/\delta$ .  $[\delta]$  est Euclidienne si et seulement si  $\delta$  l'est et de plus  $(I, \delta)$  et  $(I/\delta, [\delta])$  ont la même dimension.

#### II.4. EXISTENCE D'UN PLONGEMENT EUCLIDIEN D'UN ESPACE QUASI-MÉTRIQUE $(I, \delta)$

La détermination de conditions nécessaires d'existence d'un plongement isométrique d'un espace métrique, semi-métrique ou quasi-métrique  $(I, \delta)$  dans un espace vectoriel  $E = (\mathbb{R}^p, d)$  a fait l'objet de nombreux travaux dans le domaine de la géométrie des espaces distanciés. On se propose alors de construire une représentation exacte de  $I$ . A chaque élément  $i$  est associé un point  $P_i$  de coordonnées  $\underline{z}_i$  dans la base  $\mathfrak{E}$  de  $E$  et l'on recherche l'application

$$\begin{aligned} z : I &\rightarrow E \\ i &\mapsto z(i) = \underline{z}_i \end{aligned}$$

qui réalise exactement l'égalité

$$d(\underline{z}_i, \underline{z}_j) = \delta_{ij}$$

La recherche de représentations approchées est de nature statistique et CRITCHLEY (1985) précise à juste raison qu'il convient de ne pas identifier les deux problèmes. Nous traitons ci-dessous le problème posé par le plongement Euclidien de  $(I, \delta)$  afin de montrer la communauté des outils employés dans les deux approches du problème. Nous reprenons tout d'abord l'approche de CAYLEY (1841) systématisée et étendue par MENGER (1928). L'étude de ces auteurs s'exprime dans un cadre de géométrie affine et repose sur les propriétés des déterminants. Leur interprétation en termes de volume du simplexe engendré par  $n$  points dans un espace affine Euclidien est naturelle lorsqu'on s'intéresse aux relations entre distance ( $p=1$ ), surface ( $p=2$ ) et volume ( $p=3$ ), mises en évidence par DUBIKAJTIS (1977) entre autres. Nous rappelons ensuite les principaux résultats fondés sur les propriétés des formes quadratiques et les fonctions de type positif. Ceux-ci traitent le problème dans un cadre vectoriel et caractérisent l'apport de FRECHET (1935) et surtout de SCHOENBERG (1935, 1937, 1938) sur le sujet. L'intérêt de cette approche est lié à l'importance accordée ici à l'Analyse en Coordonnées Principales, dont les bases résultent très directement des travaux de YOUNG et HOUSEHOLDER (1938).

Soit  $\{P_i, i \in [n]\}$  un ensemble de  $n$  points d'un espace affine Euclidien. Ils ont toujours pour support un espace affine  $E = (\mathbb{R}^p, m)$  où  $p$  est inférieur ou égal à  $(n - 1)$  et  $m$  est une forme quadratique associée à la distance  $d$ . Alors pour tout couple  $(i, j)$  d'éléments de  $I$

$$d(\underline{z}_i, \underline{z}_j) = \|\underline{z}_i - \underline{z}_j\|_M$$

où  $\underline{z}_i$  est le vecteur de coordonnées de  $P_i$  dans  $E$ . Sans perte de généralité nous pouvons supposer que la matrice  $M$  est de plein rang et égale à l'identité. Les points  $P_i$  sont les sommets d'un simplexe (qui peut être dégénéré) contenu dans  $E$  et dont le volume (généralisé si  $p > 3$ ) s'exprime en termes de déterminant :

$$V((P_i)) = \frac{1}{(n-1)!} \det \begin{pmatrix} \mathbf{1} \\ Z \end{pmatrix}$$

où  $Z$  est le tableau de  $i$ -ème colonne  $\underline{z}_i$  et  $\mathbf{1}$  est le vecteur de coordonnées constantes égales à 1. Cette écriture montre que l'on regarde  $E$  comme un hyperplan d'un espace vectoriel  $\hat{E}$  construit comme au paragraphe §II.2. Remarquons que  $V((P_i))$  est égale à  $d(\underline{z}_i, \underline{z}_j)$  si  $n = 2$ , à la surface du triangle de sommets  $P_1, P_2$  et  $P_3$  si  $n=3$ , au volume du tétraèdre de sommets  $P_1, P_2, P_3$  et  $P_4$  si  $n=4$ . De plus, notons  $V_Z$  et  $W_Z$  les matrices définies par

$$V_Z = Z {}^t Z \quad W_Z = {}^t Z Z$$

**Proposition 7 :**

$$V^2((P_i)) = \frac{-1}{[(n-1)!]^2} \det \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & W_Z \end{pmatrix}$$

$$V^2((P_i)) = \frac{1}{[(n-1)!]^2} \det (W_Z + \mathbf{1}\mathbf{1})$$

$$V^2((P_i)) = \frac{-1}{[(n-1)!]^2} \det \begin{pmatrix} n & \mathbf{1} \\ \mathbf{1} & V_Z \end{pmatrix}$$

**Preuve :** Posons

$$Y_1 = \begin{pmatrix} 1 & \mathbf{0} \\ 0 & \mathbf{1} \\ \mathbf{0} & Z \end{pmatrix}$$

et  $Y_2$  la matrice déduite de  $Y_1$  en permutant les deux premières lignes et colonnes de  $Y_1$ .  $Y_1$  et  $Y_2$  ont le même déterminant que la matrice  $Y$  avec

$$Y = \begin{pmatrix} \mathbf{1} \\ Z \end{pmatrix}$$

et on vérifie donc que  $V^2((P_i)) = \det (Y_1 {}^t Y_2)$

Les deuxième et troisième égalités sont immédiates en remarquant que les matrices traitées sont respectivement  ${}^t Y Y$  et  $Y {}^t Y$ .  $\square$

**Corollaire (GOWER (1982)) :** Si  $Z$  est exprimée dans la base  $\mathcal{Z}$  de ses vecteurs propres et si l'on choisit pour origine de  $E$  le barycentre des points  $P_i$ , de sorte que  $Z_1 = 0$ , alors

$$V^2((P_i)) = \frac{1}{[(n-1)!]^2} \prod_{s=1}^{n-1} \lambda_s$$

**Preuve :** Il suffit de remarquer que  $V_Z$  est égal à la matrice diagonale  $\Lambda$  de terme général  $\lambda_s$  lorsqu'on fait le choix d'une telle base. En particulier le volume est nul si le simplexe est dégénéré.  $\square$

L'intérêt de la proposition 7 pour notre étude résulte des deux lemmes suivants :

**Lemme 1 :** Soit  $A$  une matrice carrée régulière décomposée en blocs

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

telle que  $\det(A_{11}) \neq 0$ , alors

$$\det(A) = \det(A_{11}) \det(A_{22} - A_{21} A_{11}^{-1} A_{12})$$

**Preuve :** Elle résulte de la décomposition suivante, qui joue un rôle central en régression linéaire multiple et en Analyse Factorielle Classique :

$$A = \begin{pmatrix} I & 0 \\ A_{21} A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21} A_{11}^{-1} A_{12} \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix} \quad \square$$

**Lemme 2 :** Soit  $A$  une matrice carrée régulière décomposée en blocs. Posons

$$A_{22.1} = A_{22} - A_{21} A_{11}^{-1} A_{12}$$

alors

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} (I + A_{12} A_{22.1}^{-1} A_{21} A_{11}^{-1}) & -A_{11}^{-1} A_{12} A_{22.1}^{-1} \\ -A_{22.1}^{-1} A_{21} A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix}$$

**Preuve :** Elle découle de façon immédiate de la décomposition de  $A$  utilisée dans la démonstration du lemme 1. Puisque si  $A$  est régulière,  $A_{22.1}$  est inversible et donc la matrice diagonale par bloc qui intervient dans l'expression de  $A$  est inversible. Formellement

$$A = U \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix} {}^tU \Rightarrow A^{-1} = U' \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22.1}^{-1} \end{pmatrix} {}^tU'$$

où

$$U = \begin{pmatrix} I & 0 \\ A_{21} A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix} \Rightarrow U' = \begin{pmatrix} I & -A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix} \quad \P$$

On vérifie alors la proposition suivante, par application directe du lemme 2, en permutant le rôle de  $A_{11}$  et  $A_{22}$

**Proposition 8 :** Soit  $C = \begin{pmatrix} 0 & {}^t\mathbf{1} \\ \mathbf{1} & H \end{pmatrix}$  où  $H$  est une matrice inversible. Alors

$$C^{-1} = \begin{bmatrix} \frac{-1}{{}^t\mathbf{1}H^{-1}\mathbf{1}} & \frac{{}^t\mathbf{1}H^{-1}}{{}^t\mathbf{1}H^{-1}\mathbf{1}} \\ \frac{H^{-1}\mathbf{1}}{{}^t\mathbf{1}H^{-1}\mathbf{1}} & H^{-1} - \frac{H^{-1}\mathbf{1}{}^t\mathbf{1}H^{-1}}{{}^t\mathbf{1}H^{-1}\mathbf{1}} \end{bmatrix} \quad (5)$$

Nous utilisons de plus dans la suite le lemme suivant, dont l'intérêt est lié aux deux précédents

**Lemme 3 :** Soient  $A$  et  $B$  deux matrices carrées régulières d'ordre respectif  $p$  et  $k$  et soit  $C$  une matrice réelle d'ordre  $p \times k$ . Alors

$$(A + CB^tC)^{-1} = A^{-1}(I - C(B^{-1} + {}^tCA^{-1}C)^{-1} {}^tCA^{-1})$$

**Preuve :** Elle résulte simplement d'une vérification de l'identité

$$(A + CB^tC)(A + CB^tC)^{-1} = I \quad \P$$

L'expression du volume du simplexe en fonction des distances résulte de l'application de l'identité (4). Posons pour cela  $H$  la matrice carrée d'ordre  $n$  de terme général, le carré des dissimilarités

$$h_{ij} = \delta_{ij}^2$$

$H$  et  $\mathbf{1}$  définissent complètement une forme 2-affine

**Proposition 9 :**

$$V^2((P_i)) = \frac{(-1)^n}{2^{n-1} [(n-1)!]^2} \det \begin{pmatrix} 0 & {}^t\mathbf{1} \\ \mathbf{1} & H \end{pmatrix}$$

**Preuve (BLUMENTHAL (1953)) :** La relation (4) donne

$$w_{ij} = \frac{1}{2} (w_{ii} + w_{jj} - h_{ij})$$

par suite l'expression i) de  $V((P_i))$  donnée dans la proposition 9 conduit au résultat en retranchant à la ligne i (resp. la colonne j) de la matrice  $\begin{pmatrix} 0 & {}^t\mathbf{1} \\ \mathbf{1} & W_Z \end{pmatrix}$  la première ligne (resp. colonne) multipliée par  $\frac{1}{2}w_{i,i}$  pour  $i = 2, 3, \dots, n$ . On obtient le résultat après simplification. ¶

**Remarque :** Si  $n = 2$ ,

$$H = \begin{pmatrix} 0 & \delta_{12}^2 \\ \delta_{12}^2 & 0 \end{pmatrix}$$

et donc

$$\det \begin{pmatrix} 0 & {}^t\mathbf{1} \\ \mathbf{1} & H \end{pmatrix} = 2 \delta_{ij}^2$$

on retrouve donc

$$V^2((P_1, P_2)) = \delta_{ij}^2$$

**Corollaire :** Si  $H$  est de plein rang, posons  $k^2 = {}^t\mathbf{1}H^{-1}\mathbf{1}$ , alors

$$V^2((P_i)) = \frac{(-1)^n k^2}{2^{n-1} [(n-1)!]^2} \det(H)$$

**Preuve :** Appliquons le lemme 1 à la matrice  $\begin{pmatrix} 0 & {}^t\mathbf{1} \\ \mathbf{1} & H \end{pmatrix}$ , en permutant le rôle de  $A_{11}$  et  $A_{22}$

on obtient

$$\begin{aligned} \det \begin{pmatrix} 0 & {}^t\mathbf{1} \\ \mathbf{1} & H \end{pmatrix} &= \det(H) \det(-{}^t\mathbf{1}H^{-1}\mathbf{1}) \\ &= -\det(H) \det({}^t\mathbf{1}H^{-1}\mathbf{1}) \end{aligned}$$

d'où le résultat puisque  $k^2 = {}^t\mathbf{1}H^{-1}\mathbf{1}$  est un scalaire. ¶

On montre de façon analogue, à l'aide de l'expression i) de  $V^2((P_i))$  que lorsque  $W_Z$  est inversible

$$V^2(\{P_i\}) = \frac{d^2}{[(n-1)!]^2} \det(W_Z)$$

où

$$d^2 = \begin{bmatrix} 1 & \\ & W_Z^{-1} & \\ & & 1 \end{bmatrix}$$

GOWER (1982) donne une interprétation géométrique de cette égalité et montre la propriété suivante :

**Proposition 10 :** *Le carré de la distance de l'origine de  $\hat{E}$  à l'hyperplan E support du simplexe  $\{P_i ; i \in [n]\}$  a pour expression*

$$h^2 = \left[ \begin{bmatrix} 1 & \\ & W_Z^{-1} & \\ & & 1 \end{bmatrix} \right]^{-1} = \frac{1}{d^2}$$

**Preuve :** Les points  $P_1, P_2, \dots, P_n$  appartiennent à un hyperplan de E. Notons h la distance de l'origine à cet hyperplan. Le volume du simplexe  $(O, P_1, P_2, \dots, P_n)$  est d'après ce qui précède

$$\frac{1}{n!} \det(Z)$$

où Z est une matrice carrée d'ordre n de plein rang. Ainsi, dans le cas  $n = 2$

$$\frac{1}{2} \det \begin{pmatrix} z_1^1 & z_2^1 \\ z_1^2 & z_2^2 \end{pmatrix}$$

est la surface du triangle  $OP_1P_2$ . On l'exprime de façon équivalente comme le demi-produit de la base par la hauteur, c'est à dire par la distance de l'origine au segment  $P_1P_2$ . Pour  $n > 2$ , le segment  $P_1P_2$  se généralise en un simplexe  $\{P_1, P_2, \dots, P_n\}$  et sa longueur en le volume  $V(\{P_i\})$  (cf. DUBIKAJTIS (1977)). Par suite on écrit dans le cas général

$$\frac{1}{n!} \det(Z) = \left[ \frac{1}{n} V(\{P_i\}) \right] \cdot h$$

et par suite

$$\det(W_Z) = [\det(Z)]^2 = [(n-1)! V(\{P_i\}) h]^2$$

Tout point appartenant à l'hyperplan support des  $P_i$  a pour équation  $Z\underline{s}$  où  $\underline{s}$  est un vecteur de  $F^*$  tel que  $\begin{bmatrix} 1 & \\ & \underline{s} & \\ & & 1 \end{bmatrix} = 1$  et donc

$$h^2 = \text{Min} \{ \begin{bmatrix} 1 & \\ & \underline{s} & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & Z & \\ & & M & \\ & & & Z \end{bmatrix} \underline{s} ; \begin{bmatrix} 1 & \\ & \underline{s} & \\ & & 1 \end{bmatrix} = 1 \}$$

$(\lambda, s)$  est donc le minimum de la fonction lagrangienne

$$L(\lambda, s) = {}^t s W_Z s - 2\lambda ({}^t s \underline{1} - 1)$$

et les équations normales conduisent à la solution

$$W_Z s = \lambda \underline{1}$$

$$\lambda_{\text{Min}} = h^2 = {}^t s W_Z s \quad \text{où } s = \lambda_{\text{Min}} W_Z^{-1} \underline{1}$$

c'est à dire

$$1 = {}^t s \underline{1} = h^2 {}^t \underline{1} W_Z^{-1} \underline{1} \Leftrightarrow h^2 = [{}^t \underline{1} W_Z^{-1} \underline{1}]^{-1} = \frac{1}{d^2}$$

ce qui établit le résultat.  $\parallel$

Ces propriétés analytiques permettent alors de montrer les deux résultats suivants, en remarquant que le nombre de points retenus précédemment est arbitraire. Notons donc

$$D(P_1, \dots, P_n) = \det \begin{pmatrix} 0 & {}^t \underline{1} \\ \underline{1} & H \end{pmatrix}$$

où H est la matrice de terme général  $\delta_{ij}^2$

**Théorème 2 (CAYLEY (1928)) :** Soit  $(I, \delta)$  un espace semi-métrique. Tout n-uple de points de I est plongeable isométriquement dans un espace Euclidien  $E = (\mathbb{R}^r, m)$  si pour tout  $s = 2, \dots, n-1$ , le déterminant  $D(P_1, \dots, P_s)$  est soit nul soit du même signe que  $(-1)^s$ . Si  $r < (n-1)$ , alors pour tout  $r < s \leq n-1$

$$D(P_1, \dots, P_s) = 0.$$

**Définition 21 :**  $n$  points d'un espace semi-métrique  $(I, \delta)$  admettent un plongement isométrique irréductible dans un espace affine Euclidien  $E = (\mathbb{R}^r, m)$  de dimension  $r$ , si et seulement si ils n'admettent aucun plongement isométrique dans un espace affine Euclidien de dimension  $r' < r$ .

**Théorème 3 (BLUMENTHAL (1953)) :** Une condition nécessaire et suffisante pour que  $p$  points ( $p \leq n$ ) d'un espace semi-métrique  $(I, \delta)$  admettent un plongement isométrique irréductible dans un espace Euclidien  $E = (\mathbb{R}^r, m)$  de dimension  $r = p-1$  est que pour tout  $k = 2, \dots, (p-1)$ , on ait

$$\text{sign}(D(P_1, \dots, P_k)) = (-1)^k.$$

D'autre part, soit  $C_k$  la matrice carrée définie par

$$C_k = \begin{pmatrix} 0 & \mathbf{1}_k \\ \mathbf{1}_k & H_k \end{pmatrix} \quad \text{où} \quad H_k = (\delta_{ij}^2)_{i,j=1,2,\dots,k}$$

cette matrice vérifie  $D(P_1, \dots, P_k) = \det(C_k)$ . Lorsque  $k=2$ , on obtient  $D(P_1, P_2) = 2\delta_{ij}^2$  de sorte que la condition  $\text{sign}(D(P_1, \dots, P_k)) = (-1)^k$  est vérifiée pour toute matrice réelle carrée, symétrique, à diagonale nulle. Lorsque  $k=3$ , posons

$$A_{ijl} = \delta_{ij} + \delta_{jl} - \delta_{il} \quad \text{pour tout } i, j, l = 1, 2, 3 \quad i \neq j \neq l$$

l'inégalité triangulaire implique que chacun des termes  $A_{ijl}$  est positif.

$$\begin{aligned} \det(C_3) &= \delta_{12}^4 + \delta_{23}^4 + \delta_{13}^4 - 2\delta_{12}^2\delta_{23}^2 - 2\delta_{12}^2\delta_{13}^2 - 2\delta_{13}^2\delta_{23}^2 \\ &= (\delta_{12}^2 + \delta_{23}^2 - \delta_{13}^2)^2 - 4\delta_{12}^2\delta_{23}^2 \\ &= (\delta_{12} - \delta_{23} - \delta_{13})(\delta_{12} - \delta_{23} + \delta_{13})((\delta_{12} + \delta_{23})^2 - \delta_{13}^2) \\ &= -A_{132}A_{213}A_{123}(\delta_{12} + \delta_{23} + \delta_{13}) \end{aligned}$$

c'est à dire  $\det(C_3) \leq 0$ . On a l'égalité à 0 si les trois points sont colinéaires.

Un rôle particulier est d'ailleurs joué par les triplets  $(i, j, k)$  formés par trois points colinéaires dans la recherche d'une axiomatique des espaces euclidiens. Dans le cas général, lorsque  $(I, \delta)$  est un espace semi-métrique, on pose la définition suivante :

**Définition 22 :** Soit  $(I, \delta)$  un espace semi-métrique de cardinal  $n > 2$  et  $(i, j, k)$  un élément de  $I^3$ . Le point  $P_j$  est entre les points  $P_i$  et  $P_k$  si

$$\delta_{ij} + \delta_{jk} = \delta_{ik}$$

On note  $B(i, j, k)$  la relation ternaire, non symétrique d'intermédiarité (ou *betweenness*) ainsi définie sur  $I^3$ .

D'autre part, puisque l'indiciation de  $I$  est arbitraire, on a la propriété suivante

**Propriété 5 :** Soit  $(I, \delta)$  un espace métrique. Si  $I$  est de cardinal  $n > 3$ , alors pour tout triplet  $(i, j, k)$  d'éléments de  $I$  de points non colinéaires, on a  $\text{sign}(D(P_i, P_j, P_k)) = -1$ .

Réciproquement, cela signifie que le théorème de Blumenthal fournit une condition suffisante applicable à tout espace quasi-métrique  $(I, \delta)$ .

La propriété précédente montre que la propriété la plus faible permettant de particulariser les espaces Euclidiens dans la classe des espaces semi-métriques et exprimée en termes de volume, porte sur les quadruplets  $(i,j,k,l)$  d'éléments de  $I$ .

**Définition 23 :** Soit  $(I, \delta)$  un espace métrique de cardinal  $n > 3$ .  $(I, \delta)$  est dit Ptolémien si tout quadruplet  $(i,j,k,l)$  de  $I^4$  formé de points deux à deux distincts vérifie l'inégalité

$$\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} \geq \delta_{il}\delta_{jk}$$

C'est dire que les produits de distances "opposées" vérifient l'inégalité triangulaire. On vérifie alors par un calcul élémentaire que cette propriété s'exprime de façon équivalente par la condition  $\det(H_4) \leq 0$ , où  $H_4$  est la sous matrice de  $H$  formée sur le quadruplet  $(i,j,k,l)$ , puisque ce déterminant s'écrit

$$- A_{ijkl} A_{ilkj} A_{ikjl} (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$$

où nous avons posé

$$A_{ijkl} = \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}$$

Cette inégalité de Ptolémée doit être comparée à l'identité quadrangulaire qui caractérise les métriques d'arbre additif (cf. supra) par la satisfaction d'une inégalité ultramétrique sur les sommes de distances "opposées". Elle ne permet cependant pas de particulariser les espaces Euclidiens, puisque (cf. BLUMENTHAL et MENGER (1970)) les espaces hyperboliques sont aussi Ptolémiens. Nous introduisons donc une propriété plus exigeante:

**Définition 24 :** Soit  $(I, \delta)$  un espace semi-métrique de cardinal  $n > 3$ .  $(I, \delta)$  vérifie la propriété Euclidienne des quatre points si et seulement si tout quadruplet  $(i,j,k,l)$  de  $I^4$  admet un plongement isométrique dans un espace affine de dimension au plus égale à 3.

En particulier si  $(I, \delta)$  est un espace métrique, il vérifie la propriété Euclidienne des quatre points si et seulement si tout quadruplet  $(i,j,k,l)$  de  $I^4$  vérifie

$$\text{sign}(D(P_i, P_j, P_k, P_l)) = -1$$

ou bien est formé de points coplanaires, auquel cas  $D(P_i, P_j, P_k, P_l) = 0$ . CRIPPEN (1977, 1981) étudie en détails les contraintes imposées au carré de la distance séparant tout couple de points issu d'un tel quadruplet.

**Remarque :** Les théorèmes démontrés par Cayley supposent que  $M$  est la métrique identité. On retrouve ses démonstrations dans le cas d'une métrique  $M$  quelconque de plein rang en posant

$$\underline{x}_i = T \underline{z}_i, \text{ où } T \text{ factorise } M \text{ en } M = {}^t T T$$

La matrice  $H$  qui intervient dans les théorèmes 1 et 2 est par définition caractérisée par le fait que  $h_{ij}$  est le carré d'une fonction de dissimilarité. En particulier cela impose que  $h_{ij}$  soit un nombre réel positif. Lorsqu'on suppose seulement que  $H$  est une matrice réelle symétrique à diagonale nulle, on dispose encore de théorèmes d'existence de plongements isométriques hermitiens du couple  $(I, \delta)$  où  $\delta$  s'interprète comme une fonction de pseudo-dissimilarité, ne vérifiant pas l'axiome de positivité imposé plus haut. Notons alors  $E_{r,s}$  un espace affine de dimension  $p=r+s$  caractérisée par  $r$  coordonnées réelles et  $s$  coordonnées imaginaires pures. Rappelons de plus que toute matrice réelle symétrique d'ordre  $n$  et de rang  $p$  peut être décomposée en

$$W = U D_\epsilon {}^t U$$

où  $U \in \mathcal{M}_{n,p}(\mathbb{R})$  est de rang  $p$  et la matrice diagonale  $D_\epsilon$  s'écrit (à une permutation près des colonnes de  $U$ )

$$D_\epsilon = \begin{pmatrix} I_r & 0 \\ 0 & -I_s \end{pmatrix}$$

Cette décomposition n'est pas unique, mais d'après les résultats de Sylvester, on sait que la signature  $r-s$  est elle invariante. En effet pour toute matrice  $A$  telle que  $\text{rang}(AW^t A) = \text{rang}(W)$ ,  $W$  et  $AW^t A$  ont même signature, puisque les  $A(UD_\epsilon {}^t U) {}^t A$  et  $(AU)D_\epsilon {}^t (AU)$  sont égales et donc  $\text{rang}(AU) = \text{rang}(U)$ . Par suite, si la matrice  $W_Z$  associée au plongement de  $(I, \delta)$  dans  $E_{r,s}$  vérifie

$$W_Z = Z D_\epsilon {}^t Z$$

où  $Z$  appartient à  $\mathcal{M}_{n,p}(\mathbb{R})$ . Alors

$${}^t \underline{x} W_Z \underline{x} = \sum_{k=1}^r ({}^t \underline{z}^k \underline{x})^2 - \sum_{k=r+1}^{r+s} ({}^t \underline{z}^k \underline{x})^2$$

et en posant  $x_i = +1$ ,  $x_j = -1$  et  $x_l = 0$  ( $i, j \neq l$ ), la relation (4) permet de vérifier que :

$$\delta_{ij}^2 = \sum_{k=1}^r (z_i^k - z_j^k)^2 - \sum_{k=r+1}^{r+s} (z_i^k - z_j^k)^2$$

COXETER (1933) montre alors le théorème suivant, présenté ici sans démonstration. Notons  $\mathcal{H}$  le sous-ensemble de  $\mathcal{M}_{n,n}(\mathbb{R})$  formé des matrices symétriques à diagonale nulle.

**Théorème 4 :** Soient  $I$  un ensemble de cardinal fini  $n$  et  $H$  un élément de  $\mathcal{H}$ . Alors il existe un espace affine  $E_{r,s}$  avec  $r+s=p \leq n-1$  et  $n$  points  $(P_i)$  dans  $E_{r,s}$  tels que pour tout couple  $(i,j)$  d'éléments de  $I$  on a

$$\| \overrightarrow{P_i P_j} \|^2 = h_{ij}$$

La détermination de la dimension  $p=r+s$  de l'espace  $E_{r,s}$  permettant un plongement isométrique irréductible de  $(I, \delta)$  suit en fait les mêmes voies que précédemment puisque COXETER déduit du résultat précédent le théorème suivant, qui montre qu'il convient de chercher  $p=(m-1)$  correspondant au sous ensemble  $J$  de  $I$  de cardinal  $m$  ( $m < n$ ) le plus grand pour lequel on obtient un déterminant non nul :

**Théorème 5 :** Soit  $I$  un ensemble de cardinal  $n$  et  $H$  un élément de  $\mathcal{H}$ . Si

$$\det \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & H \end{pmatrix} \neq 0$$

alors il existe un espace affine  $E_{r,s}$  avec  $r+s=n-1$  et un simplexe  $(P_i, i = 1, 2, \dots, n)$  dans  $E_{r,s}$  tels que

$$\| \overrightarrow{P_i P_j} \|^2 = h_{ij}$$

Nous étudions maintenant une deuxième classe de théorèmes d'existence fondés sur les propriétés des formes quadratiques définies sur un sous-espace affine de  $\vec{F}^*$ . Nous formulons ainsi dans un langage moderne des travaux entrepris par GAUSS (1831) et DIRICHLET (1850) dans le cas d'un espace de dimension 3 puis généralisés au cas d'un espace de dimension quelconque par HERMITE (1850) et MINKOWSKI (1891). FRECHET (1935) est en fait à l'origine de la formulation moderne du problème en termes de plongement Euclidien (ou Hilbertien) d'un espace quasi-métrique  $(I, \delta)$ .

Supposons que  $I$  est de cardinal fini  $n > 3$ ,  $I$  est en bijection avec  $[n]$ . Pour chaque choix d'une bijection, posons  $W^n$  la matrice carrée d'ordre  $(n-1)$ , de terme général

$$w_{ij}^n = -\frac{1}{2} [ \delta_{ij}^2 - \delta_{in}^2 - \delta_{jn}^2 ]$$

Dans  $F^*$ ,  $W^n$  est interprétée comme la matrice associée à une forme quadratique définie sur le sous-espace affine support des vecteurs  $\overrightarrow{N_n N_i}$ ,  $i = 1, \dots, (n-1)$  où  $N_i$  est le point de coordonnées  $z_i$  puisque, dans le référentiel  $(N_n, \{(f_i^* - f_n^*), i = 1, \dots, (n-1)\})$

$$\begin{aligned} w_{ij}^n &= w(f_i^* - f_n^*, f_j^* - f_n^*) \\ &= w(\overrightarrow{N_n N_i}, \overrightarrow{N_n N_j}) \end{aligned}$$

SCHOENBERG (1935,1938) démontre le théorème suivant, démontré de façon indépendante, semble t'il, par YOUNG et HOUSEHOLDER (1938). Nous reprenons ici la démonstration de WELLS et WILLIAMS (1970).

**Théorème 6 :** *L'espace quasi-métrique  $(I, \delta)$  où  $I$  est de cardinal fini  $n > 3$  est plongeable dans un espace affine Euclidien de dimension  $(n-1)$  si et seulement si  $W^n$  est une forme quadratique positive, semi-définie, c'est à dire telle que*

$$\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} w_{ij}^n x_i x_j \geq 0 \quad \forall (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1} \quad (6)$$

**Preuve :** Démontrons d'abord la condition nécessaire. Supposons le plongement possible dans un espace affine Euclidien  $E$  de dimension  $(n-1)$ . Les  $n$  points  $P_i, i \in I$  ont alors pour vecteurs de coordonnées dans le référentiel  $(O, \{e_j, j = 1, \dots, n-1\})$  les vecteurs  $z_i$ , de sorte que pour tout couple  $(i,j)$  de  $I \times I$

$$\|z_i - z_j\|_M = \delta_{ij} \quad (7)$$

On peut fixer l'origine en  $P_n$  de sorte que  $z_n = 0$ . Remarquons que

$$z_i - z_j = (z_i - z_n) + (z_n - z_j)$$

et donc on obtient par passage au carré

$$\langle z_i, z_j \rangle_M = w_{ij}^n$$

Par suite

$$0 \leq \left\| \sum_{i=1}^{n-1} x_i z_i \right\|_M^2 = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \langle z_i, z_j \rangle_M x_i x_j$$

Pour la condition suffisante, remarquons que si  $w^n$  est une forme quadratique semi-définie positive, alors  $W^n$  est une matrice semi-définie positive sur  $\mathbb{R}^{n-1}$  et donc si nous l'interprétons comme la matrice d'une application linéaire, définie sur un sous-espace vectoriel libre  $\vec{F}_n^*$ , de dimension  $(n-1)$

$$w^n : \vec{F}_n^* \rightarrow \vec{F}_n^*$$

et

$$w^n(x, x) \geq 0 \text{ pour tout } x \in \vec{F}_n^*$$

Il existe donc une matrice réelle régulière  $U$  qui factorise  $W^n$  en  $W^n = {}^t U U$  et un entier  $r$  tel que si on pose  $y = Ux$

$$w^n(\underline{x}, \underline{x}) = {}^t \underline{x} {}^t U U \underline{x} = {}^t \underline{y} \underline{y} = \sum_{s=1}^r (y_s)^2 = \|P_r \underline{x}\|^2 \quad (8)$$

Nous interprétons alors  $U$  comme la matrice d'une application linéaire  $u$  définie sur  $\vec{F}_n^*$  et à valeurs dans un espace vectoriel libre de dimension  $(n-1)$  noté  $\vec{E}$ , pour un choix approprié d'une base de  $\vec{E}$ .  $r$  est le rang de  $W^n$  et de la matrice  $P_r$  associée au projecteur orthogonal de  $\vec{F}_n^*$  sur le sous-espace  $\vec{E}_r$  de  $\vec{E}$  de dimension  $r$  et engendré par les vecteurs lignes de  $U$ . L'image par  $U$  de chaque vecteur de base  $\vec{f}_i^*$  est

$$\underline{z}_i = U \vec{f}_i^* \quad i = 1, 2, \dots, n-1$$

Posons donc successivement  $\underline{x} = \underline{z}_i$   $i = 1, 2, \dots, n-1$  puis  $\underline{x} = \underline{z}_i - \underline{z}_j$   $i, j = 1, 2, \dots, n-1$  dans (8). On obtient

$$\delta_{nj}^2 = w^n(\vec{f}_i^*, \vec{f}_i^*) = \|P_r \underline{z}_i\|^2 \quad i = 1, 2, \dots, n-1$$

$$\delta_{ij}^2 = w^n(\vec{f}_i^* - \vec{f}_j^*, \vec{f}_i^* - \vec{f}_j^*) = \|P_r(\underline{z}_i - \underline{z}_j)\|^2 \quad 1 \leq i, j \leq n-1$$

Par suite, on a construit une application linéaire  $u$  qui envoie  $\vec{f}_n^*$  en  $O$  et  $\vec{f}_i^*$  en  $P_r \underline{z}_i$  de sorte que  $u$  est une isométrie de  $\vec{F}_n^*$  sur  $\vec{E}_r$ .  $\parallel$

WELLS et WILLIAMS (1970) montrent que le théorème 6 a pour conséquence directe de montrer que les espaces de suite  $\ell_n^p$ ,  $1 \leq p \leq \infty$ , ne peuvent pas être plongés dans un espace Euclidien si  $n \geq 2$  et  $p \neq 2$ . Rappelons en effet que  $\ell_n^p$  désigne l'espace vectoriel normé  $(\mathbb{R}^n, \|\cdot\|_p)$  où

$$\|\underline{x}\|_p = \begin{cases} \left[ \sum_{j=1}^n |x_j|^p \right]^{\frac{1}{p}} & \text{si } p < +\infty \\ \text{Max}(|x_j|; 1 \leq j \leq n) & \text{si } p = +\infty \end{cases}$$

Un contre-exemple est fourni pour  $n=2$ . Si  $1 \leq p < 2$ , on considère les quatre points

$$\underline{z}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \underline{z}_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \underline{z}_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \underline{z}_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

On calcule alors

$$W^4 = \begin{pmatrix} 4 & 2^{1+\frac{2}{p}} & 4 - 2^{1+\frac{2}{p}} \\ 2^{1+\frac{2}{p}} & 2^{2+\frac{2}{p}} & 2^{1+\frac{2}{p}} \\ 4 - 2^{1+\frac{2}{p}} & 2^{1+\frac{2}{p}} & 4 \end{pmatrix}$$

Par suite, si  $x_1 = x_3 = 1$  et  $x_2 = -1$ , la forme quadratique (6) s'écrit

$$\sum_{i=1}^3 \sum_{j=1}^3 w_{ij}^4 x_i x_j = 2^3 (2 - 2^{\frac{2}{p}}) < 0$$

Pour  $2 < p \leq +\infty$ , le contre exemple est fourni par les mêmes  $x_i$  et par les quatre points

$$z_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad z_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad z_3 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad z_4 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

MORGAN (1974) montre ce résultat à l'aide d'une caractérisation des espaces affines Euclidiens en termes de platitude. Un espace métrique  $(I, \delta)$  est plat, si pour tout  $n$ -simplexe  $\{P_i; i \in [n]\}$ , ( $n \leq \text{card}(I)$ ), le volume  $V(\{P_i\})$  est un nombre réel. On vérifie ainsi que les 3-simplexes définis par les deux contre-exemples ci-dessus ont un volume imaginaire.

La démonstration du théorème 6 démontre en fait un résultat un peu plus fort

**Corollaire 1 :** *Un espace métrique fini  $(I, \delta)$  admet un plongement isométrique irréductible de  $I$  dans un espace affine Euclidien de dimension  $r$  si et seulement si la forme quadratique définie par (6) est semi-définie positive de rang  $r$ .*

La démonstration des deux résultats précédents est fondée sur les propriétés de la forme quadratique  $W^n$ . Cependant  $W^n$  fait jouer un rôle particulier à l'un des éléments de  $I$ , noté  $n$  ici. La formulation de cet énoncé peut être modifiée en considérant la forme quadratique de matrice

$$W = (w_{ij})_{i,j} \quad \text{avec} \quad w_{ij} = -\frac{1}{2} [\delta_{ij}^2 - \delta_{i.}^2 - \delta_{.j}^2 + \delta_{..}^2]$$

où

$$\delta_{i.}^2 = \frac{1}{n} \sum_{j=1}^n \delta_{ij}^2 \quad \delta_{.j}^2 = \frac{1}{n} \sum_{i=1}^n \delta_{ij}^2 \quad \delta_{..}^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_{ij}^2$$

On choisit alors pour origine des axes dans  $\vec{F}_n^*$ , le vecteur  $\underline{1}$  de coordonnées constantes et égales à 1. Dans  $\vec{E}$ , ceci est équivalent à retenir pour origine  $O$  le barycentre des points massiques  $\{(P_i, \frac{1}{n}); i \in [n]\}$ . TORGERSON (1952) attribue à Tucker, Green et Abelson la formulation du théorème fondée sur les propriétés de définie positivité de la forme quadratique  $W$ . Nous étudions plus bas les liens existant entre les analyses fondées sur  $W$  et sur  $W^n$  dans un cadre un peu plus général.

Les résultats précédents sont exprimés en fonction des matrices non observées  $W$  et  $W^n$ . On peut préférer définir des conditions d'Euclidéanité sous une forme plus directement liée aux observations. Pour cela, notons  $\mathcal{Y}_n$  le sous-ensemble de  $\mathcal{M}_{n,n}(\mathbb{R})$  formé des matrices symétriques d'ordre  $n$ . Tout élément  $H \in \mathcal{Y}_n$  est interprété comme la matrice associée à une forme quadratique définie sur l'espace vectoriel  $\overrightarrow{F}_n^*$  de dimension  $n$  défini plus haut. On considère de plus les contrastes de  $\overrightarrow{F}_n^*$ , c'est à dire les vecteurs  $\underline{c}$  de  $\overrightarrow{F}_n^*$  vérifiant

$$\sum_{i=1}^n c_i = 0$$

**Définition 25 :** La matrice  $H$  de  $\mathcal{Y}_n$  est dite conditionnellement négative définie si pour tout contraste  $\underline{c}$  de  $\overrightarrow{F}_n^*$ , on a

$$\underline{c}^t H \underline{c} = \sum_{i=1}^n \sum_{j=1}^n c_i h_{ij} c_j \leq 0$$

Lorsque la matrice  $H$  est la matrice associée à un espace semi-métrique par l'identité

$$h_{ij} = \delta_{ij}^2$$

on a clairement  $H \in \mathcal{Y}_n$ . On vérifie alors la conséquence suivante du théorème 6

**Corollaire 2 :** L'espace métrique fini  $(I, \delta)$  où  $\text{card}(I) = n > 3$  admet un plongement isométrique dans un espace affine Euclidien de dimension au plus égale à  $(n-1)$  si et seulement si la matrice  $H = (\delta_{ij}^2)$  est conditionnellement négative définie.

**Preuve :** Soit  $(x_1, x_2, \dots, x_{n-1})$  un élément quelconque de  $\mathbb{R}^{n-1}$ . On construit un contraste  $\underline{c}$  de  $\overrightarrow{F}_n^*$  en posant :

$$x_n = - \sum_{j=1}^{n-1} x_j$$

l'équation (6) devient alors

$$\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \delta_{ij}^2 x_i x_j + 2 x_n \sum_{i=1}^{n-1} \delta_{ij}^2 x_i \leq 0$$

soit de façon équivalente

$$\sum_{i=1}^n \sum_{j=1}^n \delta_{ij}^2 x_i x_j \leq 0 \quad \forall$$

Le corollaire 2 joue un rôle central dans l'étude du plongement métrique d'un espace quasi-métrique quelconque.

**Remarque :** Etant donné  $(I, \delta)$  et l'espace vectoriel libre  $F^*$ , nous avons construit une représentation de  $I$  dans  $F^*$  en associant à tout élément  $i$  de  $I$  le point  $N_i$ , extrémité du vecteur de base  $f_i^*$ . Regardons alors  $F^*$  comme un espace affine de direction  $\vec{F}^*$  muni du système  $(O, \{f_i^*, i \in [n]\})$ , de sorte que les structures affines de  $F^*$  et vectorielles de  $\vec{F}^*$  soient confondues. C'est pourquoi nous ne distinguons pas ces deux espaces et nous les notons tous deux  $F^*$  dans la suite.

Lorsque nous munissons  $F^*$  d'une géométrie euclidienne définie par la forme quadratique  $h$  telle que

$$h(\overrightarrow{N_i N_j}, \overrightarrow{N_i N_j}) = \delta_{ij}^2 \quad \forall i, j \in [n]$$

Puisque

$$h(\overrightarrow{N_i N_j}, \overrightarrow{N_i N_j}) = h(\overrightarrow{ON_i}, \overrightarrow{ON_i}) + h(\overrightarrow{ON_j}, \overrightarrow{ON_j}) - 2h(\overrightarrow{ON_i}, \overrightarrow{ON_j})$$

on est amené à prendre pour matrice associée à  $h$  dans cette base, la matrice  $H$  de terme général

$$H_{ij} = h(\overrightarrow{ON_j}, \overrightarrow{ON_j}) = -\frac{1}{2} \delta_{ij}^2 \quad \forall i, j \in [n]$$

Le support  $F_1^*$  du nuage de points  $((N_i, p_i); i \in [n])$  est de dimension  $(n-1)$  et l'on restreint  $H$  à  $F_1$  en construisant

$$W = Q_1 H^t Q_1 = -\frac{1}{2} Q_1 \Delta^{*2} Q_1$$

où

$$Q_1 = I - \frac{1}{n} \underline{1} \underline{1}^t$$

est le projecteur sur le supplémentaire orthogonal  $F_1$  de la droite des constantes de  $F$ .  $F_1^*$  est donc un hyperplan de  $F^*$  qui joue le rôle du sous-espace affine recherché. On observe alors que  $W$  est la matrice utilisée par TORGERSON (1958) et le corollaire 2 s'exprime pour  $H$  définie par (6) par une condition de définie négativité conditionnelle. De plus  $W$  vérifie par construction

$$W \underline{1} = 0$$

et donc si  $W$  est semi-définie positive, elle est factorisable en

$$W = {}^t Z Z$$

où  $Z\mathbf{1} = 0$  et donc la matrice  $Y = \begin{pmatrix} \mathbf{1} \\ Z \end{pmatrix}$  définit le simplexe  $\{P_i ; i \in [n]\}$  utilisé dans notre présentation des résultats de Cayley et Blumenthal.

**Définition 26 :** *Un espace quasi-métrique  $(X, \delta)$  est de type négatif si le sous-espace quasi-métrique  $(I, \delta)$  défini par la restriction de  $\delta$  à tout sous-ensemble  $I$  de  $X$  induit une matrice  $H = (\delta_{ij})$  conditionnellement négative.*

Considérons alors les transformations continues des dissimilarités  $\delta$ . Notons  $f$  une telle transformation et supposons de plus que  $f$  est définie sur  $\mathbb{R}^+$  et à valeurs dans  $\mathbb{R}^+$  telle que  $f(0)=0$ .

**Définition 27 :** *Soit  $(X, \delta)$  un espace quasi-métrique et  $f$  une transformation continue de métrique. L'espace quasi-métrique  $(X, \tau)$  où  $\tau = f \circ \delta$ , est appelé le  $f$ -transformé de  $(X, \delta)$ . Si  $(X, \tau)$  est de type négatif, la fonction  $f$  est dite de type négatif sur  $X$ . On note  $\mathcal{N}(X)$  la classe de telles fonctions.*

L'étude des propriétés de  $\mathcal{N}(X)$  a été entreprise par SCHOENBERG (1938). Nous donnons ici des résultats plus généraux développés par WELLS et WILLIAMS (1970). Ces auteurs montrent que l'étude de  $\mathcal{N}(X)$  est liée à la notion de fonction radiale positive définie sur  $X$ .

**Définition 28 :** *Une fonction  $f$  définie sur  $\mathbb{R}^+$  est radiale positive définie sur un espace quasi-métrique  $(X, \delta)$  si  $f$  est continue et vérifie pour toute partie finie  $I$  de  $X$  ( $\text{card}(I) = n$ ,  $n \geq 1$ ) et tout  $x$  de  $\mathbb{R}^n$*

$$\sum_{i=1}^n \sum_{j=1}^n f(\delta_{ij}) x_i x_j \geq 0$$

On note  $\text{RPD}(X)$  la classes des fonctions radiales positives définies. WELLS et WILLIAMS (1970) vérifient alors le théorème suivant, qui généralise les résultats de SCHOENBERG (1938)

**Théorème 7 :** *Soit  $(X, \delta)$  un espace quasi-métrique. Les trois propositions suivantes sont équivalentes*

- i)  $\delta$  est de type négatif
- ii) la fonction  $f(t) = \exp(-\lambda t^2)$  appartient à  $\text{RPD}(X)$  pour tout  $\lambda$  réel positif
- iii)  $(X, \delta)$  admet un plongement hilbertien.

La relation entre  $\mathcal{N}(X)$  et  $\text{RDP}(X)$  est alors donnée par le résultat suivant démontré par WELLS et WILLIAMS (1970)

**Théorème 8 :** *Soit  $(X, \delta)$  un espace quasi-métrique et  $f$  une fonction réelle positive définie sur  $\mathbb{R}^+$ . Les deux propositions suivantes sont équivalentes*

- i)  $f \in \mathcal{N}(X)$
- ii)  $\exp(-\lambda f^2) \in \text{RDP}(X)$  pour tout  $\lambda$  réel positif.

L'un des problèmes centraux pour l'étude de SCHOENBERG (1937, 1938) est la conservation des propriétés métriques d'un espace  $(I, \rho)$  par les transformations continues de  $\rho$ . La classe de transformations la plus couramment utilisée en pratique est celle des fonctions puissances. Ainsi, nous discutons au paragraphe §V.1.1.3 la méthode de construction d'une représentation Euclidienne d'un espace quasi-métrique proposée par CRITCHLEY (1978). Cet auteur cherche alors une transformation  $\delta \mapsto \delta^\alpha$  des données observées, permettant de construire un plongement Euclidien de  $(I, \delta^\alpha)$ . Une telle démarche peut être justifiée en partie par le résultat suivant dû à WELLS et WILLIAMS (1970)

**Théorème 9 :** *Soit  $f$  un élément de  $\mathcal{N}(X)$ , alors pour tout réel  $\alpha$  tel que  $0 < \alpha < 1$ , on a*

- i)  $f^\alpha \in \mathcal{N}(X)$
- ii)  $\exp(-\lambda f^{2\alpha}) \in \text{RDP}(X)$  pour tout  $\lambda$  réel positif.

On retrouve en particulier, comme corollaire immédiat de ce théorème, le résultat suivant dû à Von Neumann

**Corollaire :** *Si  $H$  est un espace de Hilbert et  $H^\alpha$  est déduit de  $H$  par le choix de la métrique*

$$\|x - y\|_{H^\alpha} = \|x - y\|_H^\alpha$$

alors

- i)  $H^\alpha$  admet un plongement dans  $H$
- ii) si  $\gamma \in \mathbb{R}$  et  $0 < \gamma \leq \alpha \leq 1$ , alors  $H^\gamma$  admet un plongement dans  $H^\alpha$

Par suite on obtient le résultat général suivant :

**Corollaire :** *Soit  $f$  une fonction homogène telle que  $\exp(-f^2)$  appartient à  $\text{RDP}(X)$ . Alors  $\exp(-f^{2\alpha})$  appartient à  $\text{RDP}(X)$  pour tout  $\alpha$  réel,  $0 < \alpha < 1$ , et en particulier  $\exp(-\|x\|^{2\alpha})$  est une fonction positive définie sur l'espace de Hilbert  $H$ .*

**Théorème 10** Soit  $(X, \delta)$  un espace quasi-métrique et  $f$  un élément de  $\mathcal{N}(X)$ , alors pour tout réel  $\alpha$  tel que  $0 < \alpha < 1$ , on a

i)  $f^\alpha \in \mathcal{N}(X)$

ii)  $\exp(-\lambda f^\alpha) \in \text{RDP}(X)$  pour tout  $\lambda$  réel positif.

Notre étude est limitée aux méthodes de représentation Euclidienne d'un espace quasi-métrique. Cependant les résultats ci-dessus illustrent le fait que de nombreux travaux ont été consacrés à l'étude du plongement de tels espaces dans une classe plus large d'espaces vectoriels normés. Dans le cas de la classe des espaces de Minkowski  $L^p$ ,  $1 \leq p \leq 2$ , le résultat essentiel est dû à BRETAGNOLLE, DACUNHA-CASTELLE et KRIVINE (1966)

**Théorème 11:** Un espace vectoriel normé réel  $\mathfrak{B}$  admet un plongement linéaire et isométrique dans  $L^p$ ,  $1 \leq p \leq 2$  si et seulement si  $\|x - y\|^{p/2}$  est de type négatif dans  $\mathfrak{B}$ .

Les théorèmes précédents fournissent des conditions d'existence d'un plongement métrique de l'espace quasi ou semi-métrique  $(I, \delta)$  où  $I$  est de cardinal fini et fixé à priori dans  $\vec{E}$ . Considérons ici, dans un souci de simplicité, le cas Euclidien.  $(I, \delta)$  est alors mis en bijection avec une partie de  $\vec{E}$  de sorte que ces théorèmes répondent aux besoins d'une analyse exploratoire des données. L'étude de  $(I, \delta)$  tire cependant souvent son intérêt d'une visée inférentielle qui nécessite des résultats plus généraux portant sur l'existence d'une congruence entre  $(I, \delta)$  et  $\vec{E}$  et donc en particulier d'une bijection entre  $I$  alors supposé quelconque et  $\vec{E}$  tout entier.

**Définition 29:** Soient  $(I, \delta)$  et  $(J, \rho)$  deux espaces quasi-métriques. Les couples  $(i, i')$  et  $(j, j')$  d'éléments de  $I$  et de  $J$  respectivement sont dits congruents si et seulement si

$$\delta_{ii'} = \rho_{jj'}$$

$I$  et  $J$  sont dits congruents s'il existe une application  $f$  de  $I$  vers  $J$  telle que les couples  $(i, i')$  et  $((f(i), f(i')))$  sont congruents, quelques soient  $i$  et  $i'$ , éléments de  $I$ .

On vérifie que la relation de congruence entre espaces métriques est une relation d'équivalence et que l'application  $f$  est un homéomorphisme.

de LEEUW et HEISER (1982) parlent alors de "problèmes d'espace". Ceux-ci sont abordés par MENGER (1928) et surtout WILSON (1931, 1932, 1935). BLUMENTHAL (1953) reste ici aussi la source principale pour l'étude de ces problèmes.

La difficulté essentielle rencontrée lors de l'établissement de tels résultats d'existence est liée à la vérification de propriétés de densité de l'espace. On est alors amené à définir les notions suivantes, qu'il est difficile de mettre à l'épreuve de données réelles

**Définition 30:** *Un espace semi-métrique  $(I, \delta)$  de cardinal  $n > 3$  est dit*

*(i) convexe si pour tout couple  $(i,k)$  de  $I \times I$ , il existe au moins un élément  $j$  de  $I$  tel que  $B(j;i,k)$ .*

*(ii) extérieurement convexe si pour tout couple  $(i,k)$  de  $I \times I$ , il existe au moins un élément  $j$  de  $I$  distinct de  $i$  et de  $k$  tel que  $B(j;i,k)$ .*

MENGER (1928) montre alors le théorème suivant, qui fait de plus intervenir une notion métrique de complétude, elle aussi difficile à vérifier en pratique.

**Définition 31:** *L'espace semi-métrique  $(I, \delta)$  est métriquement complet si l'existence de deux suites  $\{i_n, n \in \mathbb{N}\}$  et  $\{j_n, n \in \mathbb{N}\}$  de points de  $I$  vérifiant*

$$\lim_{n \rightarrow \infty} \delta_{i_n j_n} = 0$$

*implique l'existence d'un point  $i_0$  de  $I$  tel que*

$$\lim_{n \rightarrow \infty} \delta_{i_n i_0} = 0$$

**Théorème 12:** *L'espace semi-métrique  $(I, \delta)$  de cardinal  $n > 3$  est congruent à l'espace affine Euclidien  $\vec{E}$  de dimension  $p$  si et seulement si il est complet, convexe, extérieurement convexe et s'il admet un plongement isométrique dans  $\vec{E}$ .*

Un résultat plus fort est démontré par WILSON (1932)

**Théorème 13 :** *L'espace semi-métrique  $(I, \delta)$  de cardinal  $n > 3$  est congruent à l'espace affine Euclidien  $\vec{E}$  de dimension  $p$  si et seulement si il est complet, convexe, extérieurement convexe et si de plus*

*i) il vérifie la propriété Euclidienne des quatre points*

*ii) le déterminant  $\det(C_{p+1})$  associé au choix de  $(p+1)$  points quelconques de  $I$  est nul.*

La propriété ii) est exprimée ici en termes de déterminants. On peut clairement l'exprimer de façon équivalente en terme de rang de l'une des matrices  $W^n$  ou  $W$  définies plus haut. Les difficultés rencontrées dans l'évaluation des propriétés de convexité et de convexité externe des espaces semi-métriques ont conduit à rechercher des axiomatiques plus faibles. En particulier KRANTZ, LUCE, SUPPES et TVERSKY (1971) définissent une notion plus faible d'intermédiarité.

**Définition 32:** Soit  $I$  un ensemble de cardinal  $n > 2$  et  $\mathcal{R}$  une relation binaire définie sur  $I \times I$ . Pour tout triplet  $(i, j, k)$  d'éléments de  $I$  on dit que  $j$  est entre  $i$  et  $k$  et on note  $\mathbf{J}(j; i, k)$  si et seulement si

$$(i, k)\mathcal{R}(i, j) \text{ et } (i, k)\mathcal{R}(j, k)$$

Remarquons que si  $I$  est un ensemble de cardinal  $n > 2$ ,  $\mathbf{J}$  est une relation ternaire transitive sur  $I$  si pour tout quadruplet  $(i, j, k, l)$  d'éléments de  $I$ , les deux conditions suivantes sont vérifiées

- i) Pour  $j \neq k$  ( $\mathbf{J}(j; i, k)$  et  $\mathbf{J}(k; j, l)$ )  $\Rightarrow$  ( $\mathbf{J}(j; i, l)$  et  $\mathbf{J}(k; i, l)$ )
- ii) ( $\mathbf{J}(j; i, k)$  et  $\mathbf{J}(k; i, l)$ )  $\Rightarrow$  ( $\mathbf{J}(j; i, l)$  et  $\mathbf{J}(k; j, l)$ )

La relation  $\mathbf{J}$  fournit une forme faible de convexité. ORTH(1980) pose alors la définition suivante dans laquelle le terme non conditionnel traduit le fait que  $\mathcal{R}$  est une relation totale sur  $I \times I$  :

**Définition 33:** Soient  $I$  un ensemble de cardinal  $n > 2$  et  $\mathcal{R}$  une relation binaire définie sur  $I \times I$ . Le couple  $(I \times I, \mathcal{R})$  est une structure de dissimilarité non conditionnelle finie unidimensionnelle si et seulement si

- i)  $\mathcal{R}$  est une relation totale, symétrique et transitive sur  $I \times I$
- ii)  $\mathcal{R}$  est une relation propre, c'est à dire telle  $(i, j)\mathcal{R}(i, i), \forall (i, j) \in I \times I$
- iii) La relation  $\mathbf{J}$  induite par  $\mathcal{R}$  est transitive.

La définition 33 est illustrée par la relation  $\mathcal{R}$  associée à une fonction de dissimilarité propre  $\delta$  définie sur  $I \times I$ . Dans ce cas, la matrice  $\Delta$  de terme général  $\delta_{ij}$  a une forme particulière du fait de la condition iii). Dans ce cas, en effet, il existe une bijection de  $I$  dans  $[n]$  telle que la relation d'intermédiation se traduise par une croissance des valeurs de  $\delta_{ij}$  en fonction de  $i$  et de  $j$ . On retrouve ainsi une propriété très étudiée en relation avec les problèmes de sériation (cf. GAUD (1983), FICHET (1984), HUBERT (1985)).

**Définition 34 :** Une matrice carrée  $\Delta$  d'ordre  $n$  est de Robinson si et seulement si tout triplet  $(i, j, k)$  d'éléments de  $I$  tel que  $1 \leq i \leq j \leq k \leq n$  vérifie

$$\delta_{ij} \leq \delta_{ik} \text{ et } \delta_{jk} \leq \delta_{ik}$$

Elle est fortement de Robinson si de plus

$$(\delta_{ik} = \delta_{ij}) \Rightarrow (\delta_{il} = \delta_{il} \quad \forall l \geq k)$$

$$(\delta_{ik} = \delta_{ij}) \Rightarrow (\delta_{lj} = \delta_{lk} \quad \forall l \leq i)$$

ORTH (1980) démontre alors le théorème suivant :

**Théorème 14 :** Soit  $I$  un ensemble de cardinal  $n > 2$  et  $\mathcal{R}$  une relation binaire définie sur  $I \times I$ . Alors le couple  $(I \times I, \mathcal{R})$  est une structure de dissimilarité non conditionnelle finie unidimensionnelle si et seulement si il existe une fonction réelle  $\varphi$  définie sur  $I$  et une métrique  $\rho$  sur  $\varphi(I)$  telles que tout quadruplet  $(i, j, k, l)$  d'éléments de  $I$  vérifie

$$i) (i, j)\mathcal{R}(k, l) \text{ si et seulement si } \rho(\varphi(i), \varphi(j)) \geq \rho(\varphi(k), \varphi(l))$$

ii)  $\rho$  est topologiquement équivalente à la métrique Euclidienne définie sur  $\varphi(I)$ .

De plus si  $(\varphi', \rho')$  et  $(\varphi, \rho)$  sont deux solutions, il existe une fonction strictement monotone  $m$  telle que

$$\rho' = m \circ \rho$$

Le travail de ORTH est limité au cas des représentations Euclidiennes unidimensionnelles. Il se particularise cependant par des axiomes aisés à vérifier en pratique. Dans le cas multidimensionnel, on doit faire appel aux travaux de BEALS et KRANTZ (1967), KRANTZ (1968) et leur extension due à LEW (1975). La validité des résultats repose alors sur des hypothèses de densité de l'espace, qu'il n'est pas possible de tester en pratique.

**Définition 35:** Soit  $I$  un ensemble de cardinal  $n > 2$  et  $\mathcal{R}$  une relation binaire définie sur  $I \times I$ . Pour tout triplet  $(i, j, k)$  d'éléments de  $I$ , on dit que  $j$  est entre  $i$  et  $k$  et on note  $\mathcal{K}(j; i, k)$  si et seulement si pour tout élément  $m$  de  $I$

$$i) ((i, j)\mathcal{R}(i, k) \text{ et } (i, k)\mathcal{R}(i, m)) \Rightarrow ((k, l)\mathcal{R}(j, m))$$

$$ii) \text{ si } i) \text{ est vrai et si } (j, m) \sim (k, l) \text{ alors } (i, j) \sim (i, k) \text{ et } (i, m) \sim (i, l)$$

$$\text{où } ((i, j) \sim (k, l)) \Leftrightarrow ((i, j)\mathcal{R}(k, l) \text{ et } (k, l)\mathcal{R}(i, j))$$

On pose de plus  $\mathcal{H}(j; i, k)$  si et seulement si  $(\mathcal{K}(j; i, k) \text{ et } \mathcal{K}(j; k, i))$

Notons  $\mathcal{B}$  la relation ( $\mathcal{R}$  et non  $\sim$ ). Alors ces auteurs montrent que les conditions de densité impliquent l'existence de points équidistants de deux points au sens où si  $i \neq j$ ,  $(i, j)\mathcal{B}(i, l)$  implique l'existence d'un point  $k$  tel que  $(i, k) \sim (i, j)$  et  $\mathcal{H}(k; i, l)$ . Sous ces hypothèses, on montre alors l'existence d'une métrique  $\delta$  sur  $I$  telle que

$$i) (i, j)\mathcal{R}(k, l) \text{ si et seulement si } \delta_{ij} \leq \delta_{ik}$$

$$ii) \mathcal{H}(k; i, l) \text{ si et seulement si } \delta_{ik} + \delta_{kl} = \delta_{il}$$

Le problème se ramène alors au plongement Euclidien irréductible de  $(I, \delta)$ . ORTH (1980) remarque cependant que la faiblesse essentielle de l'axiomatique de Beals, Krantz et Tversky est d'imposer une structure de produit cartésien à  $I$ , supposant connue la dimension  $p$  afin d'écrire

$$I = \prod_{j=1}^p I_j$$

alors même que  $p$  est le plus souvent l'une des inconnues du problème traité.

Bien que moins connues des psychométriciens, d'autres axiomatiques des espaces Euclidiens peuvent être utilisées, et en particulier, dans le cadre ordinal, de LEEUW et HEISER (1982) font référence aux travaux de PIERI (1908). Celui-ci montre que la relation d'équidistance  $\mathcal{J}(j; i, k)$  si et seulement si  $j$  est équidistant de  $i$  et de  $k$  (c'est à dire si  $j$  est le sommet d'un triangle isocèle de base  $ik$ ) est une primitive suffisante pour les géométries Euclidiennes. Il est à noter que BEALS et KRANTZ (1967) fondent leur construction sur cette relation, comme résultante de  $\mathcal{B}$  et  $\mathcal{H}$ . SCOTT (1956) montre que la relation ternaire  $\mathcal{Y}(i, j, k)$  si et seulement si  $i, j$  et  $k$  forment un triangle rectangle fournit une primitive concurrente. D'autre part BETH et TARSKI (1956) montrent que l'on peut symétriser la relation  $\mathcal{J}(j; i, k)$  en  $\mathcal{Z}(i, j, k)$  si et seulement si  $(\mathcal{J}(j; i, k) \text{ et } \mathcal{J}(i; j, k))$ , c'est à dire si et seulement si  $i, j$  et  $k$  forment un triangle équilatéral. Cependant, ces travaux, bien que complétés par SZMIELEW (1983), n'ont pas permis de proposer de nouvelles méthodes d'analyse d'un tableau de dissimilarité et supposent des propriétés de densité de l'espace qui ne peuvent être testées en pratique. On doit donc s'interroger sur le sens à donner aux inférences déduites d'une analyse de données de dissimilarité, puisque toutes les axiomatiques permettant de démontrer des propriétés d'existence d'une représentation congruente de  $(I, \delta)$  reposent sur des conditions non vérifiables à partir de données expérimentales.



## CHAPITRE III

### STRUCTURES AFFINES ET ANALYSE D'UNE DISSIMILARITE EUCLIDIENNE

Les théorèmes d'existence d'un plongement Euclidien d'un espace métrique  $(I, \delta)$  exposés plus haut montrent à l'évidence le rôle joué par les structures affines Euclidiennes dans ce type d'étude. Nous montrons ici comment sont prises en compte ces structures affines lors de la construction explicite d'une solution.

Nous explicitons dans un premier temps quels sont les choix nécessaires à la mise en œuvre d'une méthode constructive d'un plongement Euclidien de  $(I, \delta)$  dans un espace affine Euclidien. Nous exposons ensuite l'approche classique du problème en termes d'Analyse en Coordonnées Principales. On trouve les premiers exposés de cette méthode dans YOUNG et HOUSEHOLDER (1938) et TORGERSON (1951, 1958) qui reprend la formulation de Tucker, Green et Abelson. Cette solution est la plus souvent adoptée pour analyser une fonction de dissimilarité mesurée selon une échelle de rapport et est même qualifiée par beaucoup de Classical Scaling Method (cf. GOWER (1966, 1982a, 1982b), MARDIA (1980)). Nous explicitons enfin les liens existant entre les présentations française et anglo-américaine du problème.

#### III.1. STATISTIQUES AFFINES INVARIANTES ET PLONGEMENT EUCLIDIEN

Soit  $(I, \delta)$  un espace quasi-métrique où  $I$  est de cardinal fini  $n$ . Alors, si  $\delta$  est une dissimilarité Euclidienne, nous la notons  $\delta(Z)$ . Comme plus haut, nous interprétons  $Z$  comme la matrice d'une application linéaire

$$z : F^* \rightarrow E$$

C'est la structure affine Euclidienne de  $E$  qui fonde les théorèmes d'existence présentés dans le paragraphe §II.1. Considérons donc  $\{P_0, \{P_j ; j \in \{1, \dots, p\}\}\}$  un référentiel de  $E$  avec  $p = \dim(E)$ .

Considérons de plus un espace affine  $G$  de direction  $\vec{G}$ , repéré dans le référentiel  $\{N_0, \{N_k ; k \in \{1, \dots, q\}\}\}$ .  $N_k$  (resp.  $P_j$ ) est associé au vecteur de base  $\vec{g}_k$  de  $\vec{G}$  (resp.  $\vec{e}_j$  de  $\vec{E}$ ). Alors toute application affine

$$f : E \rightarrow G$$

a pour application linéaire associée  $\vec{f}$  de  $\mathcal{L}(\vec{E}, \vec{G})$ . De façon classique, il existe un élément unique  $\hat{f}$  de  $\mathcal{L}(\hat{E}, \hat{G})$  qui prolonge  $f$  à  $\hat{E}$ .  $\vec{f}$  est alors la restriction de  $\hat{f}$  à  $\vec{E}$ . On note  $M(f)$  et  $M(\vec{f})$  les matrices respectives de  $f$  et de  $\vec{f}$  dans les bases associées aux référentiels définis plus haut.

Supposons de plus que les origines de  $E$  et de  $G$  se correspondent par

$$f(N_0) = P_0 + \sum_{j=1}^p b_j \vec{e}_j$$

Alors tout point  $N$  de  $E$ , de vecteur de coordonnées  ${}^1(1, \underline{z})$  dans  $\hat{E}$  a pour image dans  $G$  le point  $P$  de vecteurs de coordonnées dans  $\hat{G}$

$$\begin{pmatrix} 1 \\ \underline{y} \end{pmatrix} = M(f) \begin{pmatrix} 1 \\ \underline{z} \end{pmatrix} = \begin{pmatrix} 1 & \underline{0} \\ \underline{b} & M(\vec{f}) \end{pmatrix} \begin{pmatrix} 1 \\ \underline{z} \end{pmatrix}$$

Notons donc  $A = M(\vec{f})$ , on vérifie alors l'identité

$$\underline{y} = \underline{b} + A \underline{z}$$

et donc  $f$  associe au plongement Euclidien  $Z$  de  $(I, \delta)$ , le plongement Euclidien  $Y$  de  $(I, \delta)$  dans  $G$  défini par

$$Y = \underline{b} \underline{1} + A Z$$

**Remarque :** Soient  $j$  et  $k$  deux éléments de  $[n]$ , alors

$$y_j - y_k = A (z_j - z_k)$$

Par suite les distances interpoints ne dépendent pas du choix de  $\underline{b}$ . Le choix d'une origine arbitraire fait partie des problèmes rencontrés lors de la mise en œuvre d'une méthode de Codage Multidimensionnel.

Si nous considérons de plus le cas  $E=G$ , les Analyses en Composantes Principales des triples  $(Y, M, N)$  et  $(Z, M, N)$  fournissent la même solution dans le seul cas où

i)  $\underline{b} = \underline{0}$

ii)  $A$  est une matrice orthogonale.

L'Analyse en Composantes Principales n'est pas invariante par transformations affines des données, décrites par un couple  $[A, \underline{b}]$  plus général que  $[R, \underline{0}]$  où  $R$  est une rotation ou une symétrie. Nous notons dans la suite  $L(p)$  le groupe des transformations affines  $[A, \underline{b}]$  régulières (de rang  $p$ ), c'est à dire telles que  $A$  soit inversible. L'étude de telles

transformations est en particulier menée par FRASER (1968, 1979) et OBENCHAIN (1971, 1973) en Statistique Multidimensionnelle. Ces auteurs posent la définition suivante :

**Définition 1:** Une statistique  $t$  calculée à partir du tableau de données  $Z$  est dite affine invariante si et seulement si

$$t(Z) = t(\underline{b}'\mathbf{1} + A Z)$$

pour tout  $[A, \underline{b}]$  dans  $L(p)$ .

Des exemples particuliers de telles statistiques sont la statistique  $T^2$  de Hotelling, la statistique du rapport de vraisemblance de Wilks ou le critère de trace de Pillai ou encore la fonction linéaire discriminante de Fisher.  $L(p)$  partitionne  $\mathcal{M}_{p,n}(\mathbb{R})$  en classes d'équivalence, ou orbites, définies par (cf. par exemple KSHISAGAR (1978), FRASER (1979))

$$(Y \sim Z) \Leftrightarrow (\exists [A, \underline{b}] \in L(p); Y = \underline{b}'\mathbf{1} + A Z)$$

**Définition 2:** La statistique  $t$  est dite  $L(p)$  maximale affine invariante (MAI) si et seulement si elle est affine invariante et prend des valeurs distinctes sur chaque orbite de  $\mathcal{M}_{p,n}(\mathbb{R})$ .

La notion essentielle en analyse des tableaux de dissimilarité est celle d'indice de position centrale compatible avec  $L(p)$ . En particulier, considérons le sous-groupe  $L_0(p)$  de  $L(p)$  formé par les transformations  $[A, \underline{Q}]$  où  $\underline{Q}$  est le vecteur nul de  $\hat{E}$ . Considérons deux éléments  $X_1$  et  $X_2$  de  $\mathcal{M}_{p,n}(\mathbb{R})$ . Le centrage du tableau  $X_i$  ( $i=1, 2$ ) par  $\underline{m}_i = m(X_i)$  est l'opération consistant à retrancher un indice de position centrale  $\underline{m}_i$  de  $E$  à chaque colonne de  $X_i$ . Posons alors

$$Y_i = X_i - \underline{b}_i'\mathbf{1}$$

Lorsque  $X_1$  et  $X_2$  sont deux éléments d'une même orbite de  $\mathcal{M}_{p,n}(\mathbb{R})$  induite par  $L(p)$ , alors

$$Y_2 = (\underline{b} + A\underline{m}_1 - \underline{m}_2)'\mathbf{1} + AY_1$$

$Y_1$  et  $Y_2$  appartiennent donc à la même orbite de  $L_0(p)$  si et seulement si la fonction  $m$  vérifie la relation

$$m(\underline{b}'\mathbf{1} + A X) = \underline{b} + A m(X) \quad \forall [A, \underline{b}] \in L(p) \text{ et } \forall X \in \mathcal{M}_{pn}(\mathbb{R}) \quad (9)$$

qui montre en particulier que  $\underline{b}$  ne contient aucune information invariante pour  $L(p)$ .

**Définition 3:** Une statistique  $m$  est un indice de position centrale compatible avec  $L(p)$  si et seulement si  $m$  vérifie la condition (9).

Cette propriété est partagée par la majorité des indices de position centrale utilisés en Statistique dans le cas univarié  $p=1$ . Supposons de plus que la fonction  $m$  est continue sur  $\mathcal{M}_{p,n}(\mathbb{R})$ , cela signifie que l'on a pour toute suite  $(X_n)$  convergente dans  $\mathcal{M}_{p,n}(\mathbb{R})$

$$\lim_{n \rightarrow \infty} m(X_n) = m(\lim_{n \rightarrow \infty} X_n)$$

OBENCHAIN (1971) montre alors le résultat suivant, où  $\mathbf{1}$  est le vecteur de  $\mathbb{R}^n$  de coordonnées égales à 1, qu'il convient de mettre en rapport avec la proposition 1 due à GOWER (1982), présentée au paragraphe §III.2 :

**Lemme 1:** Pour  $p > 1$ , les seuls indices de position centrale continus compatibles avec  $L(p)$  sont les moyennes pondérées de la forme

$$m(X) = X\gamma = \underline{m}$$

où  $\gamma \in \mathbb{R}^n$  et vérifie  $\gamma \mathbf{1} = 1$ . De plus si  $Y = X - \underline{m} \mathbf{1}$  alors  $r = \text{rang}(Y) \leq \min(p, n-1)$ .

Puisque  $m(X) = X\gamma$  on a

$$Y = X - X\gamma \mathbf{1} = X(I - \gamma \mathbf{1}) = X {}^t Q_\gamma$$

où la matrice  ${}^t Q_\gamma$  s'écrit en fonction de la matrice diagonale  $D_\gamma$  définie par  $D_\gamma \mathbf{1} = \gamma$

$${}^t Q_\gamma = I - \gamma \mathbf{1} = I - D_\gamma \mathbf{1} \mathbf{1}$$

Notons  $F = (\mathbb{R}^n, D_\gamma)$ . La forme quadratique  $d_\gamma$  de matrice  $D_\gamma$  dans la base  $\mathcal{F}$  définit une géométrie sur  $F$ , qui est Euclidienne si et seulement si  $\gamma_i \geq 0$  pour tout  $i$  dans  $[n]$ . Dans tous les cas  ${}^t Q_\gamma$  est  $D_\gamma$ -symétrique et idempotente. C'est la matrice du projecteur  $D_\gamma$ -orthogonal associé à la décomposition additive de  $F$  et de son dual

$$F^* = F_\gamma^* \oplus \Delta_1^* \quad \Leftrightarrow \quad F = F_\gamma \oplus \Delta_1$$

$Q_\gamma$  est le projecteur  $D_\gamma$ -orthogonal sur  $F_\gamma$ , le sous-espace vectoriel supplémentaire  $D_\gamma$ -orthogonal de la droite des constantes  $\Delta_1$  dans  $F$ . Tout choix d'une origine de  $E$  se traduit dans  $F$  et dans  $F^*$  par une décomposition orthogonale de l'espace.

**Remarque :** La relation entre le choix d'une origine dans  $E$  d'une part et la détermination d'une géométrie dans  $F$ , ne semble pas avoir été relevé dans la littérature, malgré son caractère canonique. Les travaux en langue anglaise traitent de façon exclusive le premier

problème. En particulier GOWER (1982 a, b) étudie différents choix possibles de  $\gamma$  et donc de  $\underline{m}$ . En France, la littérature est consacrée au plongement Euclidien du triple  $(I, \delta, P_I)$  où  $P_I$  est une pondération :

$$P_I : I \rightarrow \mathbb{R}$$

$$i \mapsto p_i$$

et l'on impose en général  $p_i \geq 0$  (et accessoirement la contrainte de normalisation  $\sum p_i = 1$ ). Le souci de définir une géométrie Euclidienne  $D_p$  sur  $F$  semble être la raison de l'absence d'exposés montrant les liens avec la démarche anglo-américaine. La relation entre les problèmes de centrage en Analyse en Composantes Principales et le choix de métriques  $N$  sur  $F$  est pourtant classiquement évoquée dans les enseignements d'Analyse des Données.

Remarquons dans  $E$  que la relation

$$Y = X - \underline{m}'\underline{1} \Leftrightarrow y_i = z_i - \underline{m}$$

traduit les relations traitées plus haut entre un espace affine  $E$  et sa direction  $\vec{E}$ . Considérons en effet la base  $(\vec{e}_j, j \in [p])$  et le référentiel  $(P_0, (P_j; j \in [p]))$  de  $E$  associé au choix d'une forme linéaire  $\varphi$  de matrice  $\underline{m} = Z\gamma$ . Alors l'application affine de  $E$  dans  $E$ , de prolongement  $\hat{i}$  dans  $\hat{E}$ , caractérisé par le vecteur  $\underline{m}$ , permet d'écrire

$$M(\hat{i}) = I \quad M(i) = \begin{pmatrix} 1 & '0 \\ -\underline{m} & I \end{pmatrix}$$

On obtient alors pour tout  $z \in \vec{E}$  la relation souhaitée

$$y_i = z_i - \underline{m} \quad \text{où} \quad \underline{m} = Z\gamma = Z D_\gamma \underline{1}$$

où le vecteur  $\underline{m}$  s'interprète comme l'image du vecteur  $\underline{1}$  de  $F$  par l'application  $Z D_\gamma$ . Lorsqu'on munit  $\vec{E}$  d'une structure Euclidienne induite par la forme quadratique  $q$  de matrice  $M = (m_{ij})$ , on associe classiquement à  $Z$  et à  $Y$  les matrices de connexion

$$W_Z = {}^t Z M Z$$

$$W_Y = {}^t Y M Y = Q_Y {}^t Z M Z Q_Y = Q_Y W_Z {}^t Q_Y$$

Soit donc, l'image de  $\underline{m}$  par  ${}^t Z M$  est le vecteur  $\underline{w}^Z$  de  $F$  défini par

$$\underline{w}^Z = {}^t Z M \underline{m} = W_Z \gamma = W_Z D_\gamma \underline{1}$$

$\underline{w}^Z$  a pour terme général la moyenne pondérée

$$w_i^Z = \sum_{j=1}^n \gamma_j w_{ij}^Z = w_i^Z$$

et par suite  $\underline{w}^Z$  s'interprète dans  $F$  comme le barycentre des vecteurs colonnes de  $W_Z$ . De plus

$${}^t Y W_Z Y = {}^t Y \underline{w}^Z = \sum_{i=1}^n \gamma_i w_i^Z = \sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j w_{ij}^Z = \underline{w}^Z$$

Posons donc

$$-g = \underline{w}^Z - \frac{1}{2} \underline{w}^Z \underline{1} = W_Z Y - \frac{1}{2} {}^t Y W_Z Y \underline{1}$$

on vérifie alors le résultat suivant :

**Lemme 2:** Soit  $\underline{m}$  un indice de position centrale compatible avec  $L(p)$ , défini par  $\underline{m} = ZY$  où  ${}^t Y \underline{1} = 1$  et  $g = W_Z Y + \frac{1}{2} {}^t Y W_Z Y \underline{1}$ . Alors, si  $Y = X - \underline{m} \underline{1}$ , on a

$$W_Y = Q_Y W_Z {}^t Q_Y = W_Z + (g \underline{1} + \underline{1} g) \quad (10)$$

**Preuve :**

$$\begin{aligned} W_Y &= Q_Y W_Z {}^t Q_Y = (I - \underline{1} \underline{1} D_Y) W_Z (I - D_Y \underline{1} \underline{1}) \\ &= W_Z - \underline{1} \underline{1} D_Y W_Z - W_Z D_Y \underline{1} \underline{1} + \underline{1} \underline{1} D_Y W_Z D_Y \underline{1} \underline{1} \end{aligned}$$

et  $g = -W_Z Y + \frac{1}{2} {}^t Y W_Z Y \underline{1}$  donne le résultat souhaité.  $\parallel$

Remarquons de plus que l'on a

$$\begin{aligned} {}^t Y W_Z Y &= {}^t \underline{m} M \underline{m} = \|\underline{m}\|_M^2 \\ W_Z Y &= {}^t Z M Z Y = {}^t Z M \underline{m} \end{aligned}$$

De plus, puisque  $W_Z$  appartient à  $\mathfrak{L}(F^*, F)$ , il est intéressant de noter que l'expression (10) correspond à la décomposition additive

$$F = F_Y \oplus \Delta_1 \Rightarrow F \otimes F = \Delta_1 \otimes \Delta_1 \oplus F_Y \otimes \Delta_1 \oplus \Delta_1 \otimes F_Y \oplus F_Y \otimes F_Y$$

de l'espace  $F \otimes F$  qui est isomorphe à  $\mathfrak{L}(F^*, F)$ . D'autre part, dans  $\hat{E}$ , on associe à  $q$  son prolongement unique a déterminé par  $M$  et  $\underline{m}$ .

**Lemme 3:** La forme 2 affine  $a$ , définie sur  $E$  par l'unique extension de  $q$  à  $\hat{E}$  associée à  $M$  et  $\underline{m}$ , a pour expression matricielle dans le référentiel  $\{P_0, \{P_j; j \in [p]\}\}$

$$\begin{pmatrix} \|\underline{m}\|_M^2 & -{}^t\underline{m}M \\ -M\underline{m} & M \end{pmatrix} = C_m$$

Elle ne dépend que du choix d'une métrique  $M$  sur  $\vec{E}$  et d'une origine fixée dans  $E$  par la donnée du vecteur  $\underline{m}$  de  $\vec{E}$ .

**Preuve :** Dans  $\vec{E}$ , l'image  $\begin{pmatrix} 1 \\ \underline{y} \end{pmatrix}$  de  $\begin{pmatrix} 1 \\ \underline{z} \end{pmatrix}$  par  $\hat{i}$  a pour norme

$$n(\underline{y}) = (1, {}^t\underline{y}) \begin{pmatrix} 0 & {}^t\underline{Q} \\ \underline{Q} & M \end{pmatrix} \begin{pmatrix} 1 \\ \underline{y} \end{pmatrix}$$

Utilisons l'expression matricielle  $M(i)$  donnée plus haut. Il vient alors

$$n(\underline{y}) = (1, {}^t\underline{z}) C_m \begin{pmatrix} 1 \\ \underline{z} \end{pmatrix}$$

où

$$C_m = {}^tM(i) \begin{pmatrix} 0 & {}^t\underline{Q} \\ \underline{Q} & M \end{pmatrix} M(i) \quad \parallel$$

Remarquons de plus que le développement de l'expression

$$(\underline{1}_n, {}^t\underline{Y}) \begin{pmatrix} 0 & {}^t\underline{Q} \\ \underline{Q} & M \end{pmatrix} \begin{pmatrix} {}^t\underline{1}_n \\ \underline{Y} \end{pmatrix} = (\underline{1}_n, {}^t\underline{z}) C_m \begin{pmatrix} {}^t\underline{1}_n \\ \underline{Z} \end{pmatrix}$$

permet de retrouver l'identité (10) donnée dans le lemme 2.

Le choix particulier de la métrique  $M = V_Y^-$  où  $V_Y = YD_Y{}^tY$  et  $A^-$  est une inverse généralisée quelconque de  $A$ , offre un intérêt spécifique. En effet dans ce cas, l'opérateur  $P$  défini sur  $F$  par

$$P = W_Y D_Y \text{ où } W_Y = {}^tY V_Y^- Y$$

est la matrice du projecteur  $D_Y$ -orthogonal sur le sous-espace  $F_Y$  de  $F$  engendré par les vecteurs lignes de  $Y$ . Ce résultat demeure vrai pour tout autre choix de la matrice  $D_P$  prise en lieu et place de  $D_Y$  dans l'expression de  $V_Y$  et  $W_Y$ . En particulier, OBENCHAIN (1971) et RAO (1982) considèrent le cas  $D_P = \text{Identité}$ . Dans ce cas, la distance de Mahalanobis entre  $P_i$  et  $P_j$  est (cf. aussi MEULMAN (1986))

$$d_{ij}^2(Y) = {}^t(\underline{y}_i - \underline{y}_j) V_Y^- (\underline{y}_i - \underline{y}_j) = {}^t(\underline{z}_i - \underline{z}_j) V_Y^- (\underline{z}_i - \underline{z}_j) = d_{ij}^2(V_Y^{-1/2} Y)$$

et donc si  $U$  est la matrice obtenue par factorisation de  $P = W_Y$  en  $P = {}^t U U$  avec  $U {}^t U = I$  alors  $\gamma$  est vecteur propre de  $P$  associé à la valeur propre  $\lambda=0$  et donc  $m(U) = U\gamma = 0$ . De plus

$$\begin{aligned} d_{ij}^2(Y) &= {}^t(\underline{u}_i - \underline{u}_j)(\underline{u}_i - \underline{u}_j) \\ &= p_{ii} + p_{jj} - 2p_{ij} \end{aligned}$$

Notons  $H(Y)$  la matrice de terme général  $h_{ij} = d_{ij}^2(Y)$ . On vérifie alors le théorème suivant (cf. OBENCHAIN (1971))

**Théorème 1:** Si  $m(Z)$  est compatible avec  $L(p)$ , alors  $P (=W_Y)$  et  $H(Y)$  sont deux statistiques maximales invariantes pour  $L(p)$ .

OBENCHAIN montre de plus qu'alors

$$0 \leq h_{ij} \leq 2 - \frac{(\gamma_i - \gamma_j)^2}{{}^t \gamma \gamma} \quad \forall i, j \in [n]$$

L'Analyse en Composantes Principales fournit des résultats qui ne sont pas invariants par transformation affine des données. Par exemple, l'analyse du tableau  $W_Y$  résultant de la transformation  $[T, \gamma]$  où  $T$  factorise  $V_Y$  en  ${}^t T T$  fournit des solutions distinctes de l'analyse appliquée au tableau  $Z$  dans la métrique  $M = \text{Identité}$ . Par suite, la position des points et l'évaluation des distances interpoints lues sur les résultats dépendent des choix faits lors de la mise en œuvre de la méthode. Nous présentons plus bas un résultat qui montre la nature de cette dépendance.

Notons d'autre part qu'OBENCHAIN(1973) propose de construire une méthode affine invariante de Codage Multidimensionnel, en minimisant la fonction critère

$$\psi(\alpha, \beta, q_{I \times I}) = \sum_{i < j} q_{ij} |\delta_{ij}^\alpha(Z) - \delta_{ij}^\alpha(X)|^\beta \quad \alpha, \beta \in \mathbb{R}, \quad \alpha, \beta > 0$$

où  $q_{I \times I}$  est une pondération positive définie sur  $I \times I$ . L'optimisation de  $\psi(\alpha, \beta, q_{I \times I})$  sous la contrainte  $X = AZ$ , où la matrice  $A$  de  $\mathcal{M}_{q,p}(\mathbb{R})$  est supposée de rang  $q$ , fournit une solution affine invariante si  $\delta_{ij}(Z)$  est la distance de Mahalanobis de  $P_i$  à  $P_j$  calculée à partir de  $Z$ .

**Remarque :** Lorsque  $\alpha = \beta = 2$  et  $q_{I \times I}$  est la pondération triviale égale à 1 pour tout couple  $(i, j)$  d'éléments de  $I$ , OBENCHAIN (1973) observe que si l'on note

$$\|A\|^2 = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2$$

$$H(N) = (n_{ij}^2) \quad \text{si} \quad N = (n_{ij})$$

$$W = Q_1 \left(-\frac{1}{2} H(\Delta)\right) {}^t Q_1 \quad Q_1 = I - \frac{1}{n} \mathbf{1} \mathbf{1}^t$$

$$W_Z = {}^t Z M Z$$

$$r = (w_{ii} - (W_X)_{ii})_{i=1\dots n}$$

alors on obtient par un calcul simple

$$\|H(\Delta) - H(D(X))\|^2 = 4 \|R\|^2 + 2 \mathbf{1}^t r + 2n \mathbf{1}^t r$$

où  $R$  est la matrice des résidus  $R = (W_Z - {}^t X X)$  associée à  $Z$  et donc  $\|R\|^2$  est l'une des fonctions critères minimisées par l'Analyse en Composantes Principales (cf. CAILLEZ et PAGES (1976)). Il importe pour la suite de remarquer que les deux méthodes diffèrent par le rôle joué par les perturbations  $r_{ii} = (w_{ii} - (W_X)_{ii})$  de la diagonale de  $W_Z$ . Nous discutons plus en détail ce problème au chapitre V.

Le lemme 2 montre que le choix d'une origine dans  $E$ , ou de façon équivalente, le choix d'une pondération  $\gamma$  sur  $I$ , à valeurs dans  $\mathbb{R}$ , détermine une matrice  $W_Z$  particulière associée à la dissimilarité Euclidienne  $\delta(Z)$ . GOWER (1982 a, b, 1983) étudie les liens entre la matrice  $W_Z$  et la matrice  $H$  de terme général  $h_{ij}$

$$h_{ij} = \delta_{ij}^2(Z) \quad \Leftrightarrow \quad H = \Delta(Z) * \Delta(Z) \quad \Delta(Z) = (\delta_{ij}(Z))_{ij}$$

$H$  est le produit de Hadamard de la matrice  $\Delta(Z)$  par elle-même. Par suite, lorsque  $\delta(Z)$  est une dissimilarité propre, on a pour tout couple  $(i,j)$  d'éléments de  $I$

$$h_{ij} = h_{ji}$$

$$h_{ij} \geq 0$$

$$h_{ii} = 0$$

Nous présentons dans le paragraphe suivant, certains résultats de GOWER (1982 a, b, 1983, 1985) utiles pour la suite de notre exposé. Notre présentation de ces résultats explicite de plus les relations entre le choix d'une origine dans  $E$  et le choix d'un vecteur de  $F^*$ . D'autre part, l'expression de la dualité entre  $F$  et  $F^*$  montre alors de façon simple l'équivalence entre le choix d'une origine de  $E$  et le choix d'une pondération  $P_I$  de  $I$ . Nous étudions alors en détail l'expression particulière des équations mises en place lorsqu'on fait le choix particulier du centre de l'hypersphère circonscrite au (resp. inscrite dans) le simplexe  $\{P_i, i \in [n]\}$  pour origine de  $E$ . Le centre de l'hypersphère inscrite est utilisée

dans la suite, lors de l'étude d'une nouvelle méthode de plongement Euclidien d'un espace semi-métrique, que nous appelons l'Analyse en Coordonnées d'Images.

### III.2. CHOIX D'UNE ORIGINE DANS E ET D'UNE PONDÉRATION DE I

Notons  $\mathcal{H}$ , le sous-espace vectoriel de  $\mathcal{M}_{n,n}(\mathbb{R})$  formé par les matrices symétriques à diagonale nulle.  $\mathcal{H}$  constitue un exemple de structure linéaire étudiée par MAGNUS (1986).  $\mathcal{H}$  est isomorphe à un sous-espace vectoriel de  $F \otimes F$ , puisque  $\mathcal{M}_{n,n}(\mathbb{R})$  est isomorphe à  $F \otimes F$ . Notons  $\text{Vecs}$  cet isomorphisme. Lorsque  $F \otimes F$  est muni de la base  $\{f_i \otimes f_j ; (i,j) \in [n]^2\}$  associée à la base  $\{f_i ; i \in [n]\}$  de  $F$ , on associe alors à tout élément  $A$  de  $\mathcal{M}_{n,n}(\mathbb{R})$  le tenseur de vecteur représentatif

$$\underline{a} = \text{Vec} (A) = {}^t(\underline{a}_1, \dots, \underline{a}_n)$$

où  $\underline{a}_i$  est la  $i$ -ème ligne de  $A$ . De même, on vérifie que tout élément  $H$  de  $\mathcal{H}$  est représenté par le vecteur

$$\underline{h} = \text{Vecs} (H) = (h_{12}, \dots, h_{1n}, h_{23}, \dots, h_{2n}, \dots, h_{n-1 n})$$

Dans la suite, sauf risque d'erreur, nous identifions  $\mathcal{H}$  et  $\text{Vecs}(\mathcal{H})$ .

Notons  $A * B$ , le produit de Hadamard des matrices  $A$  et  $B$ . On obtient, par application de la loi du cosinus

$$\begin{aligned} h_{ij} &= \delta_{ij}^2(Z) = w_{ii}^Z + w_{jj}^Z - 2 w_{ij}^Z \\ \Leftrightarrow H &= \underline{w}^Z {}^t \underline{1} + \underline{1} {}^t \underline{w}^Z - 2 W_Z \\ &= W_Z * I {}^t \underline{1} + \underline{1} W_Z * I - 2 W_Z \end{aligned}$$

GOWER (1982a) montre que toute autre représentation  $Z$  de  $I$ , de matrice de connexion associée  $W_Z = {}^t Z M Z$  vérifie une équation spécifique analogue à l'équation (10).

**Lemme 4 :** *Toute représentation  $Z$  de  $I$ , réalisant un plongement Euclidien de l'espace quasi-métrique  $(I, \delta)$  est associée à un vecteur  $\underline{g}$  de  $F$  tel que*

$$W_Z = -\frac{1}{2} [H - (\underline{g} {}^t \underline{1} + \underline{1} {}^t \underline{g})] \quad (11)$$

**Preuve :** Si  $Y$  est une représentation de  $I$ , on a par la loi du cosinus

$$h_{ij} = w_{ii}^Z + w_{jj}^Z - 2 w_{ij}^Z$$

et puisqu'on peut toujours écrire  $W_Z$  sous la forme

$$W_Z = -\frac{1}{2}H + G \quad \text{où } H \in \mathcal{H}$$

$$h_{ij} = g_{ii} + g_{jj} + h_{ij} - 2g_{ij}$$

$G$  doit donc vérifier la relation

$$g_{ij} = \frac{1}{2}[g_{ii} + g_{jj}] \Leftrightarrow G = \frac{1}{2}[g \mathbf{1} + \mathbf{1} g] \quad \text{où } g = (g_{ii})_{i=1, \dots, n}$$

Par suite  $W_Z$  diffère de  $\frac{1}{2}H$  par une matrice de rang 2.  $Z$  est une représentation de  $I$  et donc  $W_Z$  doit être positive semi-définie. GOWER (1982a) pose le problème inverse. Etant donnée une dissimilarité Euclidienne  $\delta(Z)$ , on doit trouver  $g$  tel que  $W_Z$  soit positive semi-définie. Bien sûr, le choix n'est pas unique et Gower montre même que l'on peut ainsi générer des représentations de rang distinct (différant au plus de 1) associées à une même matrice. Il donne l'exemple

$$Z_1 = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \quad Z_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

où l'on observe que l'absence d'unicité traduit le caractère affine du problème. Comparons les équations (10) et (11). On observe alors que  $g$  est lié à un choix d'origine dans  $E$ . Nous avons associé plus haut à toute translation de l'origine de  $E$ , la relation

$$W_Y = Q_Y W_Z {}^t Q_Y$$

Une telle relation est en particulier satisfaite par le couple  $(W_Z, -\frac{1}{2}H)$ . Nous présentons ici un résultat dû à GOWER (1982a) et fondé sur l'étude de la matrice

$$W = -\frac{1}{2} Q_Y H {}^t Q_Y \tag{12}$$

$$W = -\frac{1}{2} [H - \mathbf{1} {}^t H - H \mathbf{1} + {}^t H \mathbf{1} \mathbf{1}] \tag{12bis}$$

Sous cette forme, on reconnaît la matrice  $W$  traitée par Tucker, Green et Abelson (cf. TORGERSON (1958)).

**Proposition 1:** *La matrice  $W$  a pour terme général*

$$w_{ij} = -\frac{1}{2} [\delta_{ij}^2 - \delta_{i.}^2 - \delta_{.j}^2 + \delta_{..}^2] \tag{13}$$

où le terme général de  $H \mathbf{1}$  est

$$\delta_{i.}^2 = \sum_j \gamma_j \delta_{ij}^2$$

et

$$\delta_{..}^2 = \sum_i \gamma_i \delta_{i.}^2 = \sum_i \sum_j \gamma_i \gamma_j \delta_{ij}^2 = {}^t \gamma H \gamma$$

**Preuve :** CAILLEZ et PAGES (1976) notent que si  $\delta$  est une dissimilarité Euclidienne

$$\delta_{ij}^2 = w(f_i^* - f_j^*) = w(f_i^*) + w(f_j^*) - 2 w(f_i^*, f_j^*) \quad (14)$$

et d'après le théorème de Huygens, si  $P_0$  est le barycentre des points  $P_i$  dans  $E$ , il lui correspond  $\underline{m}$  dans  $\vec{E}$  qui vérifie pour tout  $z_i$  représentant  $P_i$  dans la base de  $\vec{E}$

$$I_{z_i} = I_{\underline{m}} + \|z_i - \underline{m}\|_M^2 = I_{\underline{m}} + w(f_i^*)$$

où  $I_z$  est l'inertie du nuage de points  $\{(P_i, \gamma_i) ; i \in [n]\}$  autour du point  $P_i$ . On en déduit

$$w(f_i^*) = \|z_i - \underline{m}\|_M^2 = I_{z_i} - I_{\underline{m}} = \delta_{i.}^2 - \frac{1}{2} \delta_{..}^2$$

puisque

$$I_{z_i} = \sum_j \gamma_j \|z_i - z_j\|_M^2 = \sum_j \gamma_j \delta_{ij}^2$$

$$I_{\underline{m}} = \sum_j \gamma_j \|z_j - \underline{m}\|_M^2 = \sum_j \gamma_j I_{z_j} - I_{\underline{g}}$$

(14) fournit donc la forme fondamentale dont est issue (13).  $\quad \parallel$

On pose dans la suite

$$w_i = w(f_i^*) = \delta_{i.}^2 - \frac{1}{2} \delta_{..}^2 \quad (15)$$

soit sous forme matricielle

$$\underline{w} = H \gamma - \frac{1}{2} {}^t \gamma H \gamma \underline{1} \quad (15bis)$$

**Proposition 2:** Si  $H$  est inversible, alors .

$$\gamma = H^{-1} \underline{w} - \frac{1}{k} H^{-1} \underline{1} ({}^t \underline{1} H^{-1} \underline{w} - 1)$$

**Preuve :** Interprétons  $F_\gamma$  comme un sous-espace affine de direction  $\vec{F}_\gamma$  tel que  $F = \vec{F}_\gamma \oplus \Delta_1$ . Alors  $F = \hat{F}_\gamma$  et l'équation (15) s'écrit de façon équivalente, en termes de forme 2-affine

$$C \begin{pmatrix} -\frac{1}{2} {}^t \gamma H \gamma \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 & {}^t \mathbf{1} \\ \mathbf{1} & H \end{pmatrix} \begin{pmatrix} -\frac{1}{2} {}^t \gamma H \gamma \\ \gamma \end{pmatrix} = \begin{pmatrix} 1 \\ \underline{w} \end{pmatrix}$$

Par suite lorsque  $H$  est inversible, on obtient par application de l'équation (5), paragraphe §I.1.4

$$\begin{pmatrix} -\frac{1}{2} {}^t \gamma H \gamma \\ \gamma \end{pmatrix} = \begin{pmatrix} \frac{1}{k^2} [{}^t \mathbf{1} H^{-1} \underline{w} - 1] \\ H^{-1} \underline{w} - \frac{1}{k^2} H^{-1} \mathbf{1} [{}^t \mathbf{1} H^{-1} \underline{w} - 1] \end{pmatrix}$$

où  $k$  est défini au paragraphe §II.4 par

$$k^2 = [{}^t \mathbf{1} H^{-1} \mathbf{1}]$$

On obtient ainsi l'expression de  $\gamma$  en fonction de  $\underline{w}$ . Lorsque  $H$  est singulière, l'expression de  $\gamma$  en fonction de  $\underline{w}$  n'est pas simple. D'autre part on vérifie par substitution de  $\underline{w}$  par son expression (15) que la première coordonnée est bien  $-\frac{1}{2} {}^t \gamma H \gamma$ . ¶

Lorsque  $H$  est inversible, l'expression (15) s'écrit en fonction de  $\alpha = -\frac{1}{2} {}^t \gamma H \gamma$  sous la forme

$$\underline{w} = H \gamma + \alpha \mathbf{1} \quad \Leftrightarrow \quad \gamma = H^{-1} \underline{w} - \alpha H^{-1} \mathbf{1}$$

et donc

$$\begin{aligned} -2\alpha &= {}^t \gamma H \gamma = \alpha^2 {}^t \mathbf{1} H^{-1} \mathbf{1} + {}^t \underline{w} H^{-1} \underline{w} - 2\alpha {}^t \underline{w} H^{-1} \mathbf{1} \\ \Leftrightarrow 0 &= \alpha^2 {}^t \mathbf{1} H^{-1} \mathbf{1} + {}^t \underline{w} H^{-1} \underline{w} - 2\alpha [{}^t \underline{w} H^{-1} \mathbf{1} - 1] \end{aligned} \quad (16)$$

$\alpha$  est donc solution d'une équation du second degré de discriminant

$$\Delta(\underline{w}) = [{}^t \mathbf{1} H^{-1} \underline{w} - 1]^2 - ({}^t \underline{w} H^{-1} \underline{w}) ({}^t \mathbf{1} H^{-1} \mathbf{1})$$

qui est nul si l'équation (16) admet une racine double. GOWER (1982a) montre que  $W_Z$  est positive et semi-définie seulement quand  ${}^t \mathbf{1} \gamma = 1$ , c'est à dire quand

$${}^t \underline{w} H^{-1} \mathbf{1} - 1 = -\frac{1}{2} ({}^t \gamma H^{-1} \gamma) ({}^t \mathbf{1} H^{-1} \mathbf{1}).$$

De plus  $\delta$  est Euclidienne et tout  $\underline{w}$  tel que  $W_Z$  est positive semi-définie, vérifie  $\Delta(\underline{w}) \leq 0$ .  
On vérifie alors la proposition suivante démontrée par GOWER (1982 a, b)

**Proposition 3:** Pour tout élément  $\underline{g}$  de  $F$ , il existe un vecteur  $\underline{\gamma}$  de  $F^*$  tel que la relation (11) soit équivalente à

$$W_Z = -\frac{1}{2} Q_\gamma H {}^t Q_\gamma$$

**Preuve (GOWER) :** Associons à  $Z$  la représentation  $Y$  obtenue par translation de l'origine de  $P_0$  à  $P_0'$  où  $P_0'$  a pour vecteur représentatif  $\underline{m}$  dans  $\vec{E}$ .

$$Y = Z + \underline{m} {}^t \underline{1}$$

Choisissons  $\underline{m} = Z\underline{\gamma}$  où  ${}^t \underline{\gamma} \underline{g} = 0$  et bien sur  ${}^t \underline{1} \underline{\gamma} = 1$ . On a par application du lemme 4

$$Y = Z {}^t Q_\gamma$$

$$W_Y = Q_\gamma W_Z {}^t Q_\gamma = -\frac{1}{2} Q_\gamma [H - (\underline{g} {}^t \underline{1} + \underline{1} {}^t \underline{g})] {}^t Q_\gamma$$

$$W_Y = -\frac{1}{2} Q_\gamma H {}^t Q_\gamma$$

du fait des contraintes portant sur  $\underline{\gamma}$ . Puisque  $H$  est de trace nulle,  $W_Y$  n'est semi-définie que si  ${}^t \underline{\gamma} \underline{1} = 1$ , et donc  $\det(I - {}^t \underline{\gamma} \underline{1}) = 0$ . Sinon il existe un élément  $\underline{u}$  de  $F^*$  tel que

$${}^t \underline{u} W_Y \underline{u} = ({}^t Q_\gamma \underline{u}) - \frac{1}{2} H {}^t Q_\gamma \underline{u} < 0$$

Si  $H$  est régulière, prendre  $\underline{\gamma}$  dans le noyau de  $H$  implique  $H\underline{\gamma} = 0$  et donc  $W_Y = \frac{1}{2} H$ .  $W_Y$  n'est alors pas définie positive puisque  $\text{Trace}(H) = 0$ . Si  $\underline{\gamma}$  n'appartient pas au noyau de  $H$ ,  $W_Y$  est positive semi-définie et donc  $\delta$  est Euclidienne.  $\square$

Le lemme suivant exprime un résultat intuitif d'un grand intérêt pratique : si  $(I, \delta)$  admet un plongement Euclidien dans  $E$ , rapporté à l'origine  $P_0$ , il admet un plongement Euclidien dans  $E$ , rapporté à toute autre origine.

**Lemme 5 :** Soit  $\underline{p}$  un vecteur de  $F^*$  tel que  ${}^t \underline{p} \underline{1} = 1$ , alors

$$i) {}^t Q_p {}^t Q_\gamma = {}^t Q_p$$

$$ii) Q_p W_Y {}^t Q_p = -\frac{1}{2} Q_p H {}^t Q_p$$

**Preuve :**

$${}^t Q_p = (I - \underline{1} {}^t \underline{p})$$

$${}^tQ_p {}^tQ_\gamma = (I - {}^t1_p) (I - {}^t1_\gamma) = I - {}^t1_p - {}^t1_\gamma + {}^t1_p {}^t1_\gamma = {}^tQ_p$$

puisque  ${}^t1_p = 1$ .

La relation ii) est une application directe de ce résultat puisque  $W_\gamma = -\frac{1}{2} Q_\gamma H {}^tQ_\gamma$  ¶

Résumons ces résultats dans le théorème suivant

**Théorème 2:** *L'espace quasi-métrique  $(I, \delta)$  admet un plongement isométrique dans un espace affine Euclidien si et seulement si la forme quadratique de matrice*

$$W = -\frac{1}{2} Q_\gamma H {}^tQ_\gamma$$

*est positive, semi-définie pour tout choix de  $\gamma$  de  $F^*$  tel que  ${}^t1_\gamma = 1$  et  $\gamma$  n'appartient pas au noyau de  $H$ .*

**Corollaire :** *Pour tout choix d'un vecteur  $\gamma$  de  $F^*$  satisfaisant aux conditions du théorème 2, on a*

$$i) \quad W\gamma = 0 \quad (\Leftrightarrow W D_\gamma 1 = 0)$$

*c'est dire que les lignes et les colonnes de la matrice  $W$  ont pour somme pondérée par  $\gamma$  la valeur 0.*

$$ii) \text{ si de plus } W = {}^tZ M Z, \text{ alors } Z \text{ est un tableau centré, au sens où } Z\gamma = 0$$

**Preuve :**

$$W = -\frac{1}{2} Q_\gamma H {}^tQ_\gamma$$

et puisque  ${}^t1_\gamma = 1$

$${}^tQ_\gamma \gamma = (I - \gamma {}^t1) \gamma = \gamma - \gamma ({}^t1_\gamma) = 0$$

Par suite  $W\gamma = 0$ . D'autre part

$${}^t\gamma W \gamma = {}^t(Z\gamma) M (Z\gamma)$$

et donc il résulte de i) que  $Z\gamma = 0$  ¶

### III.3. CHOIX PARTICULIERS D'UNE ORIGINE DE E

De nombreuses méthodes sont associées à un choix particulier de  $\gamma$ . SCHOENBERG (1935) et YOUNG et HOUSEHOLDER considèrent le cas  $\gamma = f_n^*$  pour une indication donnée de  $I$  dans  $[n]$ . Dans ce cas,  $\underline{w}$  a pour  $i$ -ème coordonnée

$$w_i = \delta_{in}^2 - \frac{1}{2} \delta_{nn}^2 = \delta_{in}^2$$

TORGERSON (1958) adopte la proposition de Tucker, Green et Abelson et choisit

$$\gamma = \frac{1}{n} \mathbf{1} = \frac{1}{n} \sum_i f_i^*$$

On vérifie alors que  $\underline{w}$  a pour  $i$ -ème coordonnée

$$w_i = \delta_{i.}^2 - \frac{1}{2} \delta_{..}^2 = I_{zi} - I_m$$

En France (cf. CAILLEZ et PAGES (1976), MAILLES (1981)), la présentation classique préfère exprimer les données du problème sous forme d'un triple  $(I, \delta, P_I)$  où  $(I, \delta)$  est un espace quasi-métrique et  $P_I$  est une pondération définie sur  $I$ . Alors, la tradition probabiliste suppose en général de plus que

$$p_i \geq 0 \quad \text{et} \quad \sum_i p_i = 1$$

On note qu'alors, la métrique des poids  $D_P$  permet d'obtenir

$$\begin{aligned} \gamma &= D_P \mathbf{1} \\ \underline{m} &= Z D_P \mathbf{1} \end{aligned}$$

et  $\underline{m}$  s'interprète comme un barycentre du nuage de points massiques  $\{(P_i, p_i) ; i \in [n]\}$  dans  $E$ . Puisque les  $p_i$  sont positifs et somment à 1, cela signifie que l'on choisit pour origine des axes un point appartenant au polyèdre convexe de  $E$  de sommets  $P_i, i \in [n]$ .  $I_{zi}$  et  $I_m$  s'interprètent alors en termes d'inertie.

Les résultats précédents montrent cependant que la restriction au cas d'une mesure  $P_I$  positive est inutile. Mais si  $P_I$  est une mesure de signe quelconque,  $D_P$  ne définit plus une géométrie euclidienne sur  $F$ . Le choix d'une géométrie euclidienne définie par une matrice diagonale de trace égale à 1 est donc équivalent au choix d'une origine de  $E$  appartenant au polyèdre convexe de  $E$  de sommets  $P_i, i \in [n]$ .

Bien entendu, le choix d'une origine  $P_0$  est illimité. Cependant GOWER (1962a, b, 1985a) étudie certains cas particuliers. Nous présentons ici deux choix possibles, étudiés en détail par cet auteur et auxquels il sera fait référence plus bas.

### III.3.1. Le Centre de l'Hypersphère Circonscrite au Simplexe $\{P_i ; i \in [n]\}$

### ou Circumcentre

On se propose alors de déterminer le point  $P_0$  de  $E$  tel que  $\|\overrightarrow{P_0 - P_i}\| = r$  pour tout  $i \in I$ .

Les points  $P_i$  décrivent dans ce cas, une hypersphère de  $E$ . D'autre part puisque

$$\|\overrightarrow{P_0 - P_i}\|^2 = w_{ii}^2$$

On cherche en fait le vecteur tel que

$$-\frac{1}{2} Q_\gamma H {}^t Q_\gamma = r^2 I$$

$$\Leftrightarrow Q_\gamma H {}^t Q_\gamma = -2 r^2 I$$

**Proposition 4:** Le choix d'une origine  $P_0$  telle que  $\|\overrightarrow{P_0 - P_i}\| = r$  conduit à représenter  $I$  sur une hypersphère de  $E$  de rayon

$$r = \frac{1}{\sqrt{2k}} \quad ; \quad k^2 = {}^t \mathbf{1} H^{-1} \mathbf{1}$$

l'origine  $P_0$  est alors associée au vecteur

$$\gamma = \frac{1}{k^2} H^{-1} \mathbf{1}$$

**Preuve :** L'équation (15bis) permet d'écrire plus simplement l'équation de l'hypersphère, sous la forme

$$\underline{w} = H\gamma - \frac{1}{2} {}^t \gamma H \gamma \mathbf{1} = r^2 \mathbf{1}$$

La contrainte  ${}^t \mathbf{1} \gamma = 1$  implique

$${}^t \gamma H \gamma - \frac{1}{2} {}^t \gamma H \gamma {}^t \mathbf{1} = r^2 {}^t \mathbf{1}$$

$$\Leftrightarrow \frac{1}{2} {}^t \gamma H \gamma = r^2$$

c'est dire que le rayon  $r$  vérifie

$$H\gamma = {}^t \gamma H \gamma \mathbf{1} = 2 r^2 \mathbf{1}$$

$$\Leftrightarrow \gamma = 2 r^2 H^{-1} \mathbf{1}$$

On obtient donc par substitution

$$r^2 = 2 r^4 {}^t \mathbf{1} H^{-1} \mathbf{1}$$

c'est à dire que

$$r^2 = \frac{1}{2} [{}^t \mathbf{1} H^{-1} \mathbf{1}]^{-1} = \frac{1}{2k^2} \quad \square$$

Les résultats précédents reposent sur la régularité de  $H$ . Lorsque  $H$  n'est pas inversible, on considère une inverse généralisée  $H^-$  quelconque de  $H$ . Alors, d'après l'équation (15bis), les distances des points  $P_i$  à l'origine  $P_0$  ont pour écart à leur moyenne empirique les coordonnées du vecteur

$$Q_0 \underline{w} = Q_0 H \underline{\gamma} - \frac{1}{2} \underline{\gamma}' H \underline{\gamma} Q_0 \underline{1} = Q_0 H \underline{\gamma}$$

Par suite la variance empirique des écarts à l'origine est  $\underline{w}' Q_0 \underline{w}$ . On définit donc le circumcentre comme l'origine  $P_0$  qui réalise le minimum de  $\underline{w}' Q_0 \underline{w}$  sous la contrainte  $\underline{\gamma}' \underline{1} = 1$ . Le Lagrangien s'écrit

$$L(\underline{\gamma}, \lambda) = \underline{w}' Q_0 \underline{w} + \lambda(1 - \underline{\gamma}' \underline{1}) = \underline{\gamma}' H Q_0 H \underline{\gamma} + \lambda(1 - \underline{\gamma}' \underline{1})$$

et donc la variance est minimum lorsque

$$H Q_0 H \underline{\gamma} = \lambda \underline{1} \quad (17)$$

Si  $\lambda = 0$ , un circumcentre existe. Alors l'équation (17) implique

$$Q_0 H \underline{\gamma} = 0 \quad \Leftrightarrow \quad H \underline{\gamma} = \frac{1}{n} (\underline{1}' H \underline{\gamma}) \underline{1}$$

Si  $\lambda \neq 0$ , on doit rechercher un circumcentre généralisé. Pour cela, prémultiplions (17) par  $H H^-$ . On obtient après simplification un résultat dont la contrepartie est élémentaire pour  $H$  inversible :

$$H H^- \underline{1} = \underline{1}$$

**Théorème 3 (GOWER (1985)) :** *Pour toute dissimilarité Euclidienne et pour toute inverse généralisée on a*

$$H H^- \underline{1} = \underline{1} \quad (18)$$

**Preuve :** La seule propriété des inverses généralisées utilisée ici est  $H H^- H = H$ . Par suite (18) ne dépend pas du choix d'une inverse généralisée. En effet soit  $H^+$  une autre inverse généralisée, alors (RAO et MITRA (1971)) il existe deux matrices  $P$  et  $Q$  telles que

$$H^+ = H^- + (I - H^- H)P + Q(I - H H^-) \quad (19)$$

Par suite

$$\begin{aligned} H H^+ &= H H^- + H Q (I - H H^-) \\ H H^+ \underline{1} &= H H^- \underline{1} + H Q (I - H H^-) \underline{1} = H H^- \underline{1} = \underline{1} \quad \square \end{aligned}$$

GOWER (1985) démontre alors le théorème d'existence suivant

**Théorème 4:** Si  $\delta$  est une dissimilarité Euclidienne définie sur  $I$ , on peut choisir comme origine de l'espace de représentation le circumcentre des points  $\{P_i; i \in [n]\}$  si et seulement si pour un choix d'inverse généralisée :

$$k^2 = \langle \mathbf{1}, H^{-1} \mathbf{1} \rangle \neq 0$$

et l'on vérifie alors le corollaire immédiat de ce théorème

**Corollaire :** Si  $\delta$  est une dissimilarité Euclidienne définie sur  $I$ , alors

$$k^2 = \langle \mathbf{1}, H^{-1} \mathbf{1} \rangle \geq 0$$

La nécessité de recourir à une inverse généralisée de  $H$  résulte du rôle essentiel joué par les problèmes de rang de  $H$  dans la construction de plongements Euclidiens de  $(I, \delta)$ . Les méthodes attachées à l'étude de

$$W = -\frac{1}{2} Q_\gamma H^{-1} Q_\gamma$$

sont de plus complexes à analyser du fait de l'absence d'identité entre les noyaux  $\text{Ker}(H)$  et  $\text{Ker}(W)$ .

Rappelons que la dimension de  $(I, \delta)$  est par définition la dimension de l'espace affine Euclidien irréductible dans lequel  $(I, \delta)$  admet un plongement isométrique, GOWER (1985) montre alors le résultat suivant

**Théorème 5 :**

a)  $\text{Ker}(H) \subset \text{Ker}(W)$

b)  $(\underline{u} \in \text{Ker}(H)) \Rightarrow \langle \mathbf{1}, \underline{u} \rangle = 0$

c)  $\underline{\gamma} \in \text{Ker}(W)$  mais  $\underline{\gamma} \notin \text{Ker}(H)$

d) Si  $\text{rang}(H) = r$ , alors  $(I, \delta)$  a pour dimension

- soit  $r-1$  si et seulement si  $\langle \mathbf{1}, H^{-1} \mathbf{1} \rangle \neq 0$ . Dans ce cas, les points représentatifs décrivent la surface d'une hypersphère.

- soit  $r-2$  si et seulement si  $\langle \mathbf{1}, H^{-1} \mathbf{1} \rangle = 0$ . Dans ce cas, les points représentatifs ne peuvent appartenir à la surface d'une hypersphère et  $H^{-1} \mathbf{1}$  est un élément de  $\text{Ker}(W)$

**Preuve (GOWER (1985)) :** Tout d'abord, si  $\text{rang}(H)=r$ ,  $\text{Ker}(H)$  est de dimension  $(n-r)$  et il existe donc  $(n-r)$  vecteurs  $\underline{u}_s$  formant un système générateur de  $\text{Ker}(H)$ . Par suite, notons

$$\text{Ker}(H) = \mathfrak{L} \text{in} \langle \underline{u}_s; s = 1, \dots, n-r \rangle.$$

Si  $\underline{u}_s$  est un élément de  $\text{Ker}(H)$ , alors  $H \underline{u}_s = 0$  et donc

$${}^1 \underline{1} u_s = {}^1 \underline{1} H^{-1} H u_s = 0$$

Par suite

$$W u_s = -\frac{1}{2} Q_\gamma [H u_s - H \gamma {}^1 \underline{1} u_s] = 0$$

et donc

$$\text{Ker}(H) \subset \text{Ker}(W).$$

a) et b) sont donc acquis. Pour montrer c) considérons l'implication

$$({}^1 \underline{1} = 1) \Rightarrow (\gamma \notin \mathfrak{L} \text{ in } \langle u_s; s = 1, \dots, n-r \rangle)$$

Par contre, si l'élément  $\underline{y}$  de  $\text{Ker}(W)$  est distinct de  $\gamma$ , on a en développant  $W$  en fonction de  $H$

$$\begin{aligned} W \underline{y} = 0 &\Leftrightarrow H \underline{y} = [{}^1 \underline{1} H \underline{y} - ({}^1 \underline{1} H \gamma) ({}^1 \underline{1} \underline{y})] \underline{1} + ({}^1 \underline{1} \underline{y}) H \gamma \\ &= \alpha \underline{1} + ({}^1 \underline{1} \underline{y}) H \gamma \end{aligned} \quad (20)$$

où

$$\alpha = {}^1 \underline{1} H \underline{y} - ({}^1 \underline{1} H \gamma) ({}^1 \underline{1} \underline{y})$$

et donc

$${}^1 \underline{1} \underline{y} = {}^1 \underline{1} H^{-1} H \underline{y} = k {}^1 \underline{1} H^{-1} \underline{1} + ({}^1 \underline{1} \underline{y}) {}^1 \underline{1} \gamma = \alpha {}^1 \underline{1} H^{-1} \underline{1} + {}^1 \underline{1} \underline{y}$$

D'autre part, on a par définition

$$H \underline{y} = 0 \Leftrightarrow \underline{y} \in \text{Ker}(H) \Leftrightarrow \underline{y} \in \mathfrak{L} \text{ in } \langle u_s; s = 1, \dots, n-r \rangle$$

Par suite, deux cas sont possibles :

i) Si  $\alpha = 0$ , l'équation (19) montre alors que

$$\underline{y} = \underline{y} - {}^1 \underline{1} \underline{y} \gamma$$

appartient à  $\text{Ker}(H)$  et donc  $\underline{y}$  appartient au sous-espace  $\mathfrak{L} \text{ in } \langle u_s; s = 1, \dots, n-r \rangle$  de  $F^*$ .

Le rang de  $W$  est donc  $(r-1)$ .

ii) Si  ${}^1 \underline{1} H^{-1} \underline{1} = 0$ , alors l'équation (20) s'écrit

$$H \underline{y} = 0$$

où

$$\underline{y} = \underline{y} - k H^{-1} \underline{1} + ({}^1 \underline{1} \underline{y}) \gamma$$

c'est à dire que  $\underline{y}$  appartient au sous-espace  $\mathcal{L}in\langle \underline{u}_s; s = 1, \dots, n-r \rangle, \underline{\gamma}, H^{-1}\underline{1} \rangle$  de  $F^*$ . Mais  $\underline{\gamma}$  est linéairement indépendant des  $\underline{u}_s$ . Montrons de plus que  $H^{-1}\underline{1}$  n'appartient pas au sous-espace  $\mathcal{L}in\langle \underline{u}_s; s = 1, \dots, n-r \rangle, \underline{\gamma} \rangle$ . Dans le cas contraire, on aurait

$$H^{-1}\underline{1} = \sum_{s=1}^{n-r} b_s \underline{u}_s + b\underline{\gamma}$$

et donc

$${}^t H^{-1}\underline{1} = \sum_{s=1}^{n-r} b_s {}^t \underline{1} \underline{u}_s + b {}^t \underline{1} \underline{\gamma} = b$$

$$1 = HH^{-1}\underline{1} = \sum_{s=1}^{n-r} b_s H \underline{u}_s + bH\underline{\gamma} \Rightarrow 1 = bH\underline{\gamma}$$

On arrive donc à une contradiction, si  ${}^t \underline{1} H^{-1}\underline{1} = 0$ .

Ce résultat est cependant insuffisant. Il convient en effet de s'interroger sur l'influence du choix d'une inverse généralisée particulière sur le résultat. Mais toute inverse généralisée  $H^{-}$  s'écrit en fonction de l'inverse généralisée de Moore Penrose notée  $H^+$  sous la forme (19). Par suite posons  $\underline{p} = P\underline{1}$ , alors il résulte de (19)

$$H^{-}\underline{1} = H^+\underline{1} = (I - H^+H)\underline{p}$$

Le choix de l'inverse généralisée de Moore Penrose permet de vérifier les identités (cf. RAO et MITRA (1971))

$$I - H^+H = \sum_{s=1}^r \underline{u}_s {}^t \underline{u}_s$$

$$H^{-}\underline{1} = H^+\underline{1} + \sum_{s=1}^r b_s \underline{u}_s \quad \text{où} \quad b_s = {}^t \underline{u}_s \underline{p}$$

et donc

$$\mathcal{L}in\langle \underline{u}_s; s = 1, \dots, n-r \rangle, \underline{\gamma}, H^{-}\underline{1} \rangle = \mathcal{L}in\langle \underline{u}_s; s = 1, \dots, n-r \rangle, \underline{\gamma}, H^+\underline{1} \rangle$$

par conséquent  $\text{Ker}(W)$  est de dimension  $r-2$ .  $\square$

### III.3.2. Le Centre de l'Hypersphère Inscrite dans le Simplexe $\{P_i; i \in [n]\}$ ou Centre Inscrit.

Le simplexe  $\{P_i; i \in [n]\}$  permet de construire les  $n$  hyperplans supports de chacune de ses faces  $\{P_i; i \in [n] \setminus \{j\}\}$ . On se propose de construire le point  $P_0$  de  $E$  équidistant de ces  $n$  hyperplans. Il est par définition, le centre de l'hypersphère inscrite dans le simplexe.

Soit  $Z$  une configuration associée à la dissimilarité Euclidienne  $\delta(Z)$ . Supposons la configuration centrée au barycentre des points  $P_i$ , associé au vecteur  $\gamma = \frac{1}{n} \mathbf{1}$ .

$$W_Z = {}^t Z M Z = -\frac{1}{2} Q_1 H {}^t Q_1$$

Notons alors

$$K = H^{-1} - \frac{H^{-1} \mathbf{1} {}^t \mathbf{1} H^{-1}}{{}^t \mathbf{1} H^{-1} \mathbf{1}} \quad (21)$$

la matrice qui intervient (cf. lemme 3) dans l'expression de l'inverse  $C^{-1}$  de la matrice

$$C = \begin{pmatrix} 0 & {}^t \mathbf{1} \\ \mathbf{1} & H \end{pmatrix}$$

On vérifie alors le résultat technique suivant :

**Lemme 6 :** Si  $H$  est une matrice régulière alors  $W_Z$  a pour inverse la matrice  $K$

**Preuve :** Puisque la matrice  $Q_1$  est définie par

$$Q_1 = I - \frac{1}{n} \mathbf{1} {}^t \mathbf{1}$$

on a

$$W_Z = -\frac{1}{2} \left[ H - \frac{1}{n} \mathbf{1} {}^t \mathbf{1} H - \frac{1}{n} H \mathbf{1} {}^t \mathbf{1} + \frac{1}{n^2} \mathbf{1} {}^t \mathbf{1} H \mathbf{1} {}^t \mathbf{1} \right]$$

Par suite, on vérifie par substitution que

$$W_Z K = K W_Z = I$$

c'est à dire que  $K$  est l'inverse de la matrice  $W_Z$ , c'est à dire

$$W_Z^{-1} = H^{-1} - \frac{H^{-1} \mathbf{1} {}^t \mathbf{1} H^{-1}}{{}^t \mathbf{1} H^{-1} \mathbf{1}} \quad \parallel$$

**Remarque 1:** Le terme général de la matrice  $W_Z^{-1}$  s'écrit en fonction des éléments  $h^{ij}$  de  $H^{-1}$ :

$$(W_Z^{-1})_{ij} = h^{ij} - \frac{h^{i\cdot} h^{j\cdot}}{h^{\cdot\cdot}}, \quad h^{i\cdot} = \sum_{j=1}^n h^{ij}, \quad h^{\cdot\cdot} = \sum_{i=1}^n h^{i\cdot}$$

Il traduit donc l'écart de la valeur de la case (i,j) au produit des marges du tableau H à l'aide d'une formule de calcul analogue à celle qui est utilisée en Analyse des Correspondances.  $h^{ij}$  s'interprète alors en termes de fréquence.

**Remarque 2 :** Le calcul de  $W_Z^{-1}$  peut être fait de façon directe en observant que si la décomposition spectrale de  $W_Z$  s'écrit  $W_Z = {}^tX\Lambda X$  où  $X$  est une matrice orthogonale, puisque  $Z$  est centrée, les colonnes de  ${}^tX$  sont orthogonales au vecteur  $\frac{1}{\sqrt{n}}\mathbf{1}$  et donc

$$\left( {}^tX, \frac{1}{\sqrt{n}}\mathbf{1} \right) \begin{pmatrix} \Lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ \frac{1}{\sqrt{n}}\mathbf{1} \end{pmatrix} = {}^tX\Lambda X + \frac{1}{n}\mathbf{1}\mathbf{1} = W_Z + \frac{1}{n}\mathbf{1}\mathbf{1}$$

Cette matrice a pour inverse généralisée de Moore Penrose

$$\left( W_Z + \frac{1}{n}\mathbf{1}\mathbf{1} \right)^+ = \left( {}^tX, \frac{1}{\sqrt{n}}\mathbf{1} \right) \begin{pmatrix} \Lambda^+ & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ \frac{1}{\sqrt{n}}\mathbf{1} \end{pmatrix}$$

où  $\Lambda^+$  est de terme général

$$\lambda_i^+ = \begin{cases} \lambda_i^{-1} & \text{si } \lambda_i \neq 0 \\ 0 & \text{sinon} \end{cases}$$

et donc

$$\left( W_Z + \frac{1}{n}\mathbf{1}\mathbf{1} \right)^+ = {}^tX\Lambda^+ X + \frac{1}{n}\mathbf{1}\mathbf{1} = W_Z^+ + \frac{1}{n}\mathbf{1}\mathbf{1}$$

où  $W_Z^+$  vérifie (21) si  $H$  est inversible. De plus, nous notons dans la suite  $S^2$  la matrice diagonale

$$S^2 = (W_Z^+ * I)^{-1}$$

D'autre part, la matrice orthogonale  $\begin{pmatrix} X \\ \frac{1}{\sqrt{n}}\mathbf{1} \end{pmatrix}$  où  $X \in \mathcal{M}_{n-1,n}(\mathbb{R})$  permet de représenter  $I$

par le simplexe  $\{P_i ; i \in [n]\}$  de  $E$ . Particularisons alors un point, par exemple  $P_1$  de vecteur de coordonnées  $\underline{x}_1$  la première colonne de  $X$ , et considérons l'hyperplan de  $E$  d'équation

$$\langle \underline{b}, \underline{u} \rangle + \frac{1}{n} = 0 \quad \text{où } \underline{b} = \Lambda^{-1/2} \underline{x}_1.$$

Alors pour tout  $i$  dans  $\{2, \dots, n\}$ , le point  $P_i$  de coordonnées

$$\underline{u}_i = \Lambda^{1/2} \underline{x}_i$$

vérifie

$${}^t \underline{x}_1 \Lambda^{-1/2} \Lambda^{1/2} \underline{x}_i = {}^t \underline{x}_1 \underline{x}_i = 0$$

c'est à dire que  $P_i$  appartient à l'hyperplan  $\mathcal{P}_1$  qui contient la face du simplexe  $\{P_i ; i \in [n]\}$  opposée au point  $P_1$ . La distance à cet hyperplan de tout point  $P$  de coordonnées  $\underline{u}$  dans  $E$  est

$$s_1 \left[ \langle \underline{b}, \underline{u} \rangle + \frac{1}{n} \right], \quad s_1 = \frac{1}{\|\underline{b}\|}$$

$s_1$  s'écrit de façon équivalente

$$s_1^{-2} = \|\underline{b}\|^2 = {}^t \underline{x}_1 \Lambda^{-1} \underline{x}_1 = \|\underline{x}_1\|_{\Lambda^{-1}}^2$$

En particulier  $P_1$  a pour coordonnées après normalisation

$$\underline{u}_1 = \Lambda^{1/2} \underline{x}_1$$

et donc, puisque  $X$  est une matrice orthogonale, la distance de  $P_1$  à l'hyperplan  $\mathcal{P}_1$  est

$$d(P_1, \mathcal{P}_1) = s_1 \left[ {}^t \underline{b} \underline{u}_1 + \frac{1}{n} \right] = s_1 \left[ {}^t \underline{x}_1 \Lambda^{-1/2} \Lambda^{1/2} \underline{x}_1 + \frac{1}{n} \right] = s_1 \left[ 1 + \frac{1}{n} \right]$$

Par suite, la matrice diagonale  $S^{-2}$  fournit, au coefficient multiplicatif  $s_1 \left[ 1 + \frac{1}{n} \right]$  près, les distances de chaque sommet  $P_i$  à la face opposée du simplexe. La formule (21) (lemme 6) fournit leur expression en fonction de  $H$ .

**Proposition 5:** Soient  $\underline{u}$  le vecteur de coordonnées du centre de l'hypersphère inscrite dans le simplexe  $\{P_i ; i \in [n]\}$  et  $r$  son rayon. Alors:

$$\underline{u} = \frac{\Lambda^{1/2} X S^{-1} \underline{1}}{{}^t \underline{1} S^{-1} \underline{1}}$$

$$r = \frac{1}{{}^t \underline{1} S^{-1} \underline{1}} \Leftrightarrow \frac{1}{r} = \sum_i \frac{1}{s_i}$$

**Preuve :** L'expression de  $d(P_i, \mathcal{P}_i)$  assure que  $\underline{u}$  et  $r$  sont solutions des équations

$$s_i \left[ \langle \underline{b}_i, \underline{u} \rangle + \frac{1}{n} \right] = r \quad \forall i \in [n]$$

c'est à dire, sous forme matricielle

$$S[\Lambda^{-1/2} \mathbf{u} + \frac{1}{n} \mathbf{1}] = r \mathbf{1} \quad (22)$$

$$(S\Lambda^{-1/2}, -1) \begin{pmatrix} \mathbf{u} \\ r \end{pmatrix} = -\frac{1}{n} S \mathbf{1} \quad (23)$$

La résolution de l'équation (23) fournit alors la solution. En effet  $X\mathbf{1} = \mathbf{0}$  puisque  $X$  est centré et l'équation (22) implique

$$\Lambda^{-1/2} \mathbf{u} + \frac{1}{n} \mathbf{1} = r S^{-1} \mathbf{1}$$

et donc

$$1 = \frac{1}{n} \mathbf{1}^T \mathbf{1} = r \mathbf{1}^T S^{-1} \mathbf{1}$$

On a donc bien

$$r = \frac{1}{\mathbf{1}^T S^{-1} \mathbf{1}}$$

Par suite l'équation (22) implique encore

$$\Lambda^{-1/2} \mathbf{u} = -\frac{1}{n} \mathbf{1} + r S^{-1} \mathbf{1}$$

et donc puisque  $X$  est une matrice orthogonale, telle que  $X^T X = I$ , on obtient

$$\Lambda^{-1/2} \mathbf{u} = -\frac{1}{n} X \mathbf{1} + r X S^{-1} \mathbf{1} = r X S^{-1} \mathbf{1}$$

d'où

$$\mathbf{u} = \frac{\Lambda^{1/2} X S^{-1} \mathbf{1}}{\mathbf{1}^T S^{-1} \mathbf{1}} \quad \parallel$$

Les résultats présentés dans le chapitre III permettent de définir une stratégie d'analyse d'un tableau de dissimilarité Euclidienne, puisqu'alors, le choix d'une origine  $P_0$  de  $E$  permet de déterminer toutes les informations nécessaires à la construction d'une solution. Comme nous venons de le voir, choisir une origine équivaut à fixer une pondération sur  $I$ . C'est pourquoi, on note dans la suite  $(I, \delta, P_1)$  où  $P_1$  est une pondération sur  $I$ , le triplet qui caractérise l'étude d'un espace quasi-métrique  $(I, \delta)$ . La présentation française suppose de plus que  $P_1$  est une mesure positive alors que la présentation anglaise de la méthode en termes de choix d'une origine dans  $E$  considère en fait le cas d'une mesure de signe quelconque. Dans le chapitre IV nous présentons alors les outils algébriques mis en œuvre dans l'analyse du triple  $(I, \delta, P_1)$ .



## CHAPITRE IV

### STRUCTURES LINEAIRES ET ANALYSE DU TRIPLE $(I, \delta, P_I)$

Soit  $(I, \delta)$  un espace quasi-métrique où  $I$  est supposé de cardinal  $n$  fini. On suppose donnée une bijection de  $I$  dans  $[n]$ . L'étude des propriétés caractéristiques d'un tableau de dissimilarité  $\Delta$  de terme général  $\delta_{ij} = \delta(i, j)$  peut être menée de façon équivalente dans les espaces vectoriels isomorphes (cf. LANG (1967), GREUB(1967)):

$\mathcal{M}_{n,n}(\mathbb{R})$  = l'espace vectoriel des matrices carrées d'ordre  $n$ ,  
 $\mathcal{L}_{\mathbb{R}}(F, F^*)$  = l'espace vectoriel des applications linéaires de  $F$  dans  $F^*$ ,  
 $\mathcal{L}_{\mathbb{R}}(F^*, F)$  = l'espace vectoriel des applications linéaires de  $F^*$  dans  $F$ ,  
 $\mathcal{L}_{\mathbb{R}}^2(F^*)$  = l'espace vectoriel des applications bilinéaires définies sur  $F \times F$ ,  
 $F \otimes F$  = le produit tensoriel de  $F$  par  $F$ . C'est le dual algébrique de  $\mathcal{L}_{\mathbb{R}}^2(F^*)$ .

Lorsque  $F$  est muni d'une base  $\mathcal{F} = \{\vec{f}_i ; i \in [n]\}$  et  $F^*$  est repéré par rapport à la base duale de  $\mathcal{F}$ , notée  $\mathcal{F}^* = \{\vec{f}_i^* ; i \in [n]\}$ , on associe alors à l'application  $a$  de  $\mathcal{L}_{\mathbb{R}}(F^*, F)$  le tenseur  $\underline{a}$  de  $F \otimes F$  et la matrice  $A$  qui lui est associée dans  $\mathcal{M}_{n,n}(\mathbb{R})$ .

#### IV.1. RAPPELS ET NOTATIONS

Nous nous intéressons ici au plongement Euclidien de  $I$ . On suppose donc qu'il existe un espace Euclidien  $E = (\mathbb{R}^p, m)$  où  $m$  est une forme quadratique définie sur  $\mathbb{R}^p$  et une application

$$\begin{aligned} x : I &\rightarrow E \\ i &\mapsto x(i) \end{aligned}$$

tels que les distances reconstruites, de carré  $(d_{ij}(x))^2 = m(x(i) - x(j))$ , représentent au mieux, au sens d'un critère à définir, les dissimilarités observées. Nous supposons de plus qu'un choix d'origine a été fait, de sorte que nous identifions la structure affine de  $E$  et la structure d'espace vectoriel sous-jacente induite par ce choix. Par suite nous regardons  $E$  comme un espace vectoriel et nous notons  $O$  son origine.

Supposons tout d'abord que  $\delta$  est une dissimilarité Euclidienne, c'est à dire qu'il existe un triplet  $(p, m, z)$  tel que  $\delta$  soit évalué dans  $E$  par

$$(\delta_{ij})^2 = m(z(i) - z(j)) \quad \forall i, j \in I$$

Dans la suite, nous notons  $z$  l'application associée à une dissimilarité Euclidienne, et nous précisons le caractère Euclidien de la dissimilarité par la notation  $\delta(z)$ . L'analyse linéaire de la représentation  $Z$  de  $I$  décrit alors classiquement  $z$  comme une application linéaire

$$z : F^* \rightarrow E$$

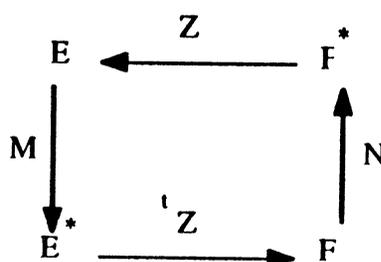
où chaque élément de  $I$  est associé à un vecteur de la base  $\{\vec{f}_i^*, i \in [n]\}$  de l'espace vectoriel réel libre  $F^*$  de dimension  $n$ .

**Remarque:** Nous adoptons dans la suite la convention usuelle en analyse des données, consistant à identifier par abus de langage  $\mathbb{R}^n$  et  $F$  (respectivement  $\mathbb{R}^p$  et  $E$ ) lorsque l'espace vectoriel  $\mathbb{R}^n$  (resp.  $\mathbb{R}^p$ ) définit la structure Euclidienne  $F = (\mathbb{R}^n, n)$  (resp.  $E = (\mathbb{R}^p, m)$ ). L'utilisation du dual algébrique  $F^*$  de l'espace vectoriel libre  $F$ , plutôt que  $F$  lui même, est arbitraire. Elle est retenue ici de façon à expliciter l'analogie des outils mis en oeuvre, avec le formalisme de l'analyse linéaire d'un tableau  $X$  de données de profil. Il est d'autre part souvent utile d'interpréter  $\mathbb{R}^p$  ( $p \leq n$ ) comme un sous-espace vectoriel de l'espace de référence  $\mathbb{R}^n$ .

Lorsqu'on a fait choix d'une base  $\{e^j, j \in [p]\}$ ,  $[p] = \{1, 2, \dots, p\}$  dans  $E$ , l'application linéaire  $z$  permet d'identifier  $z(i)$  au vecteur de coordonnées de sa représentation dans  $E$ :

$$z : \vec{f}_i^* \rightarrow z(\vec{f}_i^*) = \underline{z}_i$$

De plus  $m$  est alors décrite par la matrice  $M$  qui permet de résumer la situation par le schéma de dualité suivant



dans lequel la matrice  $N$  décrit dans la base  $\{f_i, i \in [n]\}$  la forme quadratique  $n$  qui définit la structure Euclidienne  $F = (\mathbb{R}^n, n)$ . Chaque méthode d'analyse est caractérisée en fait par l'argument utilisé pour faire le lien entre la matrice des observations  $\Delta$  et les constituants du schéma de dualité donné ci-dessus.

Associons aux deux espaces vectoriels  $E$  et  $F$ , l'ensemble  $\mathfrak{B}(F \times E, \mathbb{R})$  des formes bilinéaires définies sur  $F \times E$ . On note pour tout élément  $g$  de  $F^*$  et  $h$  de  $E^*$ ,  $g \otimes h$  l'application réelle définie sur  $F \times E$  par

$$\begin{aligned} g \otimes h &: F \times E \rightarrow \mathbb{R} \\ (\underline{x}, \underline{y}) &\mapsto g(\underline{x})h(\underline{y}) \end{aligned}$$

$g \otimes h$  est appelé le produit tensoriel de  $g$  et  $h$ . C'est une forme bilinéaire sur  $F \times E$ . Choisissons pour base de  $\mathfrak{B}(F \times E, \mathbb{R})$ , l'ensemble  $\mathcal{F}^* \otimes \mathcal{E}^*$  des  $np$  produits tensoriels  $\vec{f}_i^* \otimes \vec{e}_j^*$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, p$ , ordonnés lexicographiquement. Alors tout élément  $b$  de  $\mathfrak{B}(F \times E, \mathbb{R})$  s'écrit de façon unique

$$b = \sum_{i=1}^n \sum_{j=1}^p b_{ij} \vec{f}_i^* \otimes \vec{e}_j^* \quad (24)$$

et l'on note  $\underline{b}$  le vecteur à coordonnées réelles  $b_{ij}$  dans la base  $\mathcal{F}^* \otimes \mathcal{E}^*$ . Si  $\underline{x}$  est un élément de  $F$  et  $\underline{y}$  un élément de  $E$ , alors il résulte de (24)

$$b(\underline{x}, \underline{y}) = \sum_{i=1}^n \sum_{j=1}^p b_{ij} \vec{f}_i^*(\underline{x}) \vec{e}_j^*(\underline{y}) = \sum_{i=1}^n \sum_{j=1}^p b_{ij} x_i y_j$$

et en particulier  $b(\vec{f}_i, \vec{e}_j) = b_{ij}$ . On identifie dans la suite  $F^* \otimes E^*$  à l'espace vectoriel  $\mathfrak{B}(F \times E, \mathbb{R})$ . C'est le produit tensoriel de  $F^*$  par  $E^*$  et l'on définit l'application

$$\begin{aligned} \psi &: F^* \otimes E^* \rightarrow \mathfrak{L}_{\mathbb{R}}(F^*, E) \\ \underline{b} &\mapsto \psi(\underline{b}) : F^* \rightarrow E \\ \vec{f}_i^* &\mapsto \psi(\underline{b})(\vec{f}_i^*) = \sum_{j=1}^p b_{ij} \vec{e}_j \end{aligned}$$

$\psi$  est un isomorphisme d'espace vectoriel et l'on associe à  $\underline{b}$  la matrice  $B$  de terme général  $b_{ij}$ .  $B$  s'interprète alors aussi comme la matrice d'une application linéaire définie sur  $F^*$  à valeurs dans  $E$ , c'est à dire de l'élément  $\psi(\underline{b})$  de  $\mathfrak{L}_{\mathbb{R}}(F^*, E)$ . Alors on écrit

$$b(\underline{x}, \underline{y}) = \sum_{i=1}^n \sum_{j=1}^p b_{ij} x_i y_j = \underline{x} B \underline{y}$$

$B$  est une matrice réelle à  $n$  lignes et  $p$  colonnes et l'on note  $B \in \mathcal{M}_{n,p}(\mathbb{R})$ . Il existe donc un isomorphisme d'espace vectoriel entre  $F^* \otimes E^*$  et  $\mathcal{M}_{n,p}(\mathbb{R})$  que l'on note aussi  $\psi$ .

De plus, lorsque  $E$  et  $F$  sont munis respectivement des bases  $\{\vec{e}_j, j \in [p]\}$  et  $\{\vec{f}_i, i \in [n]\}$ , le produit tensoriel  $F \otimes E$  est muni naturellement de la base  $\{\vec{f}_i \otimes \vec{e}_j, i \in [n], j \in [p]\}$ , que nous

décrivons ici dans l'ordre lexicographique défini sur  $[n] \times [p]$ . L'application  $z$ , définie par le schéma de dualité ci-dessus, est associée à un tenseur de  $F \otimes E$ .  $z$  est alors décrite de façon équivalente soit par la matrice  $Z \in \mathcal{M}_{p \times n}(\mathbb{R})$ , soit par le vecteur  $\underline{z} \in \mathbb{R}^{np}$  déduit de  $Z$  par l'opérateur de vectorialisation  $\text{Vec}$  (cf. HENDERSON et SEARLE (1979)):

$$\underline{z} = \text{Vec}(Z) = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_i \\ \vdots \\ z_n \end{bmatrix} \quad \text{si} \quad Z = [ z_1 \quad z_2 \quad \dots \quad z_i \quad \dots \quad z_n ]$$

On vérifie alors simplement les relations élémentaires suivantes, utilisées dans la suite de ce travail, et dont on peut trouver une démonstration par exemple dans HENDERSON et SEARLE (1979), BALESTRA (1975), MAGNUS (1988) ou MAGNUS et NEUDECKER (1988) :

$$\begin{aligned} Y = AX {}^t B &\Leftrightarrow \text{Vec}(Y) = B \otimes A \text{Vec}(X) \\ &\Leftrightarrow \underline{y} = B \otimes A \underline{x} \end{aligned}$$

où l'on note  $B \otimes A$  le produit de Kronecker des matrices  $A$  et  $B$ , défini par  $B \otimes A = [b_{ij} A]_{ij}$

$$\text{Trace}({}^t A B) = {}^t \text{Vec}(A) \text{Vec}(B) = \underline{a} \underline{b}$$

et plus généralement, si  $X$ ,  $Y$ ,  $M$ , et  $N$  sont des matrices quelconques d'ordre tel que le produit  $M Y N {}^t X$  soit défini,

$$\begin{aligned} \text{Trace}(M Y N {}^t X) &= {}^t \text{Vec}(X) {}^t N \otimes M \text{Vec}(Y) = \underline{x} {}^t N \otimes M \underline{y} \\ &= {}^t \text{Vec}({}^t X) M \otimes N \text{Vec}({}^t Y) \end{aligned}$$

En particulier si les matrices  $M$  et  $N$  définissent les métriques respectives sur  $E$  et  $F$  et si  $X$  et  $Y$  sont deux éléments de  $\mathcal{M}_{p,n}(\mathbb{R})$  alors l'application

$$\langle \cdot, \cdot \rangle_{MN} : (X, Y) \mapsto \langle X, Y \rangle_{MN} = \text{Trace}({}^t X M Y N)$$

définit un produit scalaire sur  $\mathcal{M}_{p,n}(\mathbb{R})$  et

$$\begin{aligned} \langle X, X \rangle_{MN} &= \text{Trace}(M X N {}^t X) = {}^t \text{Vec}(X) N \otimes M \text{Vec}(X) = \underline{x} N \otimes M \underline{x} \\ &= \|\underline{x}\|_{N \otimes M}^2 \end{aligned}$$

La matrice symétrique semi-définie positive  $N \otimes M$  définit en effet une structure Euclidienne sur l'espace vectoriel  $\mathbb{R}^n \times \mathbb{R}^p$ , on pose alors

$$F \otimes E = (\mathbb{R}^n \times \mathbb{R}^p, n \otimes m)$$

et l'on note, avec les abus usuels de notation:

$$N \otimes M : F \otimes E \rightarrow F^* \otimes E^*$$

Le produit scalaire

$$\langle \underline{x}, \underline{y} \rangle_{N \otimes M} = \text{Trace}({}^t X M Y N)$$

permet alors de définir le coefficient de corrélation linéaire entre vecteurs centrés dans  $E \otimes F$ :

$$\begin{aligned} r(\underline{x}, \underline{y}) &= \frac{\langle \underline{x}, \underline{y} \rangle_{N \otimes M}}{\|\underline{x}\|_{N \otimes M} \cdot \|\underline{y}\|_{N \otimes M}} = \frac{\text{Trace}({}^t X M Y N)}{\sqrt{\text{Trace}({}^t X M X N)} \cdot \sqrt{\text{Trace}({}^t Y M Y N)}} \\ &= \frac{\text{Trace}({}^t X M Y N)}{\sqrt{\text{Trace}(W_X N)} \cdot \sqrt{\text{Trace}(W_Y N)}} \end{aligned}$$

c'est à dire le coefficient de corrélation entre opérateurs  $\text{COR}(X, Y)$ , étudié entre autres par MIZERE (1981).

La correspondance liant  $\langle \cdot, \cdot \rangle_{MN}$  et  $\langle \cdot, \cdot \rangle_{N \otimes M}$  définit une isométrie de structure Euclidienne entre  $\mathcal{M}_{p,n}(\mathbb{R})$  et  $F \otimes E$ . De plus, on vérifie le résultat élémentaire suivant :

**Lemme 1:** Soit  $\underline{c}$  un élément de  $\mathcal{M}_{p,1}(\mathbb{R})$  alors

$$\|X N \underline{c}\|_M \leq \|X\|_{MN} \cdot \|\underline{c}\|_N$$

et si  $\underline{x}_i$  est la  $i$ -ème colonne de la matrice  $X$ , alors

$$\sum_{i=1}^n n_i \|\underline{x}_i\|_M^2 \leq \text{Trace}(N) \cdot \|X\|_{MN}^2$$

Nous interprétons donc le plongement Euclidien de  $(I, \delta(Z))$  comme la recherche d'un sous-espace vectoriel  $\mathbb{R}^p$  de  $\mathbb{R}^n$ . Alors, pour un choix de base  $\mathfrak{Z} = \{\vec{e}_j, j \in [p]\}$  de  $\mathbb{R}^p$ , lorsque  $\mathbb{R}^p$  est muni d'une structure Euclidienne induite par la forme quadratique  $m$  de matrice  $M$  dans  $\mathfrak{Z}$ , la représentation  $Z$  de  $I$  dans  $E = (\mathbb{R}^p, M)$  définit un tenseur  $\underline{z}$  de l'espace  $F \otimes E$  qui est alors interprété comme un sous-espace vectoriel Euclidien de  $F \otimes F$  et une matrice  $Z$  de l'espace  $\mathcal{M}_{p,n}(\mathbb{R})$  qui est alors interprété comme un sous-espace vectoriel de  $\mathcal{M}_{n,n}(\mathbb{R})$ .

Nous montrons dans le paragraphe suivant que l'interprétation de  $F \otimes E$  comme sous-espace vectoriel Euclidien de  $F \otimes F$  est moins aisée qu'il n'est généralement admis. Elle repose en fait sur un isomorphisme de structures Euclidiennes dont la construction pose quelques difficultés en analyse des tableaux de dissimilarité. Cette difficulté a été notée par POLIT (1986) et d'AUBIGNY et POLIT (1988) en Analyse des Cubes de Données. Nous étudions tout d'abord la nature du problème. Puis nous donnons une formulation du problème traité par L'Analyse en Composantes Principales, exprimée dans  $F \otimes E$ . Ces résultats sont utiles à notre exposé des méthodes d'Analyse en Coordonnées Principales qui fait l'objet du paragraphe §V.1.1.

#### IV.2. LES SOUS-ESPACES $\langle E', F' \rangle$ DE $F^* \otimes E^*$

Soient  $E$  et  $F$  deux espaces vectoriels réels de dimension respective  $p$  et  $n$ . On note  $\mathcal{E}$  et  $\mathcal{F}$  les bases respectives de  $E$  et de  $F$  et l'on suppose connues les formes quadratiques  $m$  et  $n$  définissant une structure Euclidienne sur  $E$  et sur  $F$  respectivement. Leurs matrices représentatives dans les bases  $\mathcal{E}$  et  $\mathcal{F}$  sont respectivement  $M$  et  $N$ .

Soient  $E'$  et  $F'$  deux sous-espaces vectoriels de  $E$  et de  $F$ .  $E'$  (resp.  $F'$ ) est de dimension  $p'$  (resp.  $n'$ ). Alors  $(F')^* \otimes (E')^*$  est un espace vectoriel de dimension  $p'n'$ .  $(F')^* \otimes (E')^*$  est isomorphe à  $\mathcal{B}(F' \times E', \mathbb{R})$ . La difficulté provient du fait que  $(F')^* \otimes (E')^*$  n'est pas un sous-espace vectoriel de  $F^* \otimes E^*$ . En effet si  $h$  est un élément de  $(F')^* \otimes (E')^*$ , cette application a pour domaine de définition  $F' \times E'$  et non  $F \times E$ . Elle n'appartient donc pas à  $F^* \otimes E^*$  et il existe une infinité d'extensions  $h$  de  $h$  à  $F^* \otimes E^*$ . Les possibilités d'identification de  $(F')^* \otimes (E')^*$  à un sous-espace  $\langle E', F' \rangle$  de  $F^* \otimes E^*$  par isomorphisme jouent cependant un rôle central dans la démonstration des propriétés d'emboîtement des solutions du problème de plongement Euclidien d'un espace quasi-métrique  $(I, \delta)$ . La question posée s'exprime alors de la façon suivante :

Si  $\underline{x}$  (resp.  $\underline{x}'$ ) est un tenseur de  $F^* \otimes E^*$  (resp.  $(F')^* \otimes (E')^*$ ) associé à la matrice  $X$  réalisant le plongement isométrique Euclidien de  $I$  dans  $E$  (resp. dans  $E'$ ) avec  $\dim(E') = p' < p = \dim(E)$ , peut-on décrire  $E'$  comme un sous-espace affine de  $E$  ?

De façon classique, cette propriété est présentée comme une caractéristique souhaitable pour toute méthode d'Analyse des Données. Elle est en particulier satisfaite par l'Analyse en Composantes Principales et l'Analyse en Coordonnées Principales, alors qu'elle ne l'est pas par les méthodes d'Analyse Factorielle en Facteurs Communs et Spécifiques ou par les généralisations de l'Analyse en Composantes Principales aux cubes de données (cf. d'AUBIGNY et POLIT (1988)). Nous étudions la nature des choix implicites qui conduisent à satisfaire cette propriété d'emboîtement des solutions en Analyse en

Composantes Principales. Celle-ci résulte en fait d'un choix particulier d'une extension  $h$  de  $\hat{h}$  à  $F^* \otimes E^*$ . Posons en effet pour tout  $(\underline{x}, \underline{y}) \in F \times E$

$$h(\underline{x}, \underline{y}) = \hat{h}(\text{proj}_F(\underline{x}), \text{proj}_{E'}(\underline{y}))$$

où  $\text{proj}_F(\underline{x})$  (resp.  $\text{proj}_{E'}(\underline{y})$ ) est la projection  $N$ -orthogonale (resp.  $M$ -orthogonale) de  $\underline{x}$  (resp.  $\underline{y}$ ) sur  $F'$  (resp.  $E'$ ).

Cette construction d'un prolongement  $h$  nécessite, et c'est là le point fondamental, la connaissance des formes quadratiques  $m$  et  $n$ . Par suite, lorsque l'espace de représentation  $E$  est une inconnue du problème, on perd une donnée indispensable à la construction de solutions emboîtées. On vérifie de façon immédiate que  $\hat{h}$  ainsi définie est un élément de  $F \otimes E$ .

**Définition 1:** Soient  $E = (\mathbb{R}^p, m)$  et  $F = (\mathbb{R}^n, n)$  deux espaces vectoriels réels,  $E'$  et  $F'$  deux sous-espaces vectoriels de  $E$  et  $F$  respectivement. On appelle extension  $MN$ -canonique à  $F^* \otimes E^*$  de l'élément  $\hat{h}$  de  $(F')^* \otimes (E')^*$  l'application linéaire

$$\begin{aligned} h : F \times E &\rightarrow \mathbb{R} \\ (\underline{x}, \underline{y}) &\rightarrow \hat{h}(\text{Proj}_F(\underline{x}), \text{Proj}_{E'}(\underline{y})) \end{aligned}$$

où  $\text{Proj}_F$  et  $\text{Proj}_{E'}$  sont respectivement les projecteurs  $N$  et  $M$  orthogonaux sur  $F'$  et  $E'$ .

Notons  $\Phi_{MN}$  l'application définie sur  $(F')^* \otimes (E')^*$  à valeurs dans  $F^* \otimes E^*$  qui associe à tout  $\hat{h}$  son extension  $MN$ -canonique.  $\Phi_{MN}$  est clairement injective et linéaire. Elle définit un isomorphisme d'espace vectoriel entre  $(F')^* \otimes (E')^*$  et le sous-espace  $\langle E', F' \rangle = \Phi_{MN}((F')^* \otimes (E')^*)$  de  $F^* \otimes E^*$ .  $\langle E', F' \rangle$  est appelé le plongement canonique de  $(F')^* \otimes (E')^*$  dans  $F^* \otimes E^*$ .

Notons maintenant  $U = \{\underline{u}_s, s = 1, \dots, p'\}$  et  $V = \{\underline{v}_t, t = 1, \dots, n'\}$  deux bases respectives de  $E'$  et  $F'$ . Alors l'ensemble  $V^* \otimes U^* = \{\underline{v}_t^* \otimes \underline{u}_s^*, s = 1, \dots, p', t = 1, \dots, n'\}$  est une base de  $(F')^* \otimes (E')^*$ . Il lui correspond par  $\Phi_{MN}$  une base  $\Phi$  de  $\langle E', F' \rangle$  de terme général

$$\Phi_{MN}(\underline{v}_t^* \otimes \underline{u}_s^*) = \varphi_{ts}$$

Par linéarité de  $\Phi_{MN}$ , tout élément de  $\langle E', F' \rangle$  est déterminé par ses valeurs sur  $\Phi$ .  $\langle E', F' \rangle$  est donc de dimension  $p'n' \leq pn$ . Puisque  $F^* \otimes E^*$  est muni d'une structure Euclidienne de matrice  $Q = N^{-1} \otimes M^{-1}$ , on peut alors rechercher l'expression de la projection de tout tenseur  $\underline{x}$  de  $F^* \otimes E^*$  sur  $\langle E', F' \rangle$ . C'est le vecteur  $\hat{x}$  de  $\langle E', F' \rangle$  qui réalise le minimum

$$\text{Min}_{\underline{y} \in \langle E', F' \rangle} \|\underline{y} - \underline{x}\|_Q^2$$

Notons respectivement  $U$  et  $V$  les matrices de  $\mathcal{M}_{p,p}(\mathbb{R})$  et de  $\mathcal{M}_{n,n}(\mathbb{R})$  décrivant  $U$  et  $V$  dans les bases  $\mathfrak{E}$  et  $\mathfrak{F}$ . Ainsi la matrice  $U$  a pour colonne  $s$  le vecteur des coordonnées de  $u_s$  dans  $\mathfrak{E}$ . Supposons de plus que  $U$  et  $V$  sont des bases orthonormées au sens où  ${}^tUMU = I$  et  ${}^tVNV = I$

**Proposition 1 :** La matrice du projecteur orthogonal sur  $\langle E', F' \rangle$  s'écrit

$$\text{Proj}_{\langle E', F' \rangle} = NV{}^tV \otimes MU{}^tU$$

et si  $\hat{x}$  est la projection du vecteur  $x$  de  $F^* \otimes E^*$  sur  $\langle E', F' \rangle$ , alors  $\hat{x}$  a pour matrice représentative

$$\hat{X} = MU{}^tUXV{}^tVN = \text{Proj}_{E'} X {}^t\text{Proj}_{F'}$$

**Preuve :**  $\langle E', F' \rangle$  est un sous-espace vectoriel de  $F^* \otimes E^*$ , et donc il existe une solution unique

$$\hat{x} = \text{Proj}_{\langle E', F' \rangle}(x)$$

Montrons d'abord que la matrice du projecteur orthogonal sur  $\langle E', F' \rangle$  s'écrit

$$\text{Proj}_{\langle E', F' \rangle} = NV{}^tV \otimes MU{}^tU$$

On sait que

$$\text{Proj}_{E'} = U ({}^tUMU)^{-1} {}^tUM = U{}^tUM \quad (25)$$

$$\text{Proj}_{F'} = V ({}^tVNV)^{-1} {}^tVN = V{}^tVN \quad (26)$$

Le vecteur de base  $\varphi_{ts}$  est une forme linéaire particulière sur  $F \times E$  et donc

$$\begin{aligned} \Phi_{MN}(\underline{v}_t^* \otimes \underline{u}_s^*)(\vec{f}_i, \vec{e}_j) &= \underline{v}_t^* \otimes \underline{u}_s^*(\text{Proj}_{F'}(\vec{f}_i), \text{Proj}_{E'}(\vec{e}_j)) \\ &= \underline{v}_t^*(\text{Proj}_{F'}(\vec{f}_i)) \cdot \underline{u}_s^*(\text{Proj}_{E'}(\vec{e}_j)) \\ &= (NV)_{it} (MU)_{js} = (NV \otimes MU)_{ij,ts} \end{aligned}$$

puisque (25) et (26) induisent respectivement les relations

$$\underline{u}_s^*(\text{Proj}_{E'}(\vec{e}_j)) = (MU)_{js}$$

$$\underline{v}_t^*(\text{Proj}_{F'}(\vec{f}_i)) = (NV)_{it}$$

Par suite  $\Phi_{MN}$  a pour matrice représentative dans la base  $\mathfrak{F}^* \otimes \mathfrak{E}^*$  de  $F^* \otimes E^*$

$$B = NV \otimes MU$$

dont les colonnes sont les coordonnées des vecteurs de base  $\varphi_{ts}$  dans la base  $\mathfrak{F}^* \otimes \mathfrak{E}^*$ .  $B$  est de plein rang et le projecteur  $Q$ -orthogonal sur  $\langle E', F' \rangle$  s'écrit

$$\begin{aligned}
 \text{Proj}_{\langle E', F' \rangle} &= B ({}^tBQB)^{-1} {}^tBQ \\
 &= NV \otimes MU [({}^t(NV \otimes MU) N^{-1} \otimes M^{-1} (NV \otimes MU))]^{-1} ({}^t(NV \otimes MU) N^{-1} \otimes M^{-1}) \\
 &= (NV \otimes MU) [{}^tVNN^{-1}NV \otimes {}^tUMM^{-1}MU]^{-1} ({}^tVNN^{-1} \otimes {}^tUMM^{-1}) \\
 &= NV ({}^tVNV)^{-1} {}^tV \otimes MU ({}^tUMU)^{-1} {}^tU = NV {}^tV \otimes MU {}^tU
 \end{aligned}$$

et donc en appliquant les relations élémentaires présentées au paragraphe §IV.1

$$\hat{x} = \text{Proj}_{\langle E', F' \rangle}(\underline{x}) = NV {}^tV \otimes MU {}^tU \underline{x} \Leftrightarrow \hat{X} = MU {}^tUXV {}^tVN = \text{Proj}_{E'} X \text{Proj}_{F'} \quad \parallel$$

L'isomorphisme  $\psi$  entre  $(F')^* \otimes (E')^*$  et  $\langle E', F' \rangle$  permet d'associer à  $\hat{X}$  la matrice  $\hat{X}$  associée à la forme bilinéaire  $\hat{x} = \psi^{-1}(\hat{X})$

**Définition 2:** Si  $h$  est un élément de  $\langle E', F' \rangle$ , on appelle matrice centrale de  $h$ , l'élément  $\hat{H}$  de  $\mathcal{M}_{p,n}(\mathbb{R})$  associé à la forme bilinéaire réelle  $\hat{h} = \psi^{-1}(h)$ .

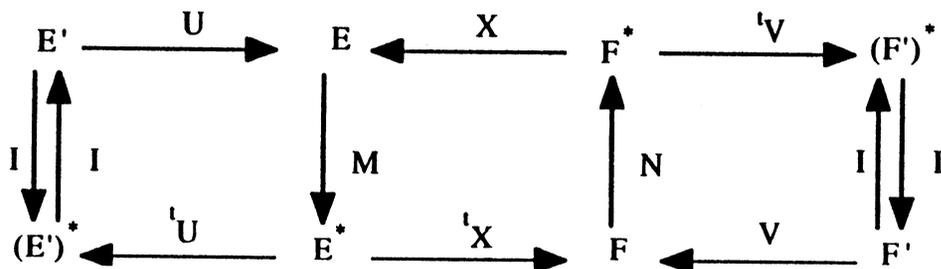
$\hat{H}$  est définie par les valeurs prises par  $h$  sur les vecteurs des bases  $U$  et  $V$ . Par suite la matrice centrale  $\hat{X}$  associée à  $\hat{x}$  vérifie

$$\begin{aligned}
 \hat{x}(\underline{u}_s, \underline{v}_t) &= {}^t\underline{u}_s (MU {}^tUXV {}^tVN) \underline{v}_t = (MU {}^tUXV {}^tVN)_{st} \\
 &= ({}^tUXV)_{st} \\
 &= {}^t\underline{u}_s X \underline{v}_t
 \end{aligned}$$

Par suite  $\hat{X} = {}^tUXV$ . C'est donc bien un élément de  $\mathcal{M}_{p,n}(\mathbb{R})$  et de plus

$$\hat{X} = MU \hat{X} {}^tVN$$

Résumons la situation par le schéma de dualité suivant



dont les éléments sont utilisés pour construire le sous-espace vectoriel le plus proche du support de  $\underline{x}$ . On parle alors du sous-espace le plus proche de  $\underline{x}$ .

### IV.3. RECHERCHE DU SOUS-ESPACE $\langle E', F' \rangle$ LE PLUS PROCHE DE $\underline{x}$

Supposons désormais  $E'$  et  $F'$  non fixés à priori et recherchons le sous-espace vectoriel  $\langle E', F' \rangle$  de  $F^* \otimes E^*$  qui réalise le minimum de

$$\| \underline{x} - \text{Proj}_{\langle E', F' \rangle}(\underline{x}) \|_Q.$$

Pour ce faire, nous notons  $O_{pp}^M$  (resp.  $O_{nn}^N$ ) le sous-ensemble de  $\mathcal{M}_{p,p}(\mathbb{R})$  (resp.  $\mathcal{M}_{n,n}(\mathbb{R})$ ), formé par les matrices  $M$ -orthogonales (resp.  $N$ -orthogonales), c'est à dire telles que  ${}^tAMA=I$  (resp.  ${}^tANA=I$ ) et

- (a)  $\Gamma_0(p', n') = \{ (E', F') ; E' \text{ sous espace vectoriel de } E \text{ et } F' \text{ sous-espace vectoriel de } F \text{ tels que } \dim(E') = p' \text{ et } \dim(F') = n' \}$   
 (b)  $\Gamma_1(p', n') = \{ (U, V) ; U \in O_{pp}^M, V \in O_{nn}^N \}$

Nous nous proposons donc de résoudre le problème (P) suivant :

Etant donné  $\underline{x} \in F^* \otimes E^*$ , trouver  $(\bar{E}', \bar{F}') \in \Gamma_0(p', n')$  et  $\hat{\underline{x}} \in \langle \bar{E}', \bar{F}' \rangle$  réalisant le minimum de

$$\| \underline{x} - \hat{\underline{x}} \|_Q.$$

Lorsque (P) admet une solution, il s'exprime de façon équivalente par la proposition suivante

**Proposition 2:** La solution  $\langle \bar{E}', \bar{F}' \rangle$  de (P) résout l'un des problèmes équivalents

$$i) \text{ Min } \{ \| \underline{x} - \text{Proj}_{\langle E', F' \rangle}(\underline{x}) \|_Q ; (\bar{E}', \bar{F}') \in \Gamma_0(p', n') \} \quad (27)$$

$$ii) \text{ Max } \{ \| \text{Proj}_{\langle E', F' \rangle}(\underline{x}) \|_Q ; (\bar{E}', \bar{F}') \in \Gamma_0(p', n') \} \quad (28)$$

$$iii) \text{ Max } \{ \| NV^tV \otimes MU^tU \underline{x} \|_C^2 ; (U, V) \in \Gamma_1(p', n') \} \quad (29)$$

**Preuve :** i) est immédiat et ii) résulte du fait que  $\text{Proj}_{\langle E', F' \rangle}$  est une projection  $Q$ -orthogonale. iii) est dû à la caractérisation d'un sous-espace vectoriel par la donnée d'une base. On pose donc  $E' = \mathcal{L} \text{ in } \langle U \rangle$  et  $F' = \mathcal{L} \text{ in } \langle V \rangle$ .  $\square$

En particulier l'optimum  $\bar{E}' = \mathcal{L} \text{ in } \langle \bar{U} \rangle$  et  $\bar{F}' = \mathcal{L} \text{ in } \langle \bar{V} \rangle$  où  $(\bar{U}, \bar{V})$  est solution de iii) et le problème (29) s'écrit de façon équivalente sous l'une des formes plus classiques

$$\text{Max } \{ \| MU^tUXV^tVN \|_{M^{-1}, N^{-1}}^2 ; (U, V) \in \Gamma_1(p', n') \} \quad (30)$$

ou encore

$$\text{Min } \{ \| X - MU^tUXV^tVN \|_{M^{-1}, N^{-1}}^2 ; (U, V) \in \Gamma_1(p', n') \} \quad (31)$$

POLIT (1986) montre le résultat suivant, dont nous donnons ici une démonstration qui corrige les erreurs de cet auteur

**Proposition 3:** *La solution de (31) est fournie par le couple  $(\bar{U}, \bar{V})$  de  $\Gamma_1(p', n')$  solution du système*

$$T_u V = NV\Lambda_1 \quad T_u = 'XU'UX \quad (32)$$

$$T_v U = MU\Lambda_2 \quad T_v = XV'V'X \quad (33)$$

où  $\Lambda_1$  et  $\Lambda_2$  sont des matrices diagonales.

**Preuve :**

$$\begin{aligned} \parallel MU 'UXV 'VN \parallel_{M^{-1}, N}^2 &= \text{Trace}(' (MU 'UXV 'VN) M^{-1} (MU 'UXV 'VN) N^{-1} ) \\ &= \text{Trace}('UXV'V'XU) = \text{Trace}('UT_v U) \\ &= \text{Trace}('V'XU'UXV) = \text{Trace}('VT_u V) \end{aligned}$$

où par définition  $T_u$  et  $T_v$  sont définies positives. Supposons par exemple que  $U$  est fixée et égale à  $\bar{U}$ . Alors la matrice  $\bar{V}$  réalise la maximum de  $\text{Trace}('VT_u V)$  si et seulement si elle est la matrice des vecteurs propres  $N$ -orthonormés associés aux  $n'$  plus grandes valeurs propres de  $N^{-1}T_u$

$$T_u \bar{V} = N\bar{V}\Lambda_1$$

Un raisonnement analogue conduit à la deuxième équation. ¶

On retrouve alors les caractérisations usuelles de la solution d'une Analyse en Composantes Principales :

**Corollaire :** *Sous l'hypothèse  $n' = p'$ , l'élément  $(\bar{U}, \bar{V})$  de  $\Gamma_1(p', n')$  est solution du système*

$$RU = UA \quad R = XN'XN = V_X M \quad (34)$$

$$SV = VA \quad S = 'XMXN = W_X N \quad (35)$$

**Preuve :** Il est clair que  $\text{rang}(R) = \text{rang}(S) = \text{rang}(X)$  (cf. CAILLEZ et PAGES (1976)). Notons  $r$  cette valeur commune. Alors par (31) on montre que  $\Lambda_1 = \Lambda_2 = \Lambda$  et si l'on tient compte de l'ordre de multiplicité des valeurs propres, il en existe  $r$  positives. Supposons donc  $n' = p' = s \leq r$  et considérons (sans perte de généralité) les  $s$  premières valeurs propres distinctes. Les matrices  $U \in O_{ps}^M$  et  $V \in O_{ns}^N$  définies par (32) et (33) vérifient les équations de transition

$$U = XNV\Lambda^{-1/2}$$

$$V = {}^tXMU\Lambda^{-1/2}$$

En effet  $XNV\Lambda^{-1/2} \in O_{ps}^M$  puisque

$${}^t(XNV\Lambda^{-1/2})M(XNV\Lambda^{-1/2}) = \Lambda^{-1/2} {}^tVN {}^tXMXNV\Lambda^{-1/2} = \Lambda^{-1/2} {}^tVNW_XNV\Lambda^{-1/2}$$

et d'après (33)

$$= \Lambda^{-1/2} {}^tVNV\Lambda^{1/2} = \Lambda^{-1/2} {}^t\Lambda^{1/2} = I$$

De plus la matrice  $XNV\Lambda^{-1/2}$  a pour colonnes des vecteurs propres de R puisque (33) permet d'écrire

$$(XN {}^tXM)(XNV\Lambda^{-1/2}) = XNSV\Lambda^{-1/2} = XNV\Lambda^{1/2} = (XNV\Lambda^{-1/2})\Lambda$$

La relation (32) conduit à des calculs analogues pour la matrice  $V = {}^tXMU\Lambda^{-1/2}$  ¶

**Proposition 4 (POLIT 1986):** Lorsque  $n'=p'=s \leq r$  et que les valeurs propres  $\lambda_i$  sont simples, alors les matrices U et V définies par (32) et (33) fournissent les solutions de (30) et (31) sous la forme

$$\bar{U} = MU \text{ et } \bar{V} = NV$$

**Preuve :** Substituons  $\bar{U}$  par son expression dans (32), on obtient à l'aide des équations de transition

$$T_{\bar{u}} = {}^tX\bar{U} {}^t\bar{U}X = {}^tXMU {}^tUMX = {}^tXMNV\Lambda^{-1/2}\Lambda^{-1/2} {}^tVN {}^tXMX$$

$$= W_XNV\Lambda^{-1} {}^tVNW_X$$

et donc

$$T_{\bar{u}}\bar{V} = W_XNV\Lambda^{-1} {}^tVNW_XNV = NV = V\Lambda = N\bar{V}\Lambda$$

On montre de façon analogue que  $T_{\bar{v}}\bar{U} = M\bar{U}\Lambda$  ¶

Par suite le triplet (U, V,  $\Lambda$ ) est solution de l'Analyse en Composantes Principales du triplet (X, M, N). Rappelons alors la propriété bien connue vérifiée par la solution de l'Analyse en Composantes Principales

**Proposition 5:** Soit  $s = \text{Min}\{r, p', n'\}$  où  $r = \text{rang}(X) \leq p' \leq p$  et  $n' \leq n$ . Si  $(U, V, \Lambda)$  est solution d'ordre  $s$  de l'Analyse en Composantes Principales du triplet  $(X, M, N)$ , alors pour toutes matrices  $A$  de  $\mathcal{M}_{p,p}(\mathbb{R})$ ,  $B$  de  $\mathcal{M}_{n,n}(\mathbb{R})$  et  $H$  de  $\mathcal{M}_{p',n'}(\mathbb{R})$  on a

$$i) \|X - AH^t B\|_{MN} \geq \|X - U\Lambda^{1/2} V^t\|_{MN}$$

$$ii) \text{Min} (\|X - AH^t B\|_{MN}^2; A, B, H) = \|X - U\Lambda^{1/2} V^t\|_{MN}^2 = \sum_{t=s+1}^r \lambda_t(XMXN)$$

L'application de ces résultats en Analyse en Coordonnées Principales impose d'étudier le cas particulier caractérisé par la recherche du sous-espace  $\langle E', F' \rangle$  de  $F^* \otimes E^*$  qui réalise la meilleure approximation de rang  $p'$  de  $X$ . Nous reprenons à cet effet, avec quelques modifications un résultat établi par POLIT (1986) :

**Proposition 6:** Soit  $s = \text{Min}\{r, p', n'\}$  et  $(U, V, \Lambda)$  une solution d'ordre  $s$  de l'Analyse en Composantes Principales du triplet  $(X, M, N)$ . Soit  $\tilde{U}$  un élément de  $O_{pp'}^M$  (resp.  $\tilde{V}$  de  $O_{nn'}^N$ ) obtenu par complétion de  $U$  (resp.  $V$ ) avec  $(p'-1)$  (resp.  $(n'-1)$ ) colonnes  $M$  (resp.  $N$ ) -orthogonales. Soient  $\bar{U}$  et  $\bar{V}$  les matrices définies par  $\bar{U} = M\tilde{U}$  et  $\bar{V} = N\tilde{V}$ , alors

i)  $(\bar{U}, \bar{V})$  réalise le minimum de (30) qui vaut alors

$$\sum_{t=s+1}^r \lambda_t(XMXN)$$

ii)  $\hat{x}$  est l'élément de  $F^* \otimes E^*$  dont la matrice associée dans les bases  $\mathcal{E}$  et  $\mathcal{F}$  de  $E$  et  $F$  est

$$\hat{X} = M\bar{U}^t \bar{U} X \bar{V}^t \bar{V} N$$

De plus  $(\mathcal{L} \text{ in } \langle \bar{U} \rangle, \mathcal{L} \text{ in } \langle \bar{V} \rangle)$  est un élément de  $\Gamma_0$  et  $\hat{x}$  est solution du problème (P).

**Preuve (POLIT (1986)):** On peut supposer sans perte de généralité  $p' \leq n'$ . Alors trois cas doivent être envisagés

- 1 :  $p' = n' = r (= s)$
- 2 :  $p' \neq n', p' \leq r (s = p')$
- 3 :  $p' \neq n', p' > r (s = r)$

Dans le cas 1, notons  $\bar{U} = MU$  et  $\bar{V} = NV$  où  $(U, V, \Lambda)$  est une solution d'ordre  $s$  de l'Analyse en Composantes Principales du triplet  $(X, M, N)$ . Supposons qu'il existe  $A$  dans  $O_{ps}^{M^{-1}}$  et  $B$  dans  $O_{ns}^{N^{-1}}$  telles que

$$\|X - M^{-1} A^t A X B^t B N^{-1}\|_{MN} < \|X - M^{-1} \bar{U}^t \bar{U} X \bar{V}^t \bar{V} N^{-1}\|_{MN}$$

Le terme de droite est égal à  $\|X - U\Lambda^{1/2}V\|_{MN}$  et on arrive donc à une contradiction du résultat i) de la proposition 5 puisque  $M^{-1}A \in \mathcal{M}_{p,p}(\mathbb{R})$ ,  $N^{-1}B \in \mathcal{M}_{n,n}(\mathbb{R})$ ,  $H = {}^tAXB \in \mathcal{M}_{p,n}(\mathbb{R})$  et  $s = p'$ .  $(\bar{U}, \bar{V})$  est donc solution de (31) et définit une solution de (P).

Dans le cas 2, les matrices  $\bar{U}$  et  $\bar{V}$  construites à partir de  $\tilde{U}$  et  $\tilde{V}$  vérifient

$$\begin{aligned} A &= \|X - M^{-1}\bar{U} {}^t\bar{U}X\bar{V} {}^t\bar{V}N^{-1}\|_{MN} = \|X - \tilde{U} {}^t\tilde{U}MXN\tilde{V} {}^t\tilde{V}\|_{MN} \\ &= \|X - \tilde{U} {}^t\tilde{U}M\tilde{U}\Lambda^{-1/2} {}^t\tilde{V}\|_{MN} \end{aligned}$$

et donc par définition de  $\tilde{U}$

$$A = \|X - U\Lambda^{1/2}V\|_{MN}$$

Appliquons comme dans le cas 1 la propriété i) de la proposition 5 qui fournit le résultat souhaité.

Dans le cas 3, complétons U et V comme précédemment. Alors

$$\tilde{U} = (U \mid U_0) \text{ et } \tilde{V} = (V \mid V_0)$$

où les colonnes de  $U_0$  (resp.  $V_0$ ) appartiennent au noyau de  ${}^tXMXN$  (resp.  $XN{}^tXM$ ). Par suite

$${}^t\tilde{U}MXN\tilde{V} = \begin{pmatrix} \Lambda^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

puisque  $XNV = U\Lambda^{1/2}$  et  $U \in O_{p,r}^M$ . On vérifie la relation

$$\|X - M^{-1}\bar{U} {}^t\bar{U}X\bar{V} {}^t\bar{V}N^{-1}\|_{MN} = \|X - \tilde{U}({}^t\tilde{U}MXN\tilde{V}) {}^t\tilde{V}\|_{MN} = \|X - U\Lambda^{1/2}V\|_{MN}$$

Comme précédemment  $(\bar{U}, \bar{V})$  fournit la solution souhaitée.  $\square$

La proposition 6 montre que si  $n' \geq s$ , la valeur de l'optimum (28) ne dépend pas de  $n'$ . Elle reste donc inchangée et égale à

$$\sum_{t=1}^s \lambda_t({}^tXMXN)$$

si l'on substitue  $n''$  à  $n'$  avec  $s \leq n'' \leq n'$ . Il en est de même pour le minimum de (27) qui est alors égal à

$$\sum_{t=s+1}^r \lambda_t ({}^t X M X N)$$

Minimiser (27) ou de façon équivalente, maximiser (28) dans  $\Gamma_0(p', n')$  ou dans  $\Gamma_0(p', n'')$  conduit à la même solution. C'est en particulier le cas pour  $F'=F$  c'est à dire pour  $n'=n$ . Mais

$$\text{Max} \{ \| \text{Proj}_{\langle E', F' \rangle}(\underline{x}) \|_Q^2; E' \text{ sous-espace vectoriel de } E, \dim(E')=p' \} = \text{Trace}(\Lambda) \quad (36)$$

où  $\Lambda$  est la matrice diagonale des valeurs propres issue de l'Analyse en Composantes Principales du triplet  $(X, M, N)$ . De façon symétrique, lorsque  $n' < p'$  on permute le rôle de  $E'$  et de  $F'$ , de sorte que l'on a aussi

$$\text{Max} \{ \| \text{Proj}_{\langle E', F' \rangle}(\underline{x}) \|_Q^2; F' \text{ sous-espace vectoriel de } F, \dim(F')=n' \} = \text{Trace}(\Lambda) \quad (37)$$

L'expression de la solution des problèmes (27), (36) et (37) repose sur la caractérisation de  $\hat{\underline{x}}$  comme projection orthogonale de  $\underline{x}$  sur un sous-espace  $\langle E', F' \rangle$  de  $F^* \otimes E^*$  de dimension donnée et dépend donc explicitement de la connaissance à priori de  $E = (\mathbb{R}^p, M)$  et  $F = (\mathbb{R}^n, N)$ . En Codage Multidimensionnel,  $\underline{x}$  est une inconnue. Etant donnée une matrice de dissimilarité, l'espace de représentation  $E$  est une inconnue du problème. Par suite il est impossible d'associer à deux espaces de représentation  $E'$  et  $E''$  de dimension respectives  $p'$  et  $p''$  avec  $p'' < p'$ , deux sous-espaces  $\langle E', F' \rangle$  et  $\langle E'', F'' \rangle$  de  $F^* \otimes E^*$  vérifiant la condition d'emboîtement  $\langle E'', F'' \rangle \subset \langle E', F' \rangle$ . On ne peut donc pas vérifier de propriété d'emboîtement des solutions d'un problème de Codage Multidimensionnel. De même, dans le cas d'une dissimilarité Euclidienne  $\delta(Z)$ , l'isomorphisme  $\psi$  étant inconnu, la solution  $\underline{x} = \hat{\underline{z}}$  ne peut être caractérisée comme la projection de  $\underline{z}$  sur un sous-espace vectoriel de  $F^* \otimes E^*$ .

#### IV.4. LES STRUCTURES LINÉAIRES EN ANALYSE DU TRIPLE $(I, \delta, P_I)$

Notons respectivement  $\mathcal{Y}_n$  l'ensemble des matrices réelles symétriques d'ordre  $n$ . C'est un sous-espace vectoriel de  $\mathcal{M}_{n,n}(\mathbb{R})$ , de dimension  $n(n+1)/2 = n^* + n$  où  $n^* = n(n-1)/2$ . De façon classique, le supplémentaire de  $\mathcal{Y}_n$  dans  $\mathcal{M}_{n,n}(\mathbb{R})$  est le sous-espace vectoriel  $\mathcal{A}_n$  de dimension  $n^*$  formé par les matrices antisymétriques et

$$\mathcal{M}_{n,n}(\mathbb{R}) = \mathcal{Y}_n \oplus \mathcal{A}_n$$

et toute matrice  $B$  élément de  $\mathcal{M}_{n,n}(\mathbb{R})$  se décompose de façon unique en

$$B = S + A$$

où

$$S = \frac{B + {}^tB}{2} \quad S \in \mathcal{Y}_n, \quad A = \frac{B - {}^tB}{2} \quad A \in \mathcal{A}_n$$

Définissons maintenant sur  $\mathcal{M}_{n,n}(\mathbb{R})$  les deux opérateurs linéaires  $\alpha$  et  $\beta$  définis pour tout élément  $A$  de  $\mathcal{M}_{n,n}(\mathbb{R})$  par

$$\alpha(A) = -\frac{1}{2} Q_1 A {}^tQ_1$$

$$\beta(A) = (A * I) \mathbb{1}_n \mathbb{1}_n + \mathbb{1}_n \mathbb{1}_n (A * I) - 2A$$

Alors

$$\text{Ker}(\beta) = \{ A \in \mathcal{M}_{n,n}(\mathbb{R}) ; a_{ij} = \frac{1}{2}(a_{ii} + a_{jj}) \}$$

et donc le vecteur  $\underline{x}$  de terme général  $x_i = \frac{1}{2}a_{ii}$  permet d'écrire le noyau de  $\beta$  sous la forme

$$\text{Ker}(\beta) = \{ A \in \mathcal{M}_{n,n}(\mathbb{R}) ; A = \underline{x} \mathbb{1}_n + \mathbb{1}_n {}^t\underline{x} \}$$

De même puisque  $Q_1$  est le projecteur orthogonal sur l'orthogonal de la droite des constantes de  $F$

$$\text{Ker}(\alpha) = \{ A \in \mathcal{M}_{n,n}(\mathbb{R}) ; \exists \underline{x}, \underline{y} \in \mathbb{R}^n \text{ tels que } A = \underline{x} \mathbb{1}_n + \mathbb{1}_n {}^t\underline{y} \}$$

Considérons le cas particulier des matrices réelles carrées à diagonale nulle. Alors  $\alpha$  est nulle pour des matrices de structure très particulières

**Proposition 7:** Soit  $A$  un élément de  $\mathcal{M}_{n,n}(\mathbb{R})$  tel que  $A * I = 0$ , alors

$$i) (A \in \text{Ker}(\alpha)) \Rightarrow (\exists \underline{x} \in \mathbb{R}^n \text{ tel que } A = \underline{x} \mathbb{1}_n - \mathbb{1}_n {}^t\underline{x})$$

ii) en particulier  $A$  appartient à  $\mathcal{A}_n$ .

**Preuve :** Si  $A$  appartient à  $\text{Ker}(\alpha)$  alors il existe deux vecteurs  $\underline{x}$  et  $\underline{y}$  de  $\mathbb{R}^n$  tels que

$$A = \underline{x} \mathbb{1}_n + \mathbb{1}_n {}^t\underline{y}$$

$$A * I = (\underline{x} \mathbb{1}_n) * I + (\mathbb{1}_n {}^t\underline{y}) * I = D_{\underline{x}} + D_{\underline{y}}$$

Par suite  $A * I = 0$  si et seulement si  $D_{\underline{x}} = -D_{\underline{y}}$ , c'est à dire  $\underline{x} = -\underline{y}$  ¶

De façon complémentaire, considérons un élément de  $\text{Ker}(\alpha) \cap \mathcal{Y}_n$ . On vérifie alors que la structure de  $A$  est très simple

**Proposition 8:** Les trois propositions suivantes sont équivalentes :

- i)  $A \in \text{Ker}(\alpha) \cap \mathcal{Y}_n$
- ii)  $A \in \text{Ker}(\alpha)$  et  $(A \underline{1}_n = {}^t A \underline{1}_n)$
- iii)  $A \in \mathcal{M}_{n,n}(\mathbb{R})$  et  $\exists \underline{x} \in \mathbb{R}^n$  tel que  $A = \underline{x} \underline{1}'_n + \underline{1}_n \underline{x}'$

**Preuve :** Soit  $A$  un élément de  $\text{Ker}(\alpha) \cap \mathcal{Y}_n$ , alors  $A \underline{1}_n = {}^t A \underline{1}_n$ .

Réciproquement si ii) est vérifiée, alors

$$\begin{aligned} \alpha(A) &= -\frac{1}{2} Q_1 A {}^t Q_1 = -\frac{1}{2} \left[ A - \frac{1}{n} \underline{1}' \underline{1} A - \frac{1}{n} A \underline{1}' + \frac{1}{n^2} \underline{1}' \underline{1} A \underline{1}' \right] \\ \alpha(A) = 0 &\Leftrightarrow A = \frac{1}{n} \underline{1}' \underline{1} A + \frac{1}{n} A \underline{1}' - \frac{1}{n^2} \underline{1}' \underline{1} A \underline{1}' \\ &= \frac{1}{n} \underline{1}' \underline{1}' A + \frac{1}{n} {}^t A \underline{1}' \underline{1} - \frac{1}{n^2} \underline{1}' \underline{1}' A \underline{1}' \underline{1}' \end{aligned}$$

et puisque par définition de  $\alpha$ ,  $\alpha({}^t A) = \alpha(A) = 0$ . On a donc  $A = {}^t A$ .

Si iii) est vraie, alors  $A \in \text{Ker}(\alpha)$  et  $A = {}^t A$ . De plus la définition de  $\text{Ker}(\alpha)$  implique iii) si i) est vérifiée.  $\square$

Les deux propositions précédentes ont pour conséquence directe :

**Corollaire :** Si  $A$  est un élément de  $\mathcal{M}_{n,n}(\mathbb{R})$  tel que  $A * I = 0$  et  $A \underline{1}_n = {}^t A \underline{1}_n$ , alors

$$(A \in \text{Ker}(\alpha)) \Leftrightarrow (A = 0)$$

**Preuve :** En effet si  $A = 0$ , clairement  $A \in \text{Ker}(\alpha)$ . Réciproquement, si  $A \in \text{Ker}(\alpha)$  et  $A \underline{1}_n = {}^t A \underline{1}_n$  alors  $A \in \text{Ker}(\alpha) \cap \mathcal{Y}_n$  par la proposition 8. Elle doit appartenir à  $\mathcal{A}_n$  puisque  $A * I = 0$ , par la proposition 7. On a donc  $A = 0$ .  $\square$

De même

**Lemme 2:** Si  $A \underline{1}_n = 0$ , ou bien  ${}^t A \underline{1}_n = 0$  ou encore  $A * I = 0$ , alors  $A \in \text{Ker}(\beta) \Leftrightarrow A = 0$

**Preuve :** Si  $A = 0$ , clairement  $A \in \text{Ker}(\beta)$ . Réciproquement,

$$A \in \text{Ker}(\beta) \Leftrightarrow A = \frac{1}{2} [(A * I) \underline{1}_n \underline{1}'_n + \underline{1}_n \underline{1}'_n (A * I)]$$

de sorte que le résultat est trivialement vérifié si  $A * I = 0$ .

$$A \underline{1}_n = \frac{1}{2} [n ((A * I) \underline{1}_n + \underline{1}_n \underline{1}'_n (A * I) \underline{1}_n)] \Rightarrow \underline{1}'_n A \underline{1}_n = n [\underline{1}'_n (A * I) \underline{1}_n]$$

donc  $A \underline{1}_n = 0$  ou  ${}^t A \underline{1}_n = 0$  impliquent  $A * I = 0$  et donc le résultat souhaité.  $\square$

Ces propriétés de symétrie et d'antisymétrie d'une matrice carrée d'ordre  $n$  sont complétées par le résultat suivant dû à CRITCHLEY (1982a)

**Proposition 9:**

i)  $\alpha({}^tA) = {}^t\alpha(A)$  et  $\beta({}^tA) = {}^t\beta(A)$

ii)  $\alpha(A) = {}^t\alpha(A) \Leftrightarrow \exists \underline{c} \in \mathbb{R}^n$  tel que  $\sum_{i=1}^n c_i = 0$  et  $A - {}^tA = \underline{c} \mathbf{1}_n - \mathbf{1}_n \underline{c}$

iii)  $\beta({}^tA) = {}^t\beta(A) \Leftrightarrow A = {}^tA$

iv) Si  $A * I = 0$ , alors les trois propriétés suivantes sont équivalentes :

$\alpha(A)$  est antisymétrique

$A$  est antisymétrique,

$\beta(A)$  est antisymétrique

La décomposition de  $\mathcal{M}_{n,n}(\mathbb{R})$  correspond par isomorphisme à la décomposition de  $F \otimes F$  en

$$F \otimes F = (F \vee F) \oplus (F \wedge F).$$

Le sous-espace vectoriel  $F \vee F$  de  $F \otimes F$  est isomorphe à  $\mathfrak{V}_n$ . De même le sous-espace  $F \wedge F$  de  $F \otimes F$  est isomorphe à  $\mathfrak{A}_n$  et est appelé le produit extérieur de  $F$  par lui-même. Si  $F$  est muni de la base  $\mathcal{F} = \{f_i ; i \in [n]\}$ , alors  $F \vee F$ , (resp.  $F \wedge F$ ) a pour base

$$\mathcal{F} \vee \mathcal{F} = \{f_i \vee f_j ; i, j \in [n]\} \text{ (resp. } \mathcal{F} \wedge \mathcal{F} = \{f_i \wedge f_j ; i, j \in [n]\})$$

où

$$f_i \vee f_j = \frac{1}{2} [f_i \otimes f_j + f_j \otimes f_i]$$

$$f_i \wedge f_j = \frac{1}{2} [f_i \otimes f_j - f_j \otimes f_i] \quad i \neq j$$

Dans la suite, on suppose que les couples  $(i, j)$  indiquant ces bases sont décrits dans l'ordre lexicographique.

Ces deux relations définissent les matrices des opérateurs  $\tau_{F \vee F}$  et  $\tau_{F \wedge F}$  définis par

$$\tau_{F \vee F} : F \otimes F \rightarrow F \vee F$$

$$f_i \otimes f_j \mapsto f_i \vee f_j$$

$$\tau_{F \wedge F} : F \otimes F \rightarrow F \wedge F$$

$$f_i \otimes f_j \mapsto f_i \wedge f_j \quad i \neq j$$

Par suite l'isomorphisme de structure vectorielle  $\psi$  entre  $\mathcal{M}_{n,n}(\mathbb{R})$  et  $F \otimes F$ , noté  $\text{Vec}$  plus haut, induit deux isomorphismes de structure vectorielle :

$$\begin{aligned} \psi_S : \mathcal{Y}_n &\rightarrow F \vee F \\ N = (n_{ij}) &\mapsto \underline{n} = \tau_{F \vee F} \circ \psi(N) \quad (= n_{ij}; i \leq j \quad i, j = 1, 2, \dots, n) \end{aligned}$$

$$\begin{aligned} \psi_A : \mathcal{A}_n &\rightarrow F \wedge F \\ A = (a_{ij}) &\mapsto \underline{a} = \tau_{F \wedge F} \circ \psi(A) \quad (= a_{ij}; i < j \quad i, j = 1, 2, \dots, n) \end{aligned}$$

et pour tout  $\underline{x} \in F \otimes F$ , on vérifie alors que

$$\begin{aligned} \tau_{F \vee F}(\underline{x}) &= (x_{ij}; i \leq j \quad i, j \in [n]) \\ \tau_{F \wedge F}(\underline{x}) &= (x_{ij}; i < j \quad i, j \in [n]) \end{aligned}$$

MAGNUS (1986) décrit les outils algébriques d'étude des sous-espaces vectoriels qu'il appelle des structures linéaires sur  $F \otimes F$  et étudie en détail les propriétés des représentations matricielles de  $\tau_{F \wedge F}$  et  $\tau_{F \vee F}$ . Notons de plus (avec une convention différente de celle de Magnus)

$$\begin{aligned} \text{Vecs} &= \tau_{F \vee F} \circ \text{Vec} \\ \text{Veca} &= \tau_{F \wedge F} \circ \text{Vec} \end{aligned}$$

Par suite si  $\underline{x} = \text{Vec}(X)$ ,  $X \in \mathcal{M}_{n,n}(\mathbb{R})$ , on vérifie que  $\text{Vecs}(X)$  (resp.  $\text{Veca}(X)$ ) est le vecteur de dimension  $(n^* + n)$  (resp.  $n^*$ ) formé en empilant la partie triangulaire inférieure de  $X$  diagonale comprise (resp. diagonale exclue). Cette décomposition de  $F \otimes F$  est complétée en considérant la décomposition de  $\mathcal{Y}_n$  en somme directe

$$\mathcal{Y}_n = \mathcal{Y}_H \oplus \mathcal{Y}_*$$

où  $\mathcal{Y}_H$  est le sous-espace de  $\mathcal{Y}_n$  de dimension  $n^*$ , formé par les matrices symétriques à diagonale nulle et  $\mathcal{Y}_*$  est le sous-espace de  $\mathcal{Y}_n$  de dimension  $n$ , formé par les matrices diagonales. Par suite si  $\mathbf{O}$  désigne la matrice d'ordre  $n$  identiquement nulle alors

$$\forall H \in \mathcal{Y}_H, \forall D \in \mathcal{Y}_* \quad H * D = \mathbf{O}$$

Dans  $F \vee F$ , on définit par isomorphisme les termes de la décomposition en somme directe

$$F \vee F = F_{VH}F \oplus F_{V*}F$$

où  $\dim(F_{VH}F) = n^*$  et  $\dim(F_{V*}F) = n$ . Il existe alors de façon immédiate un isomorphisme entre  $F_{VH}F$  et le sous-espace de  $(F \wedge F)^* \otimes (F \wedge F)^*$  décrivant les formes bilinéaires diagonales définies sur  $(F \wedge F) \times (F \wedge F)$ , c'est à dire telles que

$$B(f_i \wedge f_j, f_k \wedge f_l) = \begin{cases} 0 & \text{si } (i, j) \neq (k, l) \\ b_{ij} & \text{si } (i, j) = (k, l) \end{cases}$$

Nous montrons par la suite que cet isomorphisme joue un rôle central dans l'étude des méthodes de Codage Multidimensionnel.

Introduisons maintenant le sous-espace vectoriel de  $\mathcal{Y}_n$  défini par

$$\mathcal{Y}_\gamma = \{ W \in \mathcal{Y}_n ; W D_\gamma \underline{1} = \underline{0} \} = \{ W \in \mathcal{Y}_\gamma ; W \gamma = \underline{0} \}$$

Toute matrice  $W$  de  $\mathcal{Y}_\gamma$  qui est semi-définie positive s'interprète comme une matrice de produits scalaires associée à une configuration réelle  $Z$  centrée par rapport à la pondération  $\gamma$ . On parle alors de matrice de produits scalaires centrés. On vérifie de plus que  $\mathcal{Y}_\gamma$  est de dimension  $n^*$ .

Enfin, étant donné un plongement Euclidien de  $(I, \delta)$  de matrice de représentation  $Z$ , notons  $\Delta$  l'application

$$\begin{aligned} \Delta : \mathcal{M}_{n,n}(\mathbb{R}) &\rightarrow \mathcal{D}_E(I) \\ Z &\mapsto \Delta(Z) = (\delta_{ij}(Z))_{ij} \end{aligned}$$

où  $\delta_{ij}(Z) = \|z_i - z_j\|_M$  et  $\mathcal{D}_E(I)$  le sous-ensemble de  $\mathcal{Y}_H$  formé par les matrices de distance Euclidienne. Comme il a été vu plus haut,  $\Delta$  est invariante par translation de l'origine des axes et plus généralement, si  $X = {}^t \underline{m} \underline{1} + BZ$  alors

$$\delta_{ij}^2(X) = \|B(z_i - z_j)\|_M^2 = \|z_i - z_j\|_M^2 = \delta_{ij}^2(Z)$$

si et seulement si  $B$  est une matrice de rotation ou de  $M$ -symétrie (c'est à dire telle que  ${}^t B M B = M$ ).

On vérifie de même que si  $\dot{X} = X - {}^t \underline{m} \underline{1}$ , alors

$$W_X = {}^t \dot{X} M \dot{X} = {}^t Z {}^t B M B Z = {}^t B M B = W_Z$$

Alors l'ensemble des configurations normalisées, associées à un choix de  $\gamma$  et d'une matrice diagonale  $\Lambda = \text{diag}(\lambda_s)$ ,  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_n = 0$  par

$$\mathcal{X}(\gamma, \Lambda) = \{ X \in \mathcal{M}_{n,n}(\mathbb{R}); X \gamma = 0, X {}^t X = \Lambda \}$$

permet de construire la relation d'équivalence

$$\forall X, Z \in \mathcal{X}(\gamma, \Lambda) \quad (X \sim Z) \Leftrightarrow (\Delta(X) = \Delta(Z))$$

Notons  $\mathcal{X}(\gamma, \Lambda) / \sim$  l'espace quotient et  $[X]$  la classe d'équivalence de  $X$ .

Considérons l'application

$$\begin{aligned} \sigma_\gamma : \mathfrak{X}(\gamma, \Lambda) &\rightarrow \mathfrak{Y}_\gamma \\ X &\mapsto W_X \end{aligned}$$

On peut alors lui associer l'application bijective

$$\begin{aligned} [\sigma_\gamma] : \mathfrak{X}(\gamma, \Lambda)/\sim &\rightarrow \mathfrak{Y}_\gamma \\ [X] &\mapsto W_X = \sigma_\gamma(X) \end{aligned}$$

qui définit de façon unique la matrice de produits scalaires centrés (on dit aussi la matrice de connexion dans la littérature soviétique cf. MIRKIN (1979)). Cette application a pour inverse l'application

$$\begin{aligned} [\pi_\gamma] : \mathfrak{Y}_\gamma &\rightarrow \mathfrak{X}(\gamma, \Lambda)/\sim \\ W &\mapsto [X] \end{aligned}$$

où  $[X]$  est l'unique classe d'équivalence  $\mathfrak{X}(\gamma, \Lambda)/\sim$  telle que  $W_X = W$  pour tout  $X$  de  $[X]$ .

Les problèmes de plongement Euclidien nécessitent la définition du sous-ensemble  $\mathfrak{Y}_n^+$  de  $\mathfrak{Y}_n$  formé par les matrices symétriques semi-définies positives.  $\mathfrak{Y}_n^+$  définit un cône convexe fermé dans  $\mathfrak{Y}_n$  et FICHET (1983) montre que  $\mathfrak{Y}_n^+$  est d'intérieur non vide formé par les matrices symétriques définies positives, alors que sa frontière est décrite par les matrices symétriques semi-définies positives de rang inférieur ou égal à  $(n-1)$ .  $\mathfrak{Y}_n^+$  a pour cône polaire (ou dual) (cf. ROCKAFELLAR( 1971))

$$(\mathfrak{Y}_n^+)^* = \mathfrak{Y}_n^- = \{S \in \mathfrak{Y}_n ; S \text{ est semi-définie négative}\} = -\mathfrak{Y}_n^+$$

L'étude des conditions d'existence d'un plongement Euclidien de  $(I, \delta)$  présentée plus haut accorde une attention particulière aux opérateurs

$$\begin{aligned} \alpha_\gamma : \mathfrak{Y}_H &\rightarrow \mathfrak{Y}_\gamma \\ H &\mapsto \alpha_\gamma(H) = -\frac{1}{2} Q_\gamma H' Q_\gamma \end{aligned}$$

$$\begin{aligned} \beta_\gamma : \mathfrak{Y}_\gamma &\rightarrow \mathfrak{Y}_H \\ W &\mapsto \beta(W) = (W+I) \mathbb{1}_n \mathbb{1}_n + \mathbb{1}_n \mathbb{1}_n (W+I) - 2W \end{aligned}$$

Notons que si  $\alpha_\gamma$  dépend du choix d'une origine  $P_0$  dans  $E$ , ou de façon équivalente d'une pondération réelle  $\gamma$  de  $I$ ,  $\beta_\gamma$  traduit en terme algébrique la loi du cosinus.  $\beta_\gamma$  est donc la restriction à  $\mathfrak{Y}_\gamma$  de la fonction  $\beta$  qui elle ne dépend pas de  $\gamma$ . De plus

$$\text{Vecs}(\alpha_\gamma(H)) = -\frac{1}{2} (Q_\gamma \otimes Q_\gamma) \text{Vecs}(H)$$

et donc l'expression des applications  $\alpha_\gamma$  et  $\beta_\gamma$  dans  $F \otimes F$  montre que ce sont deux applications linéaires. Particularisons leur étude dans le cas où  $H$  est la matrice associée à une dissimilarité Euclidienne  $\delta(Z)$  et a pour terme général

$$h_{ij} = \delta_{ij}^2(Z)$$

Pour ce faire, introduisons l'ensemble

$$\mathcal{Y}_\gamma^+ = \mathcal{Y}_n^+ \cap \mathcal{Y}_\gamma = \{W \in \mathcal{Y}_n^+ ; W\gamma = \underline{0}\}$$

$\mathcal{Y}_\gamma^+$  est un cône convexe fermé dans  $\mathcal{Y}_n$  et tout élément  $W$  de  $\mathcal{Y}_\gamma^+$  de rang  $p$  est une matrice symétrique, semi-définie positive. Par suite, elle admet une décomposition spectrale fournissant  $p$  valeurs propres strictement positives et  $n-p$  valeurs propres nulles. L'une de ces dernières est associée au vecteur propre  $\gamma$  puisque  $W\gamma = \underline{0}$ .

Notons  $\underline{u}_s$  le vecteur propre de norme 1 associé à la valeur propre  $\lambda_s > 0$  ( $s = 1, \dots, p$ ) et choisissons pour représentant de  $[\pi_\gamma](W)$  dans  $\mathcal{X}(\gamma, \Lambda)$  la matrice de  $s$ -ième ligne

$$\underline{x}_s = \begin{cases} \sqrt{\lambda_s} \underline{u}_s & \text{si } s = 1, 2, \dots, p \\ \underline{0} & \text{sinon} \end{cases}$$

De façon complémentaire, notons  $\epsilon_\gamma$  la restriction de  $\Delta$  à  $\mathcal{X}(\gamma, \Lambda)$  et  $[\epsilon_\gamma]$  l'application induite sur  $\mathcal{X}(\gamma, \Lambda)/\sim$ . Alors  $[\epsilon_\gamma]$  définit une bijection entre  $\mathcal{X}(\gamma, \Lambda)/\sim$  et le cône  $\mathcal{D}_E(I)$  des dissimilarités Euclidiennes définies sur  $I \times I$ . Notons de plus  $[\epsilon_\gamma]^{-1}$  la bijection inverse. D'autre part, pour tout réel  $a$ , notons  $\xi_a$  l'opérateur  $a$ -ième puissance de Hadamard défini sur  $\mathcal{M}_{n,n}(\mathbb{R})$  par

$$B \in \mathcal{M}_{n,n}(\mathbb{R}) \mapsto \xi_a(B) = (b_{ij}^a)_{ij}$$

on note souvent  $B^{*a} = \xi_a(B)$ .

Puisque le contre-exemple dû à FICHET (1983) et exposé au paragraphe §II.1.3 montre que  $\mathcal{D}_E(I)$  n'est pas convexe dans  $\mathcal{Y}_H$ , on introduit le cône convexe fermé dans  $\mathcal{Y}_H$  décrit par les carrés de dissimilarités Euclidiennes. Notons le  $\mathcal{D}_E^{*2}(I)$ .

$$\mathcal{D}_E^{*2}(I) = \xi_2(\mathcal{D}_E(I)) \quad (\Leftrightarrow \mathcal{D}_E(I) = \xi_{1/2}(\mathcal{D}_E^{*2}(I))$$

et définissons l'opérateur

$$\begin{aligned} \alpha_\gamma^+ &: \mathcal{D}_E^{*2}(I) \rightarrow \mathcal{Y}_\gamma^+ \\ \alpha_\gamma^+ &= [\sigma_\gamma] \circ [\epsilon_\gamma]^{-1} \circ \xi_{1/2} \end{aligned}$$

$\alpha_\gamma^+$  est la restriction de  $\alpha_\gamma$  à  $\mathfrak{D}_E^{*2}(\mathbb{I})$  et sauf risque d'erreur, nous identifions ces deux opérateurs dans la suite. Nous agissons de même pour  $\beta_\gamma$  et sa restriction  $\beta_\gamma^+$  à  $\mathfrak{Y}_\gamma^+$

$$\beta_\gamma^+ : \mathfrak{Y}_\gamma^+ \rightarrow \mathfrak{D}_E^{*2}(\mathbb{I})$$

$$\beta_\gamma^+ = \xi_2 \circ [\varepsilon_\gamma] \circ [\pi_\gamma]$$

**Remarque :** On a clairement pour toute dissimilarité Euclidienne  $\delta(Z)$

$$[Z] = [\varepsilon_\gamma]^{-1}(\Delta(Z))$$

où

$$[\varepsilon_\gamma]^{-1} = [\pi_\gamma] \circ \alpha_\gamma^+ \circ \xi_2$$

De façon analogue  $\beta_\gamma^+$  intervient dans la définition de l'application associant à  $W_Z$  la classe d'équivalence  $[Z]$  et donc les applications  $\alpha_\gamma^+$  et  $\beta_\gamma^+$  jouent un rôle central dans la formulation du problème du plongement Euclidien d'une dissimilarité  $\delta$ .

CRITCHLEY (1982a) montre les deux résultats suivants dont la vérification par substitution est immédiate :

**Lemme 3 :** Si  $H \in \mathfrak{Y}_H \cap \mathfrak{Y}_\gamma$ , alors  $\alpha_\gamma(H) = -\frac{1}{2}H$  et  $\beta_\gamma(H) = -2H$

En fait CRITCHLEY suppose toujours que  $\gamma = \frac{1}{n} \mathbb{1}$ . Mais le cas général reçoit une démonstration analogue mutatis mutandis. Bien entendu ce résultat n'a d'intérêt que pour  $n > 3$  car alors

$$\dim(\mathfrak{Y}_H \cap \mathfrak{Y}_\gamma) = n^* - n.$$

Dans le cas contraire cet ensemble est réduit à la matrice nulle  $\mathbf{O}$  et la conclusion du lemme 3 est trivialement vraie.

**Lemme 4 :** Si  $\gamma = \frac{1}{n} \mathbb{1}$ , alors

$$\text{i) } \alpha_1(\mathbb{1}^t \mathbb{1} - \mathbb{I}) = \frac{1}{2} Q_1$$

$$\text{ii) } \beta_1(Q_1) = 2(\mathbb{1}^t \mathbb{1} - \mathbb{I})$$

Les matrices  $Q_1$  et  $(\mathbb{1}^t \mathbb{1} - \mathbb{I})$  jouent un rôle privilégié dans l'étude de la correspondance entre  $\mathfrak{D}_E^{*2}$  et  $\mathfrak{Y}_1^+$ . Plus généralement si  $\gamma$  est une pondération quelconque, un rôle analogue est accordé aux matrices  $W_\gamma = D_\gamma^{-1} - \mathbb{1}^t \mathbb{1}$  et  $H = Q_\gamma D_\gamma = D_\gamma^t Q_\gamma$ . En effet, la

fonction de dissimilarité  $\delta(\gamma)$  induite par la pondération  $\gamma$  (supposée telle que  $\gamma_i \neq 0$  pour tout  $i$  de  $[n]$ ) et définie par

$$\delta_{ij}^2(\gamma) = \frac{1}{\gamma_i} + \frac{1}{\gamma_j}, \quad \forall (i,j) \in [n] \times [n] \quad i \neq j$$

vérifie la propriété suivante où nous posons  $H(\gamma) = \Delta(\gamma) \times \Delta(\gamma)$

**Propriété (FICHET (1983)) :** Soit  $\gamma$  un élément de  $F^*$  tel que  $\sum \gamma_i = 1$ , alors la matrice  $W_\gamma = \alpha_\gamma^+ (H(\gamma))$  s'écrit dans la base  $\mathcal{F}$  de  $F$

$$W_\gamma = D_\gamma^{-1} - \frac{1}{n} \mathbf{1} \mathbf{1}$$

**Preuve :**

$$\delta_{ij}^2(\gamma) = \sum_{j \neq i} \gamma_j \delta_{ij}^2(\gamma) = \sum_{j \neq i} \gamma_j \left( \frac{1}{\gamma_i} + \frac{1}{\gamma_j} \right) = \frac{1}{\gamma_i} \sum_{j \neq i} \gamma_j + (n-1) = \frac{1}{\gamma_i} + (n-2)$$

$$\delta_{ii}^2(\gamma) = \sum_i \gamma_i \left( \frac{1}{\gamma_i} + (n-2) \right) = 2(n-1)$$

et donc si  $i \neq j$

$$w_{ij}^\gamma = -\frac{1}{2} \left[ \frac{1}{\gamma_i} + \frac{1}{\gamma_j} - \frac{1}{\gamma_i} - (n-2) - \frac{1}{\gamma_j} - (n-2) + 2(n-1) \right] = -1$$

et

$$w_{ii}^\gamma = \delta_{ii}^2(\gamma) - \delta_{ii}^2(\gamma) = \frac{1}{\gamma_i} - 1 \quad \square$$

De façon complémentaire, les éléments de  $I$  sont représentés dans  $F^*$  par les points  $N_i$  extrémités des vecteurs de base  $f_i^*$  de  $F^*$ . Par suite on vérifie le résultat suivant

**Propriété :** Soit  $\gamma$  un élément de  $F^*$  tel que  $\sum \gamma_i = 1$ , alors la matrice de dispersion  $N_\gamma$  du nuage de points  $\{(N_i, \gamma_i) ; i \in [n]\}$  de  $F^*$  s'écrit dans la base  $\mathcal{F}^*$  de  $F$

$$N_\gamma = D_\gamma Q_\gamma = {}^t Q_\gamma D_\gamma = D_\gamma - D_\gamma \frac{1}{n} \mathbf{1} \mathbf{1} D_\gamma$$

**Preuve :** La  $i$ -ème coordonnée de  $N_i$  dans  $F^*$  est

$$(f_i^*)^j = \delta_i^j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$$

Par suite le barycentre  $G_i$  du nuage de points massiques  $\{(N_i, \gamma_i) ; i \in [n]\}$  a pour  $j$ -ème coordonnée  $\gamma_j$  et par suite pour tout couple  $(i,k)$  d'éléments de  $[n]$

$$(N_\gamma)_{ik} = \langle \overrightarrow{GN_i}, \overrightarrow{GN_k} \rangle = \sum_j \gamma_j (f_i^*)^j (f_i^*)^k = \gamma_i \delta_i^k - \gamma_i \gamma_k = \gamma_i [\delta_i^k - \gamma_k]$$

c'est dire que

$$N_\gamma = D_\gamma Q_\gamma = {}^t Q_\gamma D_\gamma \quad \parallel$$

Ces deux propriétés montrent d'une part que  $N_\gamma$  et  $W_\gamma$  ne sont pas inverses l'une de l'autre si  $\gamma$  est différent de  $\frac{1}{n} \mathbf{1}$ . Cependant un calcul élémentaire montre que

**Proposition 10:**  $W_\gamma N_\gamma = Q_\gamma$

$N_\gamma$  et  $W_\gamma$  sont inverses l'une de l'autre dans le sous-espace vectoriel  $F_\gamma$  de  $F$  complémentaire  $D_\gamma$ -orthogonal de la droite des constantes  $\Delta_{1n}$ . Ce résultat repose sur le choix d'une fonction de dissimilarité très particulière  $\delta(\gamma)$ . Considérons le cas général

**Proposition 11:** Soit  $\gamma$  un élément de  $F^*$  tel que  ${}^t \gamma \mathbf{1} = 1$  et  $H$  la matrice de  $\mathcal{Y}_H$  associée dans  $\mathcal{F}$  à la forme quadratique  $h$  vérifiant

$$h \langle \overrightarrow{N_i N_j}, \overrightarrow{N_i N_j} \rangle = h_{ij}$$

alors l'opérateur principal associé au nuage de points massiques  $\{(N_i, \gamma_i) ; i \in [n]\}$  de  $F^*$  dans les bases  $\mathcal{F}$  et  $\mathcal{F}^*$  vérifie

$$U_\gamma = H N_\gamma = \widetilde{W}_H D_\gamma$$

où  $\widetilde{W}_H$  est la matrice non symétrique

$$\widetilde{W}_H = H - D_H \mathbf{1} \mathbf{1}^t \quad \text{où } D_H = \text{diag}(h_{i+}) \text{ et } h_{i+} = \sum_{j \neq i} h_{ij} \gamma_j$$

**Preuve :** L'opérateur principal  $U_\gamma$  a pour représentation  $HV$  où  $V$  est la matrice de dispersion du nuage. Mais nous venons de voir qu'ici  $V = N_\gamma$  et donc

$$HV = H [ D_\gamma - D_\gamma \mathbf{1} \mathbf{1}^t D_\gamma ] = H D_\gamma - H D_\gamma \mathbf{1} \mathbf{1}^t D_\gamma$$

$$H D_\gamma = (h_{ij} \gamma_j)_{ij}$$

$$H D_\gamma \mathbf{1} \mathbf{1}^t = \left( \sum_k h_{ik} \gamma_k \gamma_j \right)_{ij}$$

Posons

$$h_{i.} = \sum_k h_{ik} \gamma_k$$

Puisque  $h_{ii} = 0$ , alors  $h_{i.} = h_{i+}$ . Par suite  $HV$  a pour terme général

$$(HV)_{ij} = h_{ij} \gamma_j - h_{i.} \gamma_j = (\widetilde{W}_H D_\gamma)_{ij}$$

$$\widetilde{W}_H = H - \underline{h} \mathbf{1} \mathbf{1}^t \quad \parallel$$

Les opérateurs  $\alpha_1$  et  $\beta_1$  sont inverses l'un de l'autre. Il est cependant important de remarquer que ce résultat ne s'étend pas au cas général. En effet si  $W = {}^t Q_\gamma$ , alors  $Q_\gamma \gamma = \underline{0}$  et donc puisque  $Q_\gamma = I - \underline{1} \underline{1} D_\gamma$

$$\beta_\gamma ({}^t Q_\gamma) = ({}^t Q_\gamma * I) \underline{1} \underline{1} + \underline{1} \underline{1} ({}^t Q_\gamma * I) - 2 {}^t Q_\gamma$$

et puisque  ${}^t Q_\gamma * I = I - D_\gamma$ , on obtient après simplification

$$\beta_\gamma ({}^t Q_\gamma) = 2 (\underline{1} \underline{1} - I) + \underline{1} \underline{1} D_\gamma - D_\gamma \underline{1} \underline{1}$$

Appliquons  $\alpha_\gamma$  à cette matrice, on obtient puisque  $Q_\gamma \underline{1} = \underline{0}$

$$\alpha_\gamma (\beta_\gamma ({}^t Q_\gamma)) = -2 Q_\gamma {}^t Q_\gamma$$

Le deuxième membre de cette égalité est égal à  $-2 Q_\gamma$  si et seulement si  $\gamma = \frac{1}{n} \underline{1}$ . On vérifie dans ce cas particulier que  $\alpha_1$  et  $\beta_1$  sont inverses l'un de l'autre.

D'autre part si  $H$  appartient à  $\mathfrak{D}_E^{*2}(I)$ , notons  $W = \alpha_\gamma(H)$  et  $P_\gamma = \underline{1} \underline{1} D_\gamma$ . On vérifie, dans ce cas le résultat suivant dans l'énoncé duquel on pose  $h_{..} = \sum_i \sum_j \gamma_i \gamma_j h_{ij}$

**Lemme 5:**

$$i) \frac{1}{2} {}^t \gamma H \gamma = \text{Trace}(W D_\gamma) = h_{..}$$

$$ii) \frac{1}{2} D_\gamma H D_\gamma = \text{Trace}(W D_\gamma) \underline{1} \underline{1}$$

**Preuve :**  $H = (W * I) \underline{1} \underline{1} + \underline{1} \underline{1} (W * I) - 2 W$

et puisque  $D_\gamma$  est diagonale et  $W \gamma = \underline{0}$ .

$$H \gamma = (W * I) \underline{1} + \underline{1} \underline{1} (W * I) D_\gamma \underline{1}$$

$$\begin{aligned} {}^t \gamma H \gamma &= {}^t \underline{1} D_\gamma (W * I) \underline{1} + {}^t \underline{1} (W * I) D_\gamma \underline{1} \\ &= {}^t \underline{1} (D_\gamma W * I) \underline{1} + {}^t \underline{1} (W D_\gamma * I) \underline{1} \\ &= 2 {}^t \underline{1} (W D_\gamma * I) \underline{1} = 2 \text{Trace}(W D_\gamma) \end{aligned}$$

Le point ii) résulte alors de l'égalité  $P_\gamma = \underline{1} \underline{1} \gamma$   $\square$

Enfin notons  $\underline{w}$  le vecteur formé par les termes diagonaux de  $W$  et  $\underline{h}$  le vecteur des sommes pondérées par  $\gamma$  des lignes de  $H$  :

$$\underline{w} = (W * I) \underline{1}$$

$$\underline{h} = H \gamma$$

On vérifie alors les résultats suivants traités dans le cas particulier  $\gamma = \frac{1}{n} \underline{1}$  par CRITCHLEY (1982).

**Lemme 6:**

$$i) \text{ Si } W = \alpha_\gamma(H), \text{ alors } \underline{w} = (I - \frac{1}{2} P_\gamma) \underline{h}$$

$$ii) \text{ Si } H = \beta_\gamma(W), \text{ alors } \underline{h} = (I + P_\gamma) \underline{w}$$

**Preuve :**

$$\begin{aligned} W = \alpha_\gamma(H) &= -\frac{1}{2} [H - \underline{1}' \gamma H - H \gamma \underline{1}' + \underline{1}' \gamma H \gamma \underline{1}'] \\ &= -\frac{1}{2} [H - \underline{1}' \underline{h} - \underline{h}' \underline{1} + P_\gamma \underline{h}' \underline{1}] \end{aligned}$$

et donc

$$\underline{w} = (W * I) \underline{1} = -\frac{1}{2} [H * I - \underline{1}' \underline{h} * I - \underline{h}' \underline{1} * I + P_\gamma \underline{h}' \underline{1} * I] \underline{1}$$

Par suite, puisque  $H * I = 0$ ,  $\underline{1}' \underline{h} * I = D_h$  et  $P_\gamma \underline{1}' \underline{h} * I = D_{P_\gamma h}$  on vérifie

$$\underline{w} = D_h \underline{1} - \frac{1}{2} D_{P_\gamma h} \underline{1} = \underline{h} - \frac{1}{2} P_\gamma \underline{h}$$

Le point ii) résulte du fait que  $\underline{1}' \gamma = 1$ . En effet :

$$H = \beta_\gamma(W) = \underline{w}' \underline{1} + \underline{1}' \underline{w} - 2 \underline{w}$$

$$H \gamma = \underline{w}' \underline{1} \gamma + \underline{1}' \underline{w} \gamma - 2 \underline{w} \gamma$$

et donc puisque  $W \gamma = 0$

$$H \gamma = \underline{w} + \underline{1}' \underline{w} \gamma - 2 \gamma \underline{w} = \underline{w} + P_\gamma \underline{w} \quad \parallel$$

On en déduit que  $\underline{w}$  est nul si  $\underline{h}$  l'est et réciproquement. Par suite  $H$  est un élément de  $\mathcal{Y}_H \cap \mathcal{Y}_\gamma$  si et seulement si il en est de même pour  $W$ . Mais  $\underline{h}$  et  $\gamma$  appartiennent respectivement à l'image de  $H$ ,  $\text{Im}(H)$  et à  $\text{Ker}(W)$  et

$$(H \in \mathcal{Y}_H \cap \mathcal{Y}_\gamma) \Leftrightarrow (\underline{h} = 0)$$

$$(W \in \mathcal{Y}_H \cap \mathcal{Y}_\gamma) \Leftrightarrow (\underline{w} = 0)$$

on en déduit que si  $\underline{h}$  est un vecteur multiple de  $\underline{1}_n$ , alors il existe  $h \in \mathbb{R}$  tel que

$$\underline{h} = h \underline{1}_n \Rightarrow \underline{w} = h \underline{1}_n - \frac{h}{2} \underline{1}_n \gamma \underline{1}_n = \frac{h}{2} \underline{1}_n$$

et donc  $\underline{w}$  est colinéaire à  $\underline{1}_n$ . De même s'il existe  $w$  dans  $\mathbb{R}$  tel que  $\underline{w} = w \underline{1}_n$  alors  $\underline{h} = 2w \underline{1}_n$ . Par suite,  $\underline{h}$  et  $\underline{1}_n$  sont colinéaires si et seulement si  $\underline{w}$  et  $\underline{1}_n$  sont colinéaires.

Nous analysons dans les paragraphes suivants les relations algébriques nécessaires à l'élaboration d'un formalisme unificateur permettant de décrire d'une part les méthodes d'Analyse en Coordonnées Principales, formulées usuellement dans  $F$  et d'autre part, les méthodes de Codage Multidimensionnel, qui reçoivent ici une formulation dans  $F \wedge F$ , qui nous semble canonique. Les outils mis en place à cet effet sont complétés par une écriture matricielle des opérateurs associés à  $\alpha$  et  $\beta$  dans  $F \otimes F$ . Le regroupement de ces deux problèmes distincts nous semble intéressant puisqu'il montre la complémentarité des outils mis en œuvre.

#### IV.5. REPRÉSENTATION MATRICIELLE DE $\alpha$ ET $\beta$ DANS $F \otimes F$

Nous interprétons ici dans un cadre algébrique les calculs faits par CRITCHLEY (1982). Considérons l'application  $r$  définie sur  $(F \vee F)^*$ , à valeurs dans  $F^*$ , telle que pour tout vecteur  $f_i^* \vee f_j^*$  de la base  $\mathcal{F}^* \vee \mathcal{F}^*$  de  $(F \vee F)^*$

$$r(f_i^* \vee f_j^*) = f_i^* + f_j^*$$

Alors la matrice associée à  $r$  dans les bases  $\mathcal{F}^* \vee \mathcal{F}^*$  et  $\mathcal{F}^*$  a pour transposée

$${}^tR = \begin{pmatrix} \tilde{S}_1 \\ O_2 \tilde{S}_2 \\ \vdots \\ O_{n-1} \tilde{S}_{n-1} \end{pmatrix} \quad \text{où } \tilde{S}_i = (\underline{1}_{n-i} \mid I_{n-i}) \quad i = 1, \dots, n-1$$

$\underline{1}_{n-i}$  est le vecteur constant de  $\mathbb{R}^{n-i}$  de terme général 1,  $I_{n-i}$  la matrice identité de rang  $(n-i)$  et  $O_{n-i}$  la matrice nulle de  $\mathcal{M}_{n-i, i-1}(\mathbb{R})$ . Alors  ${}^tR$  est d'ordre  $n \times n$  et vérifie l'égalité

$${}^tR \underline{x} = \underline{y}$$

où  $\underline{y}$  donne les coordonnées du vecteur de  $F \vee F$  associé à la matrice carré d'ordre  $n$   $Y = \underline{x} \underline{1} + \underline{1} \underline{x}$  puisqu'alors

$$y_{ij} = x_i + x_j \quad i \leq j \quad i, j \in [n] \quad (38)$$

**Remarque :** Si  $F$  est décomposée en somme directe induite par la droite des constantes  $\Delta_1$

$$F = F_1 \oplus \Delta_1$$

alors

$$F \otimes F = (\Delta_1 \otimes \Delta_1) \oplus [(F_1 \otimes \Delta_1) \oplus (\Delta_1 \otimes F_1)] \oplus (F_1 \otimes F_1) \quad (39)$$

et donc si nous posons  $G_1 = (F_1 \otimes \Delta_1) \oplus (\Delta_1 \otimes F_1)$ , on vérifie que  ${}^tR(F_1)$  est un sous-espace propre de  $G_1$ . L'inclusion de  $G_1$  dans  ${}^tR(F_1)$  n'est pas assurée à cause de la contrainte de symétrie imposée par l'expression  $\underline{y} = \text{Vecs}(Y)$  où  $Y = \underline{x}{}^t\underline{1} + \underline{1}{}^t\underline{x}$ . Formellement cela s'exprime sous la forme

$${}^tR(F_1) = G_1 \cap (F \vee F) \quad (40)$$

Notons alors

$$F_1 \vee F_1 = (F_1 \otimes F_1) \cap (F \vee F)$$

et  $\Delta_1 \vee \Delta_1$  la droite des constantes dans  $F \vee F$

$$\Delta_1 \vee \Delta_1 = (\Delta_1 \otimes \Delta_1) \cap (F \vee F)$$

de telle sorte que (39) induit sur  $F \vee F$  la décomposition en somme directe

$$F \vee F = \Delta_1 \vee \Delta_1 \oplus {}^tR(F_1) \oplus F_1 \vee F_1 \quad (41)$$

Si  $Q$  est un élément de  $\mathcal{Y}_n$  de terme général  $q_{ij}$ , nous notons alors  $q_i$  la  $i$ -ème composante de  $RQ$  où  $\underline{q} = \text{Vecs}(Q)$

$$q_{i+} = \sum_{j \neq i} q_{ij} = \sum_{j=1}^n q_{ij} - q_{ii} = q_{i.} - q_{ii}$$

$$q_{.+} = \sum_{i=1}^n q_{i+} = \frac{1}{2} ({}^t\underline{1}_n \underline{q}) = \frac{1}{2} {}^t\underline{1}_n Q \underline{1}_n$$

Alors  $q_i = q_{i+}$  si  $Q \in \mathcal{Y}_H$  et  $q_i = -q_{ii}$  si  $Q \in \mathcal{Y}_1$ .

De plus si  $\underline{h} = \text{Vecs}(H)$ , pour tout élément  $H$  de  $\mathcal{Y}_n$ , alors  $\underline{h} = Rh$ . La correspondance entre  $F$  et  $F \vee F$  est résumée par le schéma de dualité

$$\begin{array}{ccc}
 F^* & \xleftarrow{\quad r \quad} & (F \vee F)^* \\
 \uparrow N & & \uparrow \Theta \\
 F & \xrightarrow{\quad {}^t r \quad} & F \vee F
 \end{array}$$

**Remarque :** L'interprétation de l'équation (38) dans le cadre de l'Analyse de la Variance appliquée au tableau croisé  $Y$  de terme général  $y_{ij}$  permet d'appeler  $R$  la matrice d'additivité simple

L'analyse algébrique du schéma de dualité précédent est partiellement éclairante lorsqu'on fixe  $\gamma = \frac{1}{n} \mathbf{1}$ . Le changement d'origine associé à un choix différent de vecteur de centrage  $\gamma$  correspond en effet au choix d'une matrice  $N = D_\gamma$  distincte de  $\frac{1}{n} I_n$ . Nous reprenons ici les résultats de CRITCHLEY (1982) que nous nous proposons de comparer à l'analyse du triple  $(I, \delta, Q_{|x|})$  présentée plus bas.

Notons  $P_{n^*}$  le projecteur orthogonal sur la droite des constantes de  $F \otimes F$  engendrée par  $\mathbf{1}_{n^*} \otimes \mathbf{1}_{n^*}$

$$P_{n^*} = \frac{1}{n^*} \mathbf{1}_{n^*} \mathbf{1}_{n^*}'$$

Alors CRITCHLEY (1982) montre le résultat suivant où  $I_{n^*}$  désigne la matrice identité d'ordre  $n^*$

**Théorème 1:**

i)  $\alpha_1$  a pour matrice représentative dans  $F \vee F$  la matrice

$$A = -\frac{1}{2} (I_{n^*} - \frac{1}{n} {}^tRR + \frac{n-1}{n} P_{n^*})$$

ii)  $\beta_1$  a pour matrice représentative dans  $F \vee F$  la matrice

$$B = - (2 I_{n^*} + {}^tRR)$$

En particulier  $A$  et  $B$  sont symétriques, définies, négatives et mutuellement inverses.

La matrice  $K = {}^tRR$  joue un rôle majeur dans la détermination des propriétés de  $A$  et de  $B$ . Son étude utilise les décompositions en somme directe suivantes :

$$\begin{aligned} F &= F_1 \oplus \Delta_1 \\ F \vee F &= (F \vee F)_1 \oplus \Delta_{1 \otimes 1} \end{aligned}$$

Alors CRITCHLEY (1982a) présente les résultats suivants, exprimés ici mutatis mutandis, en distinguant les espaces vectoriels et leur dual algébrique.

**Théorème 2:**

- i)  ${}^tR \mathbf{1}_n = 2 \mathbf{1}_{n^*}$ ,  ${}^tR$  applique  $F_1$  sur  $(F \vee F)_1$  et  $({}^tR \underline{x} \in (F \vee F)_1) \Leftrightarrow (\underline{x} \in F_1)$
- ii)  $R \mathbf{1}_{n^*} = (n-1) \mathbf{1}_n$ ,  $R$  applique  $(F \vee F)_1^*$  sur  $F_1^*$  et  $(R \underline{y} \in F_1^*) \Leftrightarrow (\underline{y} \in (F \vee F)_1^*)$
- iii)  $N = R {}^tR = (n-2) I_n + n P_1 = 2(n-1) P_1 + (n-2) Q_1$
- iv)  $\text{Rang}(K) = \text{rang}(N) = \text{rang}(R) = r$
- v)  $\text{Ker}(K) = \text{Ker}(R) = \{\text{Vecs}(S) ; S \in \mathcal{Y}_H \cap \mathcal{Y}_1\}$ ,  $\dim(\text{Ker}(K)) = n^* - n$
- vi)  $K = {}^tRR = 2(n-1) P_{n^*} + (n-2) (\frac{1}{n-2} {}^tR Q_1 R)$
- vii)  $\text{Im}(K) = \text{Im}({}^tR) = \{\text{Vecs}(S) ; S = \underline{x} \mathbf{1} + \mathbf{1} \underline{x}' \text{ pour un } \underline{x} \in F\}$

La décomposition spectrale de  $N$  précisée par l'énoncé iii) montre que  $2(n-1)$  est une valeur propre simple de  $N$ , associée au vecteur propre  $\mathbf{x} = \frac{1}{\sqrt{n}} \mathbf{1}_n$  et que  $(n-2)$  est une valeur propre de  $N$  de multiplicité  $(n-1) = \text{rang}(Q_1)$ , associée à toute base orthonormée  $\{\mathbf{u}_s, s = 1, \dots, (n-1)\}$  de  $F_1$ .

De même, la décomposition spectrale de  $K$  donnée dans l'énoncé vi) montre que  $2(n-1)$  est une valeur propre simple de  $K$ , associée au vecteur propre  $\mathbf{y} = \frac{1}{\sqrt{n^*}} \mathbf{1}_{n^*}$  et que  $(n-2)$  est valeur propre de  $K$  de multiplicité  $(n-1) = \text{rang}(Q_1)$ , associée en particulier aux vecteurs  $\mathbf{v}_s = \frac{1}{n-2} {}^t R \mathbf{u}_s$ . De plus  $0$  est une valeur propre de  $K$  de multiplicité  $(n^* - n)$  si  $n > 3$ , associée à toute base orthonormée de  $\text{Ker}(K)$ .

On vérifie de plus les équivalences suivantes, par application des lemmes 3 et 4

$$\begin{aligned} \mathbf{1}_{n^*} = \text{Vecs}(H) &\Leftrightarrow H = \mathbf{1} {}^t \mathbf{1} - I \\ \mathbf{1}_{n^*} = \text{Vecs}(W) &\Leftrightarrow W = -n Q_1 \end{aligned}$$

La décomposition spectrale de  $K$  respecte la restriction à  $F \vee F$  de la décomposition additive (41) de  $F \vee F$ . Notons  $E_\lambda(K)$  le sous-espace propre de  $K$  associé à la valeur propre  $\lambda$  de  $K$ . Alors cette décomposition en somme directe de  $F \vee F$  nous permet de donner une interprétation géométrique au corollaire suivant démontré par CRITCHLEY (1982)

**Corollaire :** Si  $n \geq 3$ , tout élément  $\mathbf{q}$  de  $F \vee F$  s'écrit de façon unique

$$\mathbf{q} = \mathbf{q}_{[1]} + \mathbf{q}_{[2]} + \mathbf{q}_{[3]}$$

où

$$\mathbf{q}_{[1]} \in E_{2(n-1)}(K) \equiv \Delta_{1n^*} : \mathbf{q}_{[1]} = P_{1n^*} \mathbf{q} = \frac{\mathbf{q} \cdot \mathbf{1}_{n^*}}{2n^*} \mathbf{1}_{n^*}$$

$$\mathbf{q}_{[2]} \in E_{(n-2)}(K) \equiv {}^t R(F_1) : \mathbf{q}_{[2]} \text{ a pour coordonnée } (i,j)$$

$$\frac{1}{n-2} \left( (q_i - \frac{1}{n} q_{\cdot}) + (q_j - \frac{1}{n} q_{\cdot}) \right)$$

$$\mathbf{q}_{[3]} \in E_0(K). E_0(K) = \{0\} \text{ si } n=3. \text{ Si } n>3, \mathbf{q}_{[3]} \text{ a pour coordonnée } (i,j)$$

$$q_{ij} - \frac{1}{n-2} (q_i + q_j) + \frac{1}{(n-1)(n-2)} q_{\cdot}$$

Ce résultat retrouve donc la décomposition classique d'un tableau à double entrée utilisée en Analyse de la Variance.  $\mathbf{q}_{[3]}$  est le terme d'interaction. La forme particulière de cette décomposition résulte de la contrainte de symétrie du tableau  $\Delta$ . Elle montre que le formalisme mis en place à propos de l'analyse de données de dissimilarité, présente un intérêt plus général. Il est en particulier utile pour rendre compte des méthodes d'Analyse

des tables de contingence symétriques, dans le cadre du modèle Log-linéaire et dans celui de l'Analyse des Correspondances symétriques.

Une application directe de ces résultats est fournie par la décomposition spectrale des matrices A et B (cf. CRITCHLEY (1982)) :

**Corollaire :** Si  $n \geq 3$ , on a

$$i) \forall H \in \mathcal{Y}_H, A\bar{h} = - \left[ \frac{1}{2n} h_{[1]} + \frac{1}{n} h_{[2]} + \frac{1}{2} h_{[3]} \right]$$

$$ii) \forall W \in \mathcal{Y}_1, B\underline{w} = - \left[ 2n \underline{w}_{[1]} + n \underline{w}_{[2]} + 2 \underline{w}_{[3]} \right]$$

De plus si  $n \geq 3$  et si H est un élément de  $\mathcal{Y}_H$ , notons  $H_{[k]}$  l'élément de  $\mathcal{Y}_H$  tel que  $\text{Vecs}(H_{[k]}) = h_{[k]}$ . Lorsqu'on munit  $\mathcal{Y}_H$  de la structure Euclidienne associée à la norme  $\| \cdot \|$  ( $= \| \cdot \|_{1,1}$ ) la décomposition additive de  $F \vee F$  se traduit par

$$\| H \| ^2 = \sum_{k=1}^3 \| H_{[k]} \|^2$$

et en particulier

$$\begin{aligned} \sum_{i < j} h_{ij}^2 &= n \cdot [\bar{h}]^2 + \frac{1}{n-2} \sum_{i=1}^n \left( h_{i \cdot} - \frac{1}{n} h_{\cdot i} \right)^2 \\ &\quad + \sum_{i < j} \left( h_{ij} - \frac{1}{n-2} (h_{i \cdot} + h_{\cdot j}) - \frac{1}{(n-1)(n-2)} h_{\cdot i} \right)^2 \end{aligned}$$

où

$$\bar{h} = \frac{1}{n} \sum_{i < j} h_{ij}$$

Un résultat analogue s'obtient pour tout élément W de  $\mathcal{Y}_1$ .

L'étude des ensembles mis en œuvre pour l'analyse des espaces quasi-métriques finis  $(I, \delta)$  est complétée par le résultat suivant dû à CRITCHLEY (1982a). Nous avons vu plus haut que  $\mathcal{Y}_1 \cap \mathcal{Y}_H = \{0\}$  si  $n \leq 3$ . Considérons le cas  $n > 3$  et considérons les décompositions en sommes directes orthogonales (pour une norme Euclidienne quelconque):

$$\mathcal{Y}_H = \mathcal{Y}_1 \cap \mathcal{Y}_H \oplus \mathcal{R}$$

$$\mathcal{Y}_1 = \mathcal{Y}_1 \cap \mathcal{Y}_H \oplus \mathcal{T}$$

Alors on vérifie le résultat suivant :

**Proposition 12 (CRITCHLEY 1982) :** Si  $n > 3$  et si  $D_x$  est la matrice diagonale associée au vecteur  $x$  de  $F$ , alors

$$\begin{aligned} \mathcal{R} &= \{ H \in \mathcal{Y}_H ; \exists x \in F \text{ tel que } H = (x^t \mathbb{1}_n + \mathbb{1}_n^t x) - 2 D_x \} \\ \mathcal{T} &= \{ W \in \mathcal{Y}_H ; \exists x \in F \text{ tel que } W = (x^t \mathbb{1}_n + \mathbb{1}_n^t x) - n D_x - (\mathbb{1}_n^t x) I_n \} \end{aligned}$$

#### IV.6. CARACTÉRISATION DES OPÉRATEURS ADJOINTS DE $\alpha_1$ ET $\beta_1$

Considérons  $F \otimes F$  muni de sa structure Euclidienne induite par celle de  $F$ . Si  $F = (\mathbb{R}^n, N)$  alors  $F \otimes F = (\mathbb{R}^n \otimes \mathbb{R}^n, N \otimes N)$ . On munit donc  $\mathcal{M}_{n,n}(\mathbb{R})$  de la structure Euclidienne  $\langle \cdot, \cdot \rangle_{N, N}$ , par isomorphisme de structure Euclidienne. Alors nous avons vu précédemment que tout changement d'origine  $P_0$  de l'espace affine Euclidien de représentation se traduit dans  $F$  par le choix d'une matrice  $N = D_\gamma$ . Si  $\gamma$  est un vecteur à coordonnées positives,  $D_\gamma$  induit une structure Euclidienne sur  $F$ . Nous supposons donc cette condition réalisée dans la suite. De plus, dans un but de simplification des notations, nous ne précisons pas la métrique  $N$  et notons  $\langle \cdot, \cdot \rangle$  le produit scalaire ambiant. Alors l'adjointe de  $\alpha_1$  est définie par

$$\begin{aligned} \text{i) } \alpha_1^* &: \mathcal{Y}_1 \rightarrow \mathcal{Y}_H \\ \text{ii) } \forall H \in \mathcal{Y}_H, \forall W \in \mathcal{Y}_1 & \langle H, \alpha_1^*(W) \rangle = \langle \alpha_1(H), W \rangle \end{aligned} \quad (42)$$

De même l'adjointe de  $\beta_1$  est définie par

$$\begin{aligned} \text{i) } \beta_1^* &: \mathcal{Y}_H \rightarrow \mathcal{Y}_1 \\ \text{ii) } \forall H \in \mathcal{Y}_H, \forall W \in \mathcal{Y}_1 & \langle H, \beta_1^*(W) \rangle = \langle \beta_1(H), W \rangle \end{aligned} \quad (43)$$

Les propriétés mathématiques des adjointes sont classiques et nous ne les rappelons pas ici. Le lecteur intéressé peut consulter LUENBERGER (1969), ROCKAFELLAR (1971) pour leur utilisation en programmation convexe et CAILLEZ (1984) pour leur utilisation en Analyse des Données. Puisque  $\mathcal{D}_E^{*2}(I)$  (resp.  $\mathcal{Y}_1^+$ ) est un cône convexe de  $\mathcal{Y}_H$  (resp.  $\mathcal{Y}_1$ ) notons  $[\mathcal{D}_E^{*2}(I)]^*$  et  $[\mathcal{Y}_1^+]^*$  leur cône polaire (ou dual). Le résultat suivant est classique et montre l'existence d'une bijection entre ces deux cônes fermés :

**Lemme 7 :**

- i)  $[\mathcal{D}_E^{*2}(I)]^*$  est l'image réciproque de  $[\mathcal{Y}_1^+]^*$  par  $\beta_1^*$
- ii)  $[\mathcal{Y}_1^+]^*$  est l'image réciproque de  $[\mathcal{D}_E^{*2}(I)]^*$  par  $\alpha_1^*$

**Preuve :** Elle est une conséquence immédiate du fait que  $\alpha_1^*$  et  $\beta_1^*$  sont inversibles et mutuellement inverses puisque  $\alpha_1$  et  $\beta_1$  le sont.  $\square$

Notons d'autre part  $\alpha_1^*$  (resp.  $\beta_1^*$ ) l'opérateur décrivant  $\alpha_1^*$  (resp.  $\beta_1^*$ ) dans  $F \otimes F$ , et  $A^*$  (resp.  $B^*$ ) sa matrice associée dans la base  $\mathcal{F} \otimes \mathcal{F}$ . Puisque  $\beta_1$  a pour matrice associée dans  $F \vee F$ , la matrice  $B = -(2I_{n^*} + {}^1RR)$ , CRITCHLEY (1982) montre que l'écriture matricielle de  $\alpha_1^*$  et  $\beta_1^*$  est simplifiée lorsqu'on munit  $\mathbb{R}^n \vee \mathbb{R}^n$  du produit scalaire de matrice  $-B$ . Nous adoptons ici cette convention et notons encore  $F \vee F$  l'espace vectoriel Euclidien  $(\mathbb{R}^n \vee \mathbb{R}^n, -B)$ , qui est encore isomorphe à  $\mathcal{Y}_n$ . CRITCHLEY (1982) vérifie alors par application directe de (42) et (43) :

**Théorème 3** (CRITCHLEY 1982) :

$$i) B^* = -2I_{n^*} \text{ et } A^* = -\frac{1}{2} I_{n^*}$$

ii) ou de façon équivalente

$$\forall H \in \mathcal{Y}_H \quad \beta_1^*(H) = -2(H - H \underline{1}_n \underline{1}_n^* I)$$

$$\forall W \in \mathcal{Y}_1 \quad \alpha_1^*(W) = -\frac{1}{2}(W - W^* I)$$

L'opérateur  $\beta_1^*(H)$  joue en effet un rôle central dans l'élaboration de méthodes de Codage Multidimensionnel (cf. §VI.3) et l'opérateur  $\alpha_1^*(W)$  permet d'interpréter la méthode de résolution du problème MINRES 2 en termes de perturbation de la matrice  $W$  par une matrice diagonale  $D_{\underline{w}}$ , proposée par BROWNE (1985) et étudiée au paragraphe §VI.3.

La matrice diagonale  $H \underline{1}_n \underline{1}_n^* I$  a pour terme général  $h_i$  et donc on obtient les matrices  $N_H$  et  $N_W$  qui présentent une forme particulière et sont très utilisées dans la suite de notre exposé :

$$\begin{aligned} \beta_1^*(H) &= 2 N_H \quad \text{où} \quad N_H = D_{\underline{h}} - H \\ \alpha_1^*(W) &= \frac{1}{2} N_W \quad \text{où} \quad N_W = D_{\underline{w}} - W \quad \underline{w} = (w_{ii}) \end{aligned}$$

Lorsqu'on applique à  $\alpha_1^*$  et  $\beta_1^*$  des résultats analogues établis plus haut pour  $\alpha_1$  et  $\beta_1$ , on obtient de façon immédiate le lemme suivant (CRITCHLEY (1982))

**Lemme 8:**

$$i) (H \in \mathcal{Y}_1 \cap \mathcal{Y}_H) \Leftrightarrow (\alpha_1^*(W) = -\frac{1}{2} H) \Leftrightarrow (\beta_1^*(W) = -2 H)$$

$$ii) \alpha_1^*(Q_1) = \frac{1}{2n} (\underline{1} \underline{1}^* - I)$$

$$iii) \beta_1^*(\underline{1} \underline{1}^* - I) = 2n Q_1$$

On vérifie de plus (CRITCHLEY 1982) que pour tout couple  $(H, W)$  vérifiant  $W = \beta_1^*(H)$  ou de façon équivalente  $H = \alpha_1^*(W)$ , les vecteurs  $\underline{w}$  et  $\underline{h}$  se correspondent de façon simple :

**Proposition 13:** Si  $W = \beta_1^*(H)$ , on vérifie

$$i) \underline{h} = \frac{1}{2} \underline{w}$$

$$ii) (\underline{h} = \underline{0}) \Leftrightarrow (H \in \mathcal{Y}_1 \cap \mathcal{Y}_H) \Leftrightarrow (W \in \mathcal{Y}_1 \cap \mathcal{Y}_H) \Leftrightarrow (\underline{w} = \underline{0})$$

Par suite si  $\underline{h} = \lambda \underline{1}_n$ , où  $\lambda$  est un réel positif, alors  $\lambda$  est valeur propre de  $H$ , associée au vecteur propre  $\frac{1}{\sqrt{n}} \underline{1}_n$ . On écrit alors  $N_H = (\lambda I - H)$  de sorte que les décompositions spectrales de  $N_H$  et de  $H$  sont étroitement liées dans ce cas particulier. En général, ces deux matrices ont des spectres de valeurs propres très différents, en particulier lorsque  $\delta$  n'est pas Euclidienne.

#### IV.7. ANALYSE DES RELATIONS ALGÈBRIQUES ENTRE F ET $F \wedge F$

Nous supposons dans la suite qu'est donnée une bijection de  $I$  dans  $[n]$  où  $n = \text{card}(I)$ . Soit  $Q$  un élément de  $\mathcal{Y}_n^+$ . Les éléments de  $Q$  vérifient pour tout  $(i,j) \in [n] \times [n]$

$$(C1) \quad q_{ij} = q_{ji}$$

$$(C2) \quad q_{ij} \geq 0$$

Posons de plus :

$$q_{i+} = \sum_{j \neq i}^n q_{ij} \quad \text{et} \quad q_{.+} = \sum_i^n q_i.$$

On peut alors interpréter simplement du point de vue géométrique le lemme suivant dû à COXETER (1948) :

**Lemme 9:** Soit  $Q$  une matrice carrée d'ordre  $n$  vérifiant les conditions (C1) et (C2). On a pour tout élément  $x$  de  $F$

$$\sum_i \sum_j q_{ij} (x_i - x_j)^2 = 2 \left[ \sum_i q_i x_i^2 - \sum_i \sum_j q_{ij} x_i x_j \right] \quad (44)$$

**Preuve:** Elle résulte d'une double application de l'égalité suivante

$$\begin{aligned} \sum_i \sum_j q_{ij} x_i x_j &= \sum_j \left[ (q_j - \sum_i q_{ij}) x_j - \sum_i q_{ij} x_i \right] x_j \\ &= \sum_j q_j x_j^2 - \sum_i \sum_j q_{ij} (x_j - x_i) x_j \end{aligned}$$

dans laquelle on permute le rôle de  $i$  et  $j$  dans le second terme de droite. On obtient ainsi le résultat souhaité.  $\square$

En effet, les conditions (C1) et (C2) permettent d'associer à l'élément  $Q$  de  $\mathcal{F}_n^+$  une forme bilinéaire diagonale sur  $F \wedge F$ , de matrice représentative  $D_Q$ . Celle-ci induit elle-même une structure Euclidienne sur  $F \wedge F$  pour laquelle la norme de tout vecteur  $\underline{y}$  de  $F \wedge F$  associé à un élément  $Y$  de  $\mathfrak{A}_n$  de la forme  $(\underline{x}^t \underline{1} - \underline{1}^t \underline{x})$ , où  $\underline{x}$  appartient à  $F$ , s'écrit

$$\|\underline{y}\|_{D_Q}^2 = \sum_{i < j} \sum q_{ij} (x_i - x_j)^2 \quad \underline{y} = (x_i - x_j)_{ij}$$

Le lemme de Coxeter permet d'associer à l'espace  $F \wedge F$  muni de la géométrie Euclidienne de matrice  $D_Q$ , une structure Euclidienne sur  $F$ . Notons en effet

$$N = D_q - Q \quad Q = (q_{ij})_{i,j=1}^n$$

$$n_{ij} = \begin{cases} -q_{ij} & \text{si } i \neq j \\ \sum_{k \neq i} q_{ik} & \text{si } i = j \end{cases} \quad D_q = \text{diag}(q_i)_{i=1}^n$$

$$D_Q = \text{diag}(q_{ij})_{i,j=1}^n$$

$N$  est la matrice dans  $F$  d'une forme quadratique semi-définie positive et détermine donc une géométrie Euclidienne sur  $F$ . Plus précisément, la relation (44) s'écrit encore dans  $F$

$$\|\underline{x}\|_N^2 = \frac{1}{2} \sum_i \sum_j q_{ij} y_{ij}^2 = \sum_i \sum_{j>i} q_{ij} y_{ij}^2 \quad (45)$$

en notant  $y_{ij} = x_i - x_j$  le terme général de la matrice  $Y = \underline{x}^t \underline{1} - \underline{1}^t \underline{x}$ . Remarquons de plus que  ${}^t Y = -Y$  et donc  $Y$  définit bien une forme bilinéaire alternée sur  $F$  de sorte que si  $\underline{y} = \text{Veca}(Y)$  ( $\underline{y} \in F \wedge F$ ), le lemme de Coxeter s'écrit de façon équivalente :

$$\|\underline{x}\|_N^2 = \|\underline{y}\|_{D_Q}^2$$

**Remarque :** La contrainte (C1) n'est pas fondamentale ici. On l'impose de sorte à assurer le caractère symétrique de la matrice  $N$  que l'on interprète alors comme une matrice de produits scalaires sur  $F$ . On se limite ainsi au cadre usuel de l'Analyse Statistique Multivariée.

L'étude de la matrice antisymétrique  $Y$  conduit à expliciter les relations liant  $F$  à l'espace vectoriel produit extérieur de  $F$  par lui-même, noté  $F \wedge F$ .  $F \wedge F$  est un sous-espace vectoriel de dimension  $n^* = \frac{n(n-1)}{2}$  du produit tensoriel  $F \otimes F$  que l'on peut décrire comme plus haut à l'aide des vecteurs de base  $f_i \wedge f_j$  associés à la base  $\mathcal{F} = \{f_i, i \in [n]\}$  de  $F$  par la relation :

$$f_i \wedge f_j = f_i \otimes f_j - f_j \otimes f_i \quad i < j \quad i, j \in [n]$$

Pour tout  $(i,j) \in I \times I$ ,  $i \neq j$ ,  $f_i \wedge f_j$  définit une application bilinéaire alternée, dont la matrice dans la base  $\{f_i \otimes f_j, i,j \in I\}$  de l'espace  $F \otimes F$  a pour seuls termes non nuls le terme  $(i,j)$  égal à 1 et le terme  $(j,i)$  égal à -1.

**Remarque :** D'un point de vue statistique, les éléments  $f_i \wedge f_j$  fournissent une interprétation algébrique aux codages classiquement utilisés lors du calcul des statistiques liées à l'indice TAU B proposé par Kendall pour mesurer l'association entre variables ordinales. Les codages de la forme  $Y = \mathbf{x} \mathbf{1} - \mathbf{1} \mathbf{x}$  offrent une première généralisation des codages de Kendall, à l'aide de codages antisymétriques des paires et cette généralisation permet d'exprimer tous les indices d'association se déduisant de l'indice de corrélation linéaire, par transformation des données. Ce codage correspond en particulier à l'écriture du coefficient de corrélation linéaire empirique sous la forme

$$r(X,Y) = \frac{\sum_{i < j} (x_i - x_j)(y_i - y_j)}{\sqrt{\sum_{i < j} (x_i - x_j)^2} \sqrt{\sum_{i < j} (y_i - y_j)^2}}$$

Plus généralement, les codages  $y_{ij}$  qui s'expriment comme terme général d'une matrice antisymétrique d'ordre  $n$  fondent l'étude de l'association entre variables aléatoires développée par DANIELS (1944) (cf. aussi HUBERT (1977, 1983, 1985), LE CALVE (1976), LERMAN (1973)). Tous ces indices se décomposent dans la base  $\mathcal{F} \wedge \mathcal{F}$  de  $F \wedge F$ . Lorsqu'on munit cet espace d'une géométrie Euclidienne induite par une métrique  $\Phi$ , on obtient une interprétation géométrique des travaux de cet auteur dans un cadre plus général puisque cet auteur se limite au cas où  $\Phi$  est l'identité ainsi qu'une généralisation immédiate de ses résultats au cas de procédures d'échantillonnage complexes ( $\Phi$  est alors associée à une mesure de dénombrement non nécessairement diagonale) ou de données manquantes ou structurellement nulles comme dans le cas des problèmes de déploiement multidimensionnel (cf. d'AUBIGNY (1979), de LEEUW et HEISER (1980, 1982), DAVISON (1984)) prises en compte par le choix d'une pondération nulle des paires manquantes où plus généralement (si  $\Phi$  n'est pas diagonale) des couples de paires manquantes.

La correspondance algébrique entre  $F$  et  $F \wedge F$  peut être résumée par le schéma de dualité suivant, dont l'analogie avec le schéma résumant la correspondance entre  $F$  et  $F \vee F$  induite par l'application  $r$  doit être soulignée

$$\begin{array}{ccc}
 F^* & \xleftarrow{t} & (F \wedge F)^* \\
 \uparrow N & & \uparrow \Phi \\
 F & \xrightarrow{t_t} & F \wedge F
 \end{array}$$

La correspondance  $t$  est donc caractérisée par la relation

$$t(f_i^* \wedge f_j^*) = f_i^* - f_j^* \quad (46)$$

$t$  est donc représentée par la matrice  $T$  définie par:

$${}^tT = \begin{bmatrix} S_1 \\ O_2 S_2 \\ \vdots \\ O_{n-1} S_{n-1} \end{bmatrix} \quad S_i = \begin{bmatrix} \mathbf{1}_{n-i} & \vdots & -I_{n-i} \end{bmatrix} \quad i = 1, 2, \dots, n-1$$

$\mathbf{1}_{n-i}$  (resp.  $I_{n-i}$  et  $O_{n-i}$ ) est le vecteur constant de  $\mathbb{R}^{n-i}$  de terme général égal à 1 (resp. la matrice Identité d'ordre  $(n-i)$  et la matrice nulle de  $M_{(n-i),(i-1)}(\mathbb{R})$ ).  $S_i$  est un outil classique de description d'un plan d'expérience. C'est en effet une matrice de simples contrastes entre les  $(n-i)$  dernières coordonnées d'un vecteur  $\underline{x}$  de  $F$ . Elle est donc de dimension égale à  $(n-i) \times (n-i+1)$ . Par suite la matrice  $T$  est la matrice d'ordre  $n \times n^*$  qui décrit toutes les comparaisons par paires d'éléments distincts.  $T$  vérifie l'identité

$${}^tT \underline{x} = \underline{y} = \text{veca} (\underline{x} \mathbf{1} - \mathbf{1} \mathbf{x}) \quad (47)$$

Ainsi pour tout élément  $\underline{x}$  de  $F$ ,  $\underline{y}$  donne les coordonnées du vecteur de  $F \wedge F$  associé à la matrice antisymétrique  $\underline{x} \mathbf{1} - \mathbf{1} \mathbf{x}$  de terme général :

$$y_{ij} = x_i - x_j \quad \text{si } i < j.$$

Par construction  $Y$  est une matrice de rang  $(n-1)$  et donc  ${}^tT(F)$  est un sous-espace vectoriel de dimension  $(n-1)$  de  $F \wedge F$ . Le plongement du problème dans le sous-espace vectoriel  $F \wedge F$  de  $F \otimes F$  consiste en pratique à négliger les informations apportées par l'un des termes  $(i,j)$  et  $(j,i)$ . Ceux-ci sont en effet redondants du fait de l'antisymétrie de la matrice  $Y$ .

Notons comme précédemment  $F_1$  (resp.  $F_1^*$ ) le sous-espace de  $F$  (resp.  $F^*$ ) orthogonal pour la métrique identité à la droite des constantes  $\Delta_{1F}$  (resp.  $\Delta_{1F^*}$ ).

$$F = F_1 \oplus \Delta_{1F}$$

$$F^* = F_1^* \oplus \Delta_{1F^*}$$

Puisque les vecteurs  $(f_i^* - f_j^*)$  sont linéairement indépendants de  $\underline{1}_F$  et engendrent  $F_1^*$ , la relation (46) montre que  $t$  définit une isométrie partielle entre  $(F \wedge F)^*$  et le sous-espace  $F_1^*$ . Plus précisément, lorsqu'on munit  $F$  de la géométrie Euclidienne induite par la métrique définie par la forme quadratique de rang au plus égal à  $(n-1)$

$$N = T\Phi {}^tT \quad (48)$$

$\underline{1}_F$  est un élément du noyau de  $N$ .  $N$  agit donc sur  $F_1$ , assurant une isométrie partielle entre  $(F_1, N)$  et  $(F \wedge F, \Phi)$ . On note  ${}^tT(F_1)$  le sous-espace vectoriel de  $F \wedge F$  image de  $F_1$  par  ${}^tT$ . Tout élément de  ${}^tT(F_1)$  s'écrit alors  $\underline{x} {}^t\underline{1} - \underline{1} {}^t\underline{x}$  où  $\underline{x}$  est élément de  $F_1$ .  ${}^tT(F_1)$  contient donc les éléments  $Y$  de  $\text{Ker}(\alpha)$  qui vérifient de plus

$$Y * I = 0$$

La pratique de l'Analyse des Données conduit à choisir des métriques  $\Phi$  particulières. Ainsi la minimisation de l'indice proposé par OBENCHAIN suppose le choix d'une métrique des poids, décrite dans  $F \wedge F$  par la matrice diagonale  $\Phi = D_Q$ , dont l'interprétation statistique générale est abordée plus loin. Il lui correspond par l'application  $t$ , la matrice  $N = D_Q - Q$  décrite par le lemme de Coxeter. En particulier, les résultats du paragraphe précédent montrent que toute matrice  $H$  de  $\mathcal{Y}_H$  a pour image par  $\beta_1^*$  la matrice  $2N_H$ , qui s'interprète ici comme la matrice associée à une forme quadratique définie sur  $F_1$ . De même nous avons vu que toute matrice  $W$  de  $\mathcal{Y}_1^+$  est telle que

$$\alpha_1^*(W) = \frac{1}{2} N_W.$$

Ces propriétés sont utilisées dans la suite pour l'étude des méthodes de Codage Multidimensionnel. Cependant, nous étudions deux cas particuliers d'utilisation courante en Statistique,  $Q$  est alors interprétée comme une pondération réelle positive définie sur  $I \times I$ . Supposons tout d'abord que

$$(a) \quad q_{ij} = p_i p_j \quad p_i > 0 \quad \sum p_i = 1 \quad (\text{i.e. } Q = P \otimes P)$$

Alors  $q_{i+} = p_i$  et il correspond à  $Q$  une matrice  $N_Q = T D_Q {}^tT$  définie par

$$n_{ij} = \begin{cases} p_i (1 - p_i) & \text{si } i = j \\ - p_i p_j & \text{si } i \neq j \end{cases}$$

Interprétons  $p_i$  comme la probabilité de l'évènement  $\{i\}$ , alors la donnée d'une mesure tensorielle correspond à une hypothèse d'indépendance des unités statistiques et se traduit par une matrice  $N_Q$  associée au modèle multinomial défini par  $(p_1, \dots, p_n)$ . La matrice  $N_Q$  s'écrit alors formellement

$$N_Q = D_p(I - \mathbb{1} \mathbb{1}^t D_p) = D_p Q_p \quad (49)$$

où  $Q_p$  est le projecteur  $D_p$ -orthogonal sur le supplémentaire  $D_p$ -orthogonal de  $\Delta_{1_n}$  défini plus haut.  $N_Q$  est alors égale à la matrice  $N_p = D_p Q_p = {}^t Q_p D_p$

(b) en particulier dans le cas équi pondéré ( $p_i = \frac{1}{n}$  pour tout  $i$ )

$$n_{ij} = \frac{1}{n} \left( 1 - \frac{1}{n} \right) = \frac{n-1}{n^2}$$

et donc

$$N = \frac{1}{n} \left( I - \frac{1}{n} \mathbb{1} \mathbb{1}^t \right) = \frac{1}{n} Q_1 = \frac{1}{n} I_n Q_1$$

cette égalité montre de plus que l'on peut factoriser l'opérateur  $Q_1$  en fonction de la matrice  $T$ :

$$Q_1 = T {}^t T$$

En notant  $N_Q$  la métrique  $N$  induite par les pondérations  $q_{ij}$ , l'isométrie déterminée par l'application  $t$  permet de réécrire la formule (45) sous la forme condensée:

$$\|x\|_{N_Q \otimes I}^2 = \|y\|_{D_Q}^2$$

Les conséquences de cette égalité pour la définition d'algorithmes de Codage Multidimensionnel sont traitées ci-dessous. En particulier, cette égalité se généralise par additivité au traitement d'une matrice de configuration  $X$  quelconque, c'est à dire à l'étude des quantités :

$$\begin{aligned} \eta^2(X) &= \sum_{i < j} \sum q_{ij} d_{ij}^2(X) = \sum_{s=1}^p \left[ \sum_{i < j} q_{ij} (x_i^s - x_j^s)^2 \right] \\ &= \sum_{s=1}^p \|x^s\|_N^2 = \text{Trace}(XN {}^t X) \\ &= \text{Trace}(W_X N) \end{aligned}$$

où l'on pose  $W_X = {}^t X X$ . On vérifie de plus que

$$\text{Trace}({}^t X X N) = \|x\|_{N \otimes I}^2$$

où  $\|x\|_{N \otimes I}$  désigne la norme de Hilbert Schmidt de l'opérateur  $x : F^* \rightarrow E$  regardé comme un élément de  $F \otimes E$ , muni de sa structure Euclidienne canonique associée à la semi-métrique  $N \otimes I$ .

Plus généralement, lorsque l'espace  $F \otimes E$  est muni de la structure Euclidienne de matrice  $N_Q \otimes M$ , on vérifie le résultat suivant

**Théorème 4:** Si  $X$  est une représentation Euclidienne de  $I$  dans  $E = (\mathbb{R}^P, M)$ , alors

$$\eta^2(X) = \sum_i \sum_{i < j} q_{ij} d_{ij}^2(X) = \text{Trace}(W_X N_Q) = \| \underline{x} \|_{N_Q \otimes M}^2$$

où

$$W_X = {}^t X M X$$

**Preuve :**

$$d_{ij}^2(X) = \| \underline{x}_i - \underline{x}_j \|_M^2 = \sum_{s=1}^p \sum_{t=1}^p m_{st} (x_i^s - x_j^s)(x_i^t - x_j^t)$$

et donc

$$\begin{aligned} \eta^2(X) &= \sum_{s=1}^p \sum_{t=1}^p m_{st} \left[ \sum_i \sum_{i < j} q_{ij} (x_i^s - x_j^s)(x_i^t - x_j^t) \right] \\ &= \sum_{s=1}^p \sum_{t=1}^p m_{st} \langle \underline{x}^s, \underline{x}^t \rangle_{N_Q} = \sum_{s=1}^p \sum_{t=1}^p m_{st} \text{Trace}({}^t \underline{x}^t N_Q \underline{x}^s) \\ &= \text{Trace} \left( \sum_{s=1}^p \sum_{t=1}^p \underline{x}^s m_{st} {}^t \underline{x}^t N_Q \right) = \text{Trace}({}^t X M X N_Q) \end{aligned}$$

et donc si  $\underline{x} = \text{Vec}(X)$

$$\eta^2(X) = \| \underline{x} \|_{N_Q \otimes M}^2 \quad \parallel$$

**Remarque :** L'intérêt pratique du lemme de Coxeter repose en particulier sur l'hypothèse de symétrie de la pondération  $Q_{|X|}$ . Celle-ci n'est pas toujours vérifiée en pratique. On associe alors à  $Q_{|X|}$  une matrice symétrique définie comme la projection  $\hat{Q}$  de  $Q$  sur l'espace  $\mathcal{Y}_n$ . On applique alors le lemme de Coxeter à  $\hat{Q}$ . L'opération de symétrisation se traduit simplement de façon analytique, puisque

$$\hat{Q} = \frac{Q + {}^t Q}{2}$$

Par suite, on construit la matrice  $N$  de terme général:

$$n_{ij} = \begin{cases} -\frac{1}{2} (q_{ij} + q_{ji}) & \text{si } i \neq j \\ \sum_{k \neq i} \frac{1}{2} (q_{ik} + q_{ki}) & \text{si } i = j \end{cases}$$

A notre connaissance le cas de pondérations de signe quelconque ne semble pas avoir fait l'objet d'applications concrètes. On construit alors la projection de  $Q$  sur le cône  $\mathcal{Y}_n^+$  des matrices semi-définies positives associées aux applications linéaires définies sur  $F$  à valeurs dans  $F^*$ . Cette projection n'a pas d'expression analytique simple. Mais puisque  $\mathcal{Y}_n^+$  est un cône convexe fermé,  $\hat{Q}$  existe et permet d'associer à  $Q$  une métrique Euclidienne  $N$  sur  $F$ .

Complétons alors le schéma de dualité précédent de façon à décrire totalement les outils d'analyse d'un triplet  $(I, d(X), Q_{|x|})$  associé à une distance Euclidienne. La structure algébrique du problème est résumée par les éléments suivants

$$\begin{array}{ccccc}
 E & \xleftarrow{X} & F^* & \xleftarrow{T} & (F \wedge F)^* \\
 \begin{array}{c} \downarrow M \\ \uparrow V_X \end{array} & & \begin{array}{c} \downarrow W_X \\ \uparrow N_Q \end{array} & & \begin{array}{c} \downarrow \Psi_X \\ \uparrow D_Q \end{array} \\
 E^* & \xrightarrow{{}^t X} & F & \xrightarrow{{}^t T} & (F \wedge F)
 \end{array}$$

où

$$\begin{aligned}
 V_X &= XN_Q {}^t X \\
 W_X &= -\frac{1}{2} Q_1 D^{*2}(X) {}^t Q_1 \\
 \Psi_X &= {}^t T W_X T \\
 N_Q &= T D_Q {}^t T
 \end{aligned}$$

et

$$\begin{aligned}
 \eta^2(X) &= \|x\|_{N_Q \otimes M}^2 = \|X\|_{M, N_Q}^2 \\
 &= \text{Trace}(V_X M) = \text{Trace}(W_X N_Q) = \text{Trace}(\Psi_X D_Q) \quad (50) \\
 \text{rang}(N_Q) &= \dim(F_1) = \dim({}^t T(F)) = (n-1) \\
 \dim(E) &= \text{rang}(M) = p \\
 \dim(F) &= n \\
 \dim(F \wedge F) &= n^* = \frac{n(n-1)}{2}
 \end{aligned}$$

## CHAPITRE V

### LES MÉTHODES D'ANALYSE DU TRIPLE $(I, \delta, P_I)$

Nous étudions dans ce chapitre les méthodes d'Analyse du Triple  $(I, \delta, P_I)$  où  $I$  est un ensemble de cardinal fini  $n$ . On suppose alors donnée une bijection de  $I$  dans  $[n]$ .  $P_I$  est une pondération réelle de  $I$ , que l'on suppose normalisée de sorte que

$$\sum_{i=1}^n p_i = 1$$

$\delta$  est une dissimilarité définie sur  $I \times I$ . Les résultats du chapitre I permettent de considérer que  $\delta$  est une dissimilarité propre.

Nous étudions tout d'abord le cas d'une fonction de dissimilarité Euclidienne. L'Analyse en Coordonnées Principales permet alors de réaliser un plongement irréductible de  $(I, \delta)$  dans un espace affine Euclidien. Nous nous attachons à proposer des interprétations nouvelles de la solution, susceptibles selon nous de fixer le statut statistique de la méthode. L'introduction d'une phase de réduction de dimension conduisant à une représentation résumée des données, complète alors la méthode.

Nous étudions ensuite un ensemble de méthodes, qui offre une alternative à cette démarche classique. Nous montrons en particulier que ce cadre d'étude conduit à distinguer

- (a) les méthodes d'Analyses en Coordonnées, de nature tautologique
- (b) les méthodes d'Analyse Factorielle, de nature modélisante.

Nous établissons ainsi un parallèle complet entre les méthodes d'Analyse du triple  $(I, \delta, P_I)$  et les méthodes d'analyse des données de profil caractérisées par le triple  $(J, V_X, M)$  où  $V_X$  est la matrice de covariance associé au tableau de données  $X$ .

Nous étudions enfin le cas de fonctions de dissimilarité quelconques. Nous donnons alors une interprétation nouvelle du problème dit de la constante additive. Le lecteur de cette section sera alerté par l'absence d'un statut mathématique clair des méthodes présentées dans ce cas. De notre point de vue, nous sommes alors confrontés à des heuristiques qui doivent être jugées comme telles.

Nous avons montré plus haut que la donnée d'une pondération réelle  $\gamma$  de  $F^*$  équivaut au choix d'une origine  $P_0$  dans l'espace affine de représentation de dimension  $p$ , noté  $E$ . Nous nous limitons ici au cas d'une mesure positive  $P_1$ , généralement retenu dans les ouvrages français.

### V.1. L'ANALYSE DU TRIPLE $(I, \delta(Z), P_1)$ DÉFINI PAR UNE DISSIMILARITÉ EUCLIDIENNE

Nous nous intéressons ici à l'ensemble particulier  $\mathcal{D}_E(I)$  des fonctions de dissimilarité Euclidienne définies sur  $I$ . FICHET (1983) donne de nombreux exemples de fonctions de dissimilarité d'emploi courant en Biologie et en Ecologie qui satisfont cette propriété. De plus, GOWER (1984, 1985b), GOWER et LEGENDRE (1984), FICHET (1985) et CRITCHLEY (1985) produisent une discussion détaillée des propriétés Euclidiennes ou plus généralement métriques des fonctions de dissimilarité les plus couramment utilisées. Les travaux de FICHET (1987) et de LE CALVE (1986, 1987, 1988) traduisent la vitalité des efforts de classification des espaces quasi-métriques finis en fonction de leurs propriétés mathématiques, en particulier en rapport avec les problèmes d'existence de représentations Euclidiennes ou arborées, susceptibles de guider le choix d'une méthode appropriée d'analyse d'un tableau de dissimilarité observée.

#### V.1.1. L'Analyse en Coordonnées Principales

##### V.1.1.1. Formulation de la méthode.

Soient  $\delta(Z)$  une fonction de dissimilarité Euclidienne définie sur  $I$  et  $(I, \delta(Z), P_1)$  le triple que l'on se propose d'analyser dans un cadre Euclidien. Il existe alors une application  $z$  de  $\mathcal{L}(F^*, E)$  qui fournit une représentation de  $I$  dans un espace affine  $E$  caractérisée par :

$$\begin{aligned} z : F^* &\rightarrow E \\ f_i^* &\mapsto z(f_i^*) = P_i \end{aligned}$$

La donnée de  $P_1$  impose le choix d'une origine. Aussi ne différencions nous pas la structure affine de  $E$  de sa structure d'espace vectoriel induite sur  $E$  par la fixation de cette origine. Supposons de plus que  $E$  est muni d'une structure Euclidienne induite par la forme quadratique  $m$ . Alors, pour un choix de bases  $(\mathcal{E}, \mathcal{F}^*)$  défini plus haut,  $P_i$  est représenté par le vecteur  $z_i$  tel que si  $E = (\mathbb{R}^p, M)$  où  $M$  est la matrice associée à  $m$ , on a

$$\delta_{ij}(Z) = \| z_i - z_j \|_M$$

Notons alors  $H(Z)$  le carré de Hadamard de la matrice  $\Delta(Z)$ :

$$H(Z) = \Delta(Z) * \Delta(Z)$$

$$W_Z = -\frac{1}{2} Q_p H(Z) {}^t Q_p$$

Par suite, puisque  $\delta(Z)$  est une fonction de dissimilarité Euclidienne,  $H(Z)$  est un élément du sous-ensemble  $\mathfrak{D}_E^{*2}$  de  $\mathfrak{Y}_H^+$  et l'on vérifie par application des résultats ci-dessus

$$W_Z = \alpha_p^+(H(Z)) = (w_{ij})_{i,j \in [n]}$$

où

$$w_{ij} = -\frac{1}{2} [\delta_{ij}^2(Z) - \delta_{i.}^2(Z) - \delta_{.j}^2(Z) + \delta_{..}^2(Z)]$$

$$\delta_{i.}^2(Z) = \sum_{j=1}^n p_j \delta_{ij}^2(Z)$$

$$\delta_{..}^2(Z) = \sum_{j=1}^n p_j \delta_{i.}^2(Z) = \sum_{i=1}^n \sum_{j=1}^n p_i p_j \delta_{ij}^2(Z) = {}^t p H p = \text{Trace}(H p {}^t p)$$

et  $p$  est l'élément  $p = D_p \mathbf{1}_F$  de l'orthant positif de  $F^*$ .

Puisque  $H(Z)$  appartient à  $\mathfrak{D}_E^{*2}$ ,  $W_Z$  est un élément de  $\mathfrak{Y}_p^+$  et donc  $(I, \delta(Z))$  admet un plongement Euclidien irréductible dans un espace affine  $E$  de dimension  $\text{rang}(W_Z)$ . On note alors  $(\mathcal{P}, E)$  une image Euclidienne de  $(I, \delta(Z))$  dans  $E$ .  $\mathcal{P}$  est un ensemble de points massiques  $\{(P_i, p_i); i \in [n], P_i \in E\}$ .

La méthode usuellement retenue pour construire  $(\mathcal{P}, E)$  est connue depuis YOUNG et HOUSEHOLDER (1938) lorsque la pondération  $P_i$  est construite de sorte que  $P_0$  soit l'un des points  $P_i$ . Tucker, Green et Abelson (cf. TORGERSON (1952)) traitent le problème dans un cadre analogue à celui que nous présentons ici. Ces auteurs considèrent le cas particulier  $p_i = \frac{1}{n}$  pour tout  $i$ . GOWER (1966) présente une étude très détaillée de cette approche souvent appelée Classical Scaling dans la littérature anglaise. Le cas d'une pondération positive quelconque est traité par les auteurs français (cf. BENZECRI (1973), CAILLEZ et PAGES (1976), MAILLES (1981)). La méthode classique conduit à construire un élément  $X$  de  $[X]$  (cf. § II.1) où

$$[X] = [\varepsilon_p]^{-1}(\Delta(Z)) \quad X \in \mathcal{M}_{r,n}(\mathbb{R})$$

La représentation  $X$  est obtenue par l'analyse spectrale de l'opérateur  $U_Z = W_Z D_p$  et le point  $P_i$  est alors représenté par le vecteur  $\underline{x}_i$ ,  $i$ -ème colonne de la matrice  $X$ .  ${}^t X$  a pour  $s$ -

ième colonne le vecteur propre  $\underline{x}^s$  associé à la valeur propre  $\lambda_s$  de  $U_Z$ , les valeurs propres étant rangées par ordre décroissant

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = 0 \quad \text{où } r = \text{rang}(U_Z)$$

On résume cette approche par la proposition suivante (cf. MAILLES (1981) par exemple)

**Proposition 1:** On appelle *Coordonnées Principales du Triple*  $(I, \delta(Z), P_1)$  les composantes principales  $D_p$ -orthogonales de  $W_Z$ . Ce sont les vecteurs propres de  $U_Z = W_Z D_p$  solutions de l'équation aux valeurs propres

$$U_Z {}^t X = W_Z D_p {}^t X = \Lambda {}^t X$$

La solution est complètement déterminée lorsqu'on impose une contrainte d'échelle aux vecteurs propres de  $U_Z$ . Deux conventions sont le plus souvent retenues:

- i)  $X D_p {}^t X = \Lambda$
- ii)  $X D_p {}^t X = I$

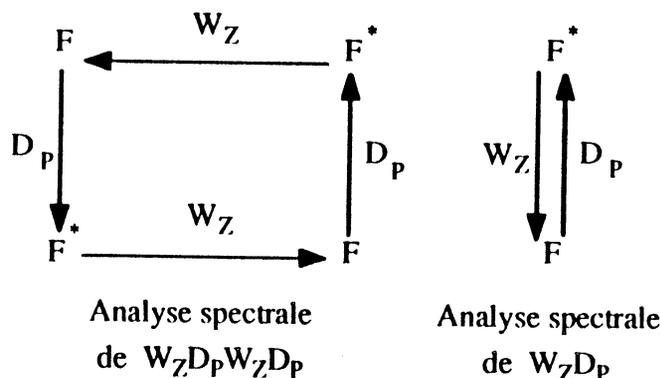
Dans le premier cas, les vecteurs propres sont  $D_p$ -normés à la valeur propre correspondante. Dans le deuxième cas, ils sont  $D_p$ -normés à 1.

Cette approche considère donc  $E$  comme un sous-espace vectoriel de dimension  $p$  de  $F$ . D'autre part, la solution est aussi solution de l'Analyse en Composantes Principales du triplet  $(W_Z, D_p, D_p)$ . Alors les vecteurs colonnes de  ${}^t X$  sont les vecteurs propres de l'opérateur

$$W D_p = (W_Z D_p W_Z) D_p = (W_Z D_p)^2$$

Par suite  $W D_p$  et  $W_Z D_p$  ont les mêmes vecteurs propres associés respectivement aux valeurs propres  $\lambda_s(W D_p)$  et  $\sqrt{\lambda_s(W D_p)}$ .  $\lambda_s(W_Z D_p) = \sqrt{\lambda_s(W D_p)}$  est positive ou nulle pour tout  $s = 1, 2, \dots, r$ ,  $r = \text{rang}(Z)$ , puisque  $\Delta(Z)$  est Euclidienne.

Cette situation est résumée par les deux schémas de dualité



L'équivalence de ces deux analyses spectrales n'est assurée que si  $W_Z$  est gramienne. Dans le cas général, nous préférons rompre avec la tradition et poser la définition suivante

**Définition 1:** On appelle *Analyse en Coordonnées Principales du triple*  $(I, \delta(Z), P_1)$ , l'analyse spectrale de l'opérateur  $W_Z D_P$ . Les coordonnées principales de  $\delta(Z)$  sont alors les vecteurs propres, solutions de

$$(W_Z D_P W_Z) D_P X = \Lambda X$$

et normés selon les contraintes i) ou ii) ci-dessus.

**Remarque 1 :**

$$W_Z D_P = ({}^t Z M Z) D_P$$

et donc

$$W_Z D_P = {}^t Z M Z D_P = {}^t Z M V_Z M Z D_P$$

où

$$V_Z = Z D_P {}^t Z$$

Choisissons donc pour métrique particulière dans  $E$  la métrique de Mahalanobis  $M = V_Z^{-1}$ . On obtient

$$W_Z D_P = {}^t Z V_Z^{-1} Z D_P = W_Y D_P$$

où l'on a posé

$$(W_Y = {}^t Y Y, Y = V_Z^{-\frac{1}{2}} Z) \Leftrightarrow (\Delta(V_Z^{-\frac{1}{2}}) = \beta_P^+(W_Y))$$

On décrit donc la dissimilarité Euclidienne  $\delta(V_Z^{-\frac{1}{2}} Z)$  dans ce cas. L'étude de ce problème dans le cadre de l'Analyse en Coordonnées Principales et surtout dans celui du Codage Multidimensionnel est menée de façon détaillée par MEULMAN (1986) qui expose les problèmes concrets justifiables d'une telle approche. Nous avons montré plus haut (cf. aussi d'AUBIGNY (1988)) qu'une telle approche est affine invariante.

**Corollaire :** Les vecteurs colonnes de  $W_Z$  engendrent le même sous-espace vectoriel de  $F$  que les vecteurs propres de  $W_Z D_P$  associés à des valeurs propres non nulles.

**Preuve :** Les colonnes de  $W_Z$  engendrent le même sous-espace vectoriel de  $F$  que les composantes principales vecteurs propres non nuls de  $W_Z D_P W_Z D_P$  et donc que les vecteurs propres non nuls de  $W_Z D_P$ . ¶

La démarche précédente est tautologique en ce sens que l'on réalise la décomposition spectrale complète de  $WD_p$  ou de  $WD_pWD_p$ . Lorsqu'on se propose de construire un résumé statistique de ces opérateurs, on retient un sous-espace vectoriel de  $E$  de dimension  $p$  ( $p < r$ ). On doit donc définir un critère d'approximation. On utilise classiquement les résultats de ECKART et YOUNG (1936) (cf. aussi YOUNG et HOUSEHOLDER (1938)) qui caractérisent la solution en termes d'ajustement de la matrice de produits scalaires au sens des moindres carrés (cf. aussi KELLER (1962) pour une synthèse de ces résultats). Appliquons au problème présent les résultats présentés plus haut sur la caractérisation de la solution d'une Analyse en Composantes Principales en termes de projection. Lorsque  $\delta(Z)$  est Euclidienne, les éléments propres de  $W_Z D_p$  fournissent les composantes principales du triplet  $(Z, M, N)$  et par suite si  $Z$  est un élément de  $\mathcal{M}_{p,n}(\mathbb{R})$  et  $s = \text{Min}(r, p', n')$  où  $p' \leq p$ ,  $n' \leq n$  et  $r = \text{rang}(Z)$ , alors la solution d'ordre  $s$  de l'Analyse en Composantes Principales du triplet  $(Z, M, N)$  vérifie pour toutes matrices  $A$  de  $\mathcal{M}_{p,p'}(\mathbb{R})$ ,  $B$  de  $\mathcal{M}_{n,n'}(\mathbb{R})$  et  $C$  de  $\mathcal{M}_{p',n'}(\mathbb{R})$

$$\text{Min} ( \|Z - AH^1B\|_{M,D_p}^2 ; A, B, H ) = \|Z - U \Lambda^{\frac{1}{2}} V\|_{M,D_p}^2 = \sum_{t=s+1}^r \lambda_t(W_Z D_p)$$

où  $\hat{Z} = U \Lambda^{\frac{1}{2}} V$  est la projection de  $Z$  sur  $\langle E', F' \rangle$  avec  $\dim(E') = p'$  et  $\dim(F') = n'$ . Puisque nous avons montré plus haut que

$$I_{p_0} = \frac{1}{2} \delta_{..}^2 = \text{Trace}(W_Z D_p) = \|Z\|_{M,D_p}^2$$

on définit un indicateur d'approximation, où l'on prend en général  $p' = n' = s < r$ :

$$\sigma(Z, \hat{Z}) = \frac{\|Z - \hat{Z}\|_{M,D_p}^2}{\|Z\|_{M,D_p}^2} = \frac{\sum_{t=s+1}^r \lambda_t(W_Z D_p)}{\text{Trace}(W_Z D_p)}$$

Ce rapport s'écrit dans  $F^* \otimes E^*$

$$\sigma(Z, \hat{Z}) = \frac{\|z - \hat{z}\|_Q^2}{\|z\|_Q^2} \quad \text{avec} \quad Q = M^{-1} \otimes D_p^{-1}$$

**Remarque 2 :**

$$\hat{Z} = U \Lambda^{\frac{1}{2}} V \Rightarrow W_{\hat{Z}} = \hat{Z} M \hat{Z} = V \Lambda^{\frac{1}{2}} U M U \Lambda^{\frac{1}{2}} V = U \Lambda^{\frac{1}{2}} V$$

et donc dans le sous-espace  $E'$  solution résumée d'ordre  $s$  de l'Analyse en Coordonnées Principales, on vérifie

$$d_{ij}^2(Z) = \|\hat{z}_i - \hat{z}_j\|_M^2$$

Par suite, le moment d'inertie du nuage projeté  $\hat{\mathcal{P}} = \{(\hat{P}_i, p_i) ; i \in [n]\}$  autour de son barycentre  $\hat{P}_0 = \text{Proj}_E(P_0)$  s'écrit:

$$I_{\hat{P}_0} = \frac{1}{2} \delta_{..}^2(\hat{Z}) = \text{Trace}(W_Z D_P) = \sum_{t=1}^s \lambda_t(W_Z D_P) = \sum_{t=1}^s \lambda_t(W_Z D_P)$$

de sorte que

$$\sigma(Z, \hat{Z}) = 1 - \frac{\delta_{..}^2(\hat{Z})}{\delta_{..}^2(Z)} = 1 - \frac{\|\hat{z}\|_Q^2}{\|z\|_Q^2}$$

D'autre part, les solutions de l'Analyse en Composantes Principales du triplet  $(Z, M, N)$  vérifient les propriétés classiques d'emboîtement des sous-espaces propres et donc si  $E_s$  et  $E_t$  sont les sous-espaces propres engendrés les vecteurs propres associés respectivement aux  $s$  et aux  $t$  plus grandes valeurs propres de  $U_Z$ , alors si  $s < t$ ,  $E_s$  est inclu dans  $E_t$ .

La propriété définissant  $\hat{z}$  comme une projection de  $z$  sur  $\langle E', F' \rangle$  admet une conséquence très importante en pratique, liée au caractère contractant des projecteurs :

**Proposition 2:** Soit  $E_1 \subset E_2 \subset \dots \subset E_s \subset \dots \subset E_r$  un drapeau de sous-espaces vectoriels de  $E$  engendré par les vecteurs propres de  $W_Z D_P$  ordonnés par valeurs décroissantes des valeurs propres  $\lambda_t(W_Z D_P)$ . Alors si  $\hat{Z}_s$  est la matrice de configuration de  $I$  associée au plongement de  $(I, \delta(Z))$  dans  $E_s$ , on vérifie la chaîne d'inéquations

$$d_{ij}^2(\hat{Z}_1) \leq d_{ij}^2(\hat{Z}_2) \leq \dots \leq d_{ij}^2(\hat{Z}_r) = \delta_{ij}^2(Z) \quad \forall (i,j) \in [n] \times [n]$$

**Preuve :** Le projecteur  $D_P$ -orthogonal  $P_s$  de  $E$  sur  $E_s$  est contractant et donc

$$d_{ij}^2(\hat{Z}_1) = \|P_s(z_i - z_j)\|_M^2 \leq \delta_{ij}^2(Z)$$

de plus si  $M = {}^t T T$  et  $Y = T Z$

$$\begin{aligned} d_{ij}^2(\hat{Z}_{s+1}) &= \sum_{t=1}^s (y_i^t - y_j^t)^2 + (y_i^{s+1} - y_j^{s+1})^2 \\ &\geq \sum_{t=1}^s (y_i^t - y_j^t)^2 = d_{ij}^2(\hat{Z}_s) \quad \parallel \end{aligned}$$

Appliquons maintenant les résultats précédents à l'Analyse en Composantes Principales du triplet  $(W_Z, D_P, D_P)$ . Du fait de la symétrie de  $W_Z$ , on obtient, mutatis mutandis, en prenant  $E' = F'$  et  $\dim(F') = n'$ ,

$$\begin{aligned} \text{Min} \{ \|W_Z - AH^tA\|_{D_P D_P}^2 ; A, H \} &= \|W_Z - V \Lambda^t V\|_{D_P D_P}^2 = \sum_{t=s+1}^r \lambda_t(W_Z D_P W_Z D_P) \\ &= \sum_{t=s+1}^r \lambda_t^2(W_Z D_P) \end{aligned}$$

On définit donc l'indicateur d'approximation suivant, utile à l'évaluation de la qualité globale d'une régression et d'une Analyse en Coordonnées Principales

$$\sigma(W_Z, \widehat{W}_Z) = \frac{\|W_Z - \widehat{W}_Z\|_{D_P D_P}^2}{\|W_Z\|_{D_P D_P}^2} = \frac{\sum_{t=s+1}^r \lambda_t^2(W_Z D_P)}{\sum_{t=1}^r \lambda_t^2(W_Z D_P)}$$

On remarque que  $\sigma(W_Z, \widehat{W}_Z)$  ne s'interprète pas en termes mécanistes d'inertie d'un nuage de points massiques  $\mathcal{P}$ . Cette interprétation semble en effet très directement liée à l'hypothèse d'euclidéanité et donc de semi-définie positivité de  $W_Z$ .  $\sigma$  s'écrit dans  $F \otimes F$

$$\sigma(\underline{w}_Z, \widehat{\underline{w}}_Z) = \frac{\|\underline{w}_Z - \widehat{\underline{w}}_Z\|_Q^2}{\|\underline{w}_Z\|_Q^2} \quad \text{avec} \quad Q = D_P \otimes D_P$$

**Remarque 3:** L'approche de MAILLES (1981) permet d'autre part d'étendre la classe des méthodes d'Analyse du Triple  $(I, \delta(Z), P_I)$ . Il suffit pour cela de retenir pour l'analyse de  $W_Z$ , un choix de métrique plus général qu'il n'est utilisé en Analyse en Coordonnées Principales. On réalise alors l'Analyse en Composantes Principales de  $(W_Z, N_1, N_2)$ . La solution vérifie alors

$$\|W_Z - \widehat{W}_Z\|_{N_1, N_2}^2 = \sum_{t=s+1}^r \lambda_t(W_Z N_1 W_Z N_2)$$

Mais on perd alors le lien unissant le choix d'une métrique sur  $F$  et le choix d'une origine dans l'espace de représentation  $E$ . D'autre part, lorsqu'on pose  $N_1 = N_2 = D_P$ , les vecteurs propres associés aux valeurs propres non nulles de  $W_Z D_P$  et de  $W_Z D_P W_Z D_P$  étant identiques, on vérifie immédiatement que

$$\widehat{W}_Z = \widehat{W}_Z$$

**Remarque 4 :** Les théorèmes d'existence d'un plongement Euclidien de  $(I, \delta)$  dans  $E$ , comme solution constructive en termes d'Analyse en Composantes Principales de  $W$  repose sur les propriétés de semi-définie positivité de  $W$ . L'introduction d'une mesure réelle  $\gamma$  non nécessairement positive sur  $I$  conduit à définir sur  $F$  une forme quadratique  $D_\gamma$  non nécessairement positive.  $F$  n'est alors plus un espace Euclidien. Cependant puisque  $W_Z$  et  $D_\gamma W_Z D_\gamma$  sont encore semi-définies positives, le sous-espace de représentation  $E$  est bien réel. On retrouve ainsi sous une forme différente, une conséquence de l'étude menée plus haut des translations associées au vecteur  $\gamma$ . L'introduction de  $D_\gamma$  traduit un changement d'origine seulement et ne peut donc pas causer la perte des propriétés Euclidiennes du support affine de  $\mathcal{P}$ . La présentation de la solution en termes de décomposition spectrale de l'opérateur  $U_Z$  rend ce constat moins aisé. C'est l'une des raisons pour lesquelles nous avons défini plus haut l'Analyse en Coordonnées Principales en terme d'analyse spectrale de  $W D_P$  et non de  $W_Z D_P$ .

#### V.1.1.2. Introduction d'éléments supplémentaires

GOWER (1968) (cf. aussi Wilkinson (1970)) enrichit la méthodologie de l'Analyse en Coordonnées Principales du Triple  $(I, \delta, P_1)$  par une technique de représentation de points supplémentaires sur le support affine du nuage de points massiques  $\{(P_i, p_i); i \in [n]\}$ . Bien que développée dans le but de déterminer la non euclidéanité potentielle de  $\delta$  dans le cas d'une pondération uniforme ( $p_i = 1/n$ ), cette démarche est aisément généralisable au cas d'une pondération quelconque. Elle présente alors un grand intérêt pratique dans la phase d'interprétation des résultats d'une étude et accorde aux techniques d'analyse des tableaux de dissimilarité un statut méthodologique comparable à ceux des méthodes d'Analyse Linéaire des tableaux de données de profil. Nous présentons ici une formulation du problème due à Mailles (1981). Celle-ci explicite en effet l'analogie de cette méthode avec les techniques de points supplémentaires utilisées en Analyse en Composantes Principales et illustre le rôle des pondérations en Analyse du Triple  $(I, \delta, P_1)$ .

Considérons en effet une pondération accordant une masse nulle aux éléments d'un sous-ensemble  $J$  de  $I$ . Décomposons en conséquence  $I$  comme la réunion disjointe de l'ensemble  $K$  des points actifs et de l'ensemble  $J$  des points supplémentaires ou illustratifs. Choisissons de plus une bijection de  $I$  dans  $[n]$  respectant la bipartition de  $I$ . Par suite on obtient, avec des notations évidentes

$$D_P = \begin{bmatrix} D_K & 0 \\ 0 & 0 \end{bmatrix}$$

si l'on compte  $K$  unités statistiques actives et  $J$  unités statistiques illustratives parmi les  $n = J+K$  éléments de  $I$ .

$$Q_p = I_p - \underline{1}_p \underline{1}_p' D_p = \begin{bmatrix} Q_K & 0 \\ -A & I_J \end{bmatrix}$$

où

$$Q_K = I_K - \underline{1}_K \underline{1}_K' D_K$$

$$A = \underline{1}_J \underline{1}_K' D_K$$

par suite lorsque la matrice H est partitionnée en

$$H = \begin{bmatrix} H_{KK} & H_{KJ} \\ {}^t H_{JK} & H_{JJ} \end{bmatrix}$$

on lui associe par l'application  $\alpha_p$  la matrice des produits scalaires centrés

$$W = -\frac{1}{2} Q_p H {}^t Q_p = \begin{bmatrix} W_{KK} & W_{KJ} \\ {}^t W_{JK} & W_{JJ} \end{bmatrix}$$

où

$$W_{KK} = -\frac{1}{2} [Q_K H_{KK} {}^t Q_K]$$

$$W_{KJ} = -\frac{1}{2} [Q_K H_{KJ} - Q_K H_{KK} {}^t A]$$

$$W_{JJ} = -\frac{1}{2} [A H_{JJ} {}^t A + H_{KK} - A H_{JK} - H_{KJ} {}^t A]$$

**Proposition 3:** Soit  $(C_K, \Lambda)$  la solution de l'Analyse en Coordonnées Principales du triplet  $(K, \delta_K, P_K)$  où  $\delta_K$  est la restriction de  $\delta$  à K. Les coordonnées des J unités statistiques supplémentaires sont alors les colonnes de la matrice

$${}^t C_J = \Lambda^{-1} {}^t C_K D_K W_{KJ}$$

**Preuve:** Réalisons l'Analyse en Coordonnées Principales du triplet  $(I, \delta, P_I)$ , où la pondération  $P_I$  affecte un poids nul aux éléments de J. Les coordonnées des n unités statistiques sont alors solutions du problème aux valeurs propres

$$W D_p C = C \Lambda \Leftrightarrow W D_p \begin{bmatrix} C_K \\ C_J \end{bmatrix} = \begin{bmatrix} C_K \\ C_J \end{bmatrix} \Lambda$$

$$\Leftrightarrow \begin{bmatrix} W_{KK} & W_{KJ} \\ {}^t W_{KJ} & W_{JJ} \end{bmatrix} \begin{bmatrix} D_K & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} C_K \\ C_J \end{bmatrix} = \begin{bmatrix} C_K \\ C_J \end{bmatrix} \Lambda$$

$$\Leftrightarrow \begin{cases} W_{KK} D_K C_K = C_K \Lambda \\ W_{JK} D_K C_K = C_J \Lambda \end{cases}$$

$$\Leftrightarrow \begin{cases} W_{KK} D_K C_K = C_K \Lambda \\ W_{JK} D_K C_K \Lambda^{-1} = C_J \end{cases}$$

La première équation traite le problème classique de plongement Euclidien de  $(K, \delta_K, P_K)$  et la seconde fournit les coordonnées souhaitées des points supplémentaires

$${}^t C_J = \Lambda^{-1} {}^t C_K D_K W_{KJ} \quad \parallel$$

L'équation précédente se ramène aux équations obtenues par GOWER (1968) lorsque la pondération  $P_K$  est uniforme.

En pratique, le choix de points supplémentaires est lié à l'une des deux préoccupations suivantes. Dans le premier cas on dispose de données observées que l'on traite comme des données passives d'une analyse, ou bien du fait de leur grande influence sur l'établissement de la solution d'une analyse précédente, ou bien dans un souci de régression des informations qu'elles contiennent sur le sous-espace affine support des points actifs. Dans le second cas il s'agit de données non observées, construites dans le but de valider une interprétation suggérée par les résultats d'une étude. Dans les deux cas, il est nécessaire de disposer des valeurs de H permettant de calculer  $W_{KJ}$ . Puisque

$$W_{KJ} = -\frac{1}{2} [ Q_K H_{KJ} - Q_K H_{KK} {}^t A ]$$

La mise en œuvre de la méthode suppose la connaissance de toutes les distances entre un point de J et un point de K. Par contre, les éléments de  $H_{JJ}$  (c'est à dire les distances entre points de J) n'interviennent pas dans le calcul des coordonnées des points supplémentaires.

**Remarque:** L'application de cette approche au problème de déploiement multidimensionnel, ou Unfolding, est étudiée dans d'AUBIGNY (1981). La donnée du problème est alors le triplet  $(I, \delta_{KJ}, P_I)$  et la notation  $\delta_{KJ}$  implique que I est partitionné en deux classes K et J, K est l'ensemble des conditions k dans lesquelles sont évaluées les dissimilarités  $\delta_{kl}$ , où l décrit J. Nous montrons alors qu'une méthode d'analyse possible consiste à calculer d'abord les éléments d'une matrice de dissimilarités dérivées et donc une matrice  $H_{KK}$ , de façon à construire une représentation des éléments de K. Alors seulement, la technique des points supplémentaires permet d'obtenir une représentation conjointe de K et J, à partir des informations brutes contenues dans  $H_{KJ}$ .

### V.1.1.3. Transformation paramétrée de la fonction de dissimilarité

CRITCHLEY (1978) enrichit la méthodologie de l'Analyse en Coordonnées Principales en considérant la famille de transformations

$$\Delta(a) = (\delta_{ij}^a) = \Delta^{*a}$$

On associe alors à  $\Delta(a)$  la matrice

$$W(a) = \alpha_p(\Delta(a))$$

que l'on soumet à l'Analyse en Coordonnées Principales. Compte tenu du caractère arbitraire du choix d'une transformation à priori du tableau de données, l'intérêt méthodologique de la proposition de Critchley tient au fait que cet auteur propose de déterminer le paramètre  $a$  de façon à minimiser la valeur du critère optimisé en Analyse en Coordonnées Principales. Puisque

$$T(a) = \text{Trace} (W(a)D_p W(a)D_p) = \sum_{s=1}^n \lambda_s^2(a)$$

on recherche la valeur  $\hat{a}$  qui réalise le maximum de  $T(a)$  dans une partie bornée  $A$  de  $\mathbb{R}$  sous contrainte:

$$T(\hat{a}) = \text{Max} \{ T(a) ; a \in A, \lambda_n(a) = 0, \frac{1}{n} \sum_{s=1}^n \lambda_s(a) = 1 \}$$

Les contraintes de normalisation utilisées dans cette formule imposent d'une part le caractère de semi-définie positivité de  $W(a)$  et d'autre part, permettent d'exprimer ce critère de façon équivalente, en termes de maximisation de la variance empirique de la série des valeurs propres de  $W(a)$ . Cet objectif doit être rapproché des méthodes de détermination de la dimension d'une représentation résumée des données en Analyse en Composantes Principales ou en Analyse Factorielle discutées par ANASTASSAKOS et d'AUBIGNY (1984). Bien qu'aucune étude systématique n'ait été menée à notre connaissance, ce critère semble susceptible de favoriser les représentations Euclidiennes en faible dimension.

De plus, l'application du théorème de SCHOENBERG (1938) au problème traité par Critchley montre que si  $\delta$  est Euclidienne, il en est de même pour  $\delta^{*a}$  avec  $0 < a \leq 1$ . Il n'est cependant pas démontré de résultats liant les variations de la dimension de l'espace de représentation de  $(I, \delta^{*a}, P_I)$  à celles de  $a$ .

## V.1.2. Interprétation de l'Analyse en Coordonnées Principales en termes de contrastes 2x2

### V.1.2.1. Le rôle des contrastes 2x2 en Analyse en Coordonnées Principales

L'approche précédente fournit un plongement Euclidien de  $(I, \delta(Z), P_1)$ , sans pour autant fournir d'interprétation concrète des axes principaux construits par l'analyse spectrale de  $W_Z D_P$  ou de  $W_Z D_P W_Z D_P$ . Nous étudions ici les termes du schéma de dualité liant  $F$  et  $F \wedge F$  et nous en déduisons l'interprétation cherchée.

L'étude des opérateurs  $\alpha_\gamma^*$  et  $\beta_\gamma^*$  présentée plus haut nous a montré l'analogie des outils développés pour l'étude des tableaux de dissimilarité Euclidienne d'une part et en Analyse de la Variance d'autre part. Ainsi nous avons montré que  $\text{Ker}(\alpha_1)$  est formé des matrices  $A$  décrivant un modèle additif

$$a_{ij} = x_i + y_j \quad \forall i \in [n], \quad \forall j \in [p].$$

Interprétons maintenant la matrice  $H = (h_{jk})$  comme un tableau à double entrée résultant de la mesure d'une variable réponse  $H$  réelle sur le produit cartésien  $J \times K$ .  $J \times K$  décrit alors les éléments d'un plan d'expérience factoriel d'ordre 2 croisant les modalités de deux facteurs  $F_J$  et  $F_K$ .  $J$  et  $K$  sont supposés de cardinal fini égal respectivement à  $t$  (les traitements) et  $b$  (les blocs).

Le cas particulier où  $J = K = I$  et  $H$  est un élément de  $\mathcal{Y}_n^+$ , fournit le cadre naturel d'étude des fonctions de dissimilarité. Notre étude complète alors l'Analyse en Coordonnées Principales. Si  $J \neq K$  et si  $H$  est cependant interprétée encore en termes de carrés de dissimilarité Euclidienne, notre étude fournit une interprétation claire des éléments de la solution algébrique du problème de déploiement multidimensionnel proposé par SCHÖNEMAN (1972).

Soit donc  $H$  un tableau croisé d'ordre  $t \times b$ . TUKEY (1949) propose d'analyser de telles données dans le cadre de l'Analyse de la Variance, à l'aide du modèle structurel suivant où  $LG(0, \sigma^2)$  désigne la loi de Gauss de moyenne 0 et de variance  $\sigma^2$ :

$$h_{jk} = \mu + \alpha_j + \beta_k + \lambda \alpha_j \beta_k + \varepsilon_{jk}, \quad \varepsilon_{jk} \in LG(0, \sigma^2) \text{ et } \varepsilon_{jk} \text{ indépendantes}$$

sous les contraintes

$$\sum_j \alpha_j = \sum_k \beta_k = 0$$

Ce modèle décrit le terme d'interaction  $\gamma_{jk}$  sous la forme d'un tenseur irréductible d'ordre 1

$$\gamma_{jk} = \lambda \alpha_j \beta_k \quad \forall (j,k) \in J \times K$$

Ce modèle est étendu par JOHNSON et GRAYBILL (1972) qui retiennent le modèle

$$h_{jk} = \mu + \alpha_j + \beta_k + \lambda u_j v_k + \varepsilon_{jk}, \quad \varepsilon_{jk} \in LG(0, \sigma^2) \text{ et } \varepsilon_{jk} \text{ indépendantes}$$

sous les contraintes

$$\begin{aligned} \sum_j \alpha_j &= \sum_k \beta_k = 0 \\ \sum_j u_j &= \sum_k v_k = 0 & \sum_j u_j^2 &= \sum_k v_k^2 = 1 \end{aligned}$$

La structure algébrique de ce modèle est reprise dans le cadre des modèles Log-linéaires, pour l'étude des tables de contingence dans GOODMAN (1979 a b, 1981, 1983). Alors l'interaction est présente si et seulement si  $\lambda \neq 0$  et lorsque H est associée par isomorphisme à un élément  $\underline{h}$  de  $F \otimes G$  où  $\dim(F) = b$  et  $\dim(G) = t$ , les décompositions présentées ci-dessus correspondent à la décomposition additive de  $F \otimes G$  en

$$F_J \otimes F_K = (\mathbb{1}_b \otimes \mathbb{1}_t) \oplus (F_{J_1} \otimes \mathbb{1}_t) \oplus (\mathbb{1}_b \otimes F_{K_1}) \oplus (F_{J_1} \otimes F_{K_1})$$

de sorte que les deux modèles proposés respectivement par GOLLOB (1968) pour l'étude d'observations répliquées

$$\begin{aligned} h_{jk,l} &= \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{jk,l}, \quad \varepsilon_{jk,l} \in LG(0, \sigma^2) \text{ et } \varepsilon_{jk,l} \text{ indépendantes} \\ \gamma_{jk} &= \sum_{s=1}^r \sqrt{\lambda_s} u_j^s v_k^s \quad (\Leftrightarrow \Gamma = U \Lambda^t V) \end{aligned}$$

sous les contraintes

$$\sum_j \alpha_j = \sum_k \beta_k = 0$$

et par MANDEL (1961) pour les plans sans réplifications

$$h_{jk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{jk} \tag{51}$$

sous les mêmes hypothèses probabilistes et les mêmes contraintes d'identification des paramètres du modèle que précédemment, décrivent analytiquement une décomposition canonique de  $F_J \otimes F_K$  en somme directe orthogonale. Les modèles analogues proposés par GOODMAN (1979 a b, 1981, 1983) pour l'étude des données qualitatives agrégées définissant une table de contingence expriment les paramètres d'interaction comme la somme de  $r$  contrastes multiplicatifs.

Notons alors  $r_{jk}$  le résidu du modèle additif

$$r_{jk} = h_{jk} - h_{j.} - h_{.k} + h_{..} \quad (52)$$

On vérifie, sous les hypothèses probabilistes faites plus haut, que

$$E(r_{jk}) = \gamma_{jk} - \gamma_{j.} - \gamma_{.k} + \gamma_{..}$$

$$\text{Var}(r_{jk}) = \frac{(t-1)(b-1)}{bt} \sigma^2$$

$$\text{Cov}(r_{jk}, r_{j'k'}) = \frac{1}{bt} (t \delta_j^{j'} - 1)(b \delta_k^{k'} - 1) \sigma^2$$

où

$$\delta_j^k = 1 \text{ si } j = k \text{ et } 0 \text{ sinon}$$

Par suite, si toute réalisation de  $H$  est un élément de  $\mathcal{Y}_n^+$ , alors la matrice  $\Gamma$  de terme général  $\gamma_{jk}$  appartient à  $\mathcal{Y}_n^+$ . De plus, la matrice  $R$  de terme général  $r_{jk}$  vérifie

$$R = \alpha_1^*(H) \text{ et } E(R) = \alpha_1^*(\Gamma)$$

De façon complémentaire si  $J \neq K$ ,

$$R_{JK} = -\frac{1}{2} Q_J H_{JK} {}^t Q_K$$

où l'on précise le caractère non symétrique des matrices traitées à l'aide des indices  $J$  et  $K$ .  $Q_J$  (resp.  $Q_K$ ) est le projecteur orthogonal sur le complémentaire orthogonal  $F_{J1}$  (resp.  $F_{K1}$ ) de la droite des constantes  $\Delta_1$  de  $F_J$  (resp.  $F_K$ ) et donc

$$(H : F_K^* \rightarrow F_J) \Rightarrow (R : F_{K1}^* \rightarrow F_{J1})$$

L'information ordinaire induite sur  $J \times K$  par le tableau  $R_{JK}$  est en général différente de celle qui est induite par  $H_{JK}$ . Cette difficulté est particulièrement importante pour l'analyse des tableaux de dissimilarité et tient au biais d'estimation exprimés par DENIS (1983) sous la forme

$$E(\hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k) = \mu + \alpha_j + \beta_k + (\gamma_{j.} + \gamma_{.k} - \gamma_{..})$$

Ce biais d'estimation est illustré par BRADU (1975) qui construit le contre-exemple suivant, caractérisé par un modèle parfaitement additif puisque  $\mu = \alpha_j = \beta_k = 0$  sauf pour les cases (1,1), (2,2) et (2,3). Le tableau de droite présente les résidus

$$\begin{pmatrix} -36 & 0 & 0 & 0 & 0 & 0 \\ 0 & 24 & 36 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} -20 & 1 & -2 & 7 & 7 & 7 \\ 0 & 9 & 18 & -9 & -9 & -9 \\ 10 & -5 & -8 & 1 & 1 & 1 \\ 10 & -5 & -8 & 1 & 1 & 1 \end{pmatrix}$$

Cet exemple montre que l'existence de résidus non nuls n'implique pas l'existence de termes d'interaction. On préfère donc évaluer le potentiel d'interaction à l'aide des contrastes 2x2, appelés aussi tétrades par GABRIEL, PUTTER et WAX (1973). C'est le contraste unique d'interaction d'une sous-table 2x2 définie par 2 lignes et 2 colonnes quelconques de H.

|           |    |             |              |
|-----------|----|-------------|--------------|
|           |    | Facteur 2   |              |
|           |    | k           | k'           |
| Facteur 1 | j  | $\mu_{jk}$  | $\mu_{jk'}$  |
|           | j' | $\mu_{j'k}$ | $\mu_{j'k'}$ |

$$\Theta : J^2 \times K^2 \rightarrow \mathbb{R}$$

$$((j,j'),(k,k')) \mapsto \theta_{jj',kk'} = \mu_{jk} - \mu_{j'k} - \mu_{jk'} + \mu_{j'k'}$$

où nous avons noté classiquement  $\mu_{jk}$  l'espérance de H sous une hypothèse d'additivité

$$\mu_{jk} = \mu + \alpha_j + \beta_k$$

On vérifie alors de façon élémentaire les équivalences suivantes

**Lemme 1:**  $H_{JK}$  admet une décomposition additive si et seulement si l'une des trois propriétés équivalentes suivantes est vérifiée

- i)  $\mu_{jk} - \mu_{jk'} = \mu_{j'k} - \mu_{j'k'}$
- ii) Toute pondération positive  $P_j$  (resp.  $P_k$ ) des lignes (resp. des colonnes) de  $H_{JK}$  permet d'écrire l'espérance d'une case sous la forme

$$\mu_{jk} = \mu_{j.} + \mu_{.k} - \mu_{..}$$

où les points indiquent une moyenne pondérée sur l'indice correspondant

- iii) les écarts des valeurs appartenant à une ligne à leur moyenne ne dépend que de la ligne

L'application t étudiée au paragraphe §IV.7, définie sur l'espace vectoriel F et à valeurs dans  $F \wedge F$  par

$${}^t T_{\underline{x}} = \underline{y} = \text{Veca} (\underline{x}' \underline{1} - \underline{1}' \underline{x})$$

permet de définir ici deux applications linéaires

$$t_J : F_J \rightarrow F_J \wedge F_J \text{ et } t_K : F_K \rightarrow F_K \wedge F_K$$

Notons respectivement  $T_J$  et  $T_K$  les matrices associées à  $t_J$  et  $t_K$  pour les bases déduites de  $\mathcal{F}$ . L'opérateur  $t_J \otimes t_K$  définit alors une application linéaire sur  $(F_J \otimes F_K)$  à valeurs dans  $(F_J \wedge F_J) \otimes (F_K \wedge F_K)$  et caractérisée par

$$t_J \otimes t_K (\Gamma) = \Psi$$

où  $\Psi$  a pour matrice représentative dans  $\mathcal{M}_{J \times K}(\mathbb{R})$

$$\Psi_{JK} = {}^t T_J R_{JK} T_K$$

de terme général

$$\Psi_{jj',kk'} = r_{jk} - r_{j'k} - r_{jk'} + r_{j'k'}$$

**Lemme 2:** Si la matrice  $\Psi_{JK}$  définie par  $\Psi_{JK} = {}^t T_J R_{JK} T_K$  alors

$$\Psi_{JK} = -\frac{1}{2} {}^t T_J H_{JK} T_K$$

**Preuve :** On vérifie que  $\Psi_{jj',kk'}$  s'exprime sous la forme souhaitée en substituant dans son expression chaque terme  $r_{jk}$  par son expression en fonction des termes de  $H$  donnée par (52).  $\square$

**Définition 2:** On appelle *contraste empirique 2x2 sur  $J \times K$*  le terme général  $\Psi_{jj',kk'}$  de la matrice

$$\Psi_{JK} = -\frac{1}{2} {}^t T_J H_{JK} T_K$$

Par suite si  $\Psi_{JK}$  est une matrice de contrastes 2x2 et  $\Theta_{JK}$  la représentation matricielle de

$$\underline{\Theta} = t_J \otimes t_K (\underline{\mu})$$

alors DENIS et SPOJTVOLL (1981) montrent le résultat suivant

**Proposition 4:**

i)  $E(\Psi_{JK}) = -\frac{1}{2} {}^t T_J \Gamma_{JK} T_K$

ii)  $\text{Var}(\Psi_{jj',kk'}) = \sigma^2$

iii) Le coefficient d'aplatissement de  $\Psi_{jj',kk'}$  est

$$\mathcal{K}(\Psi_{jj',kk'}) = \frac{1}{4} \mathcal{K}(H_{jk}) = \frac{\gamma}{4}$$

DENIS (1983) remarque que déjà pour  $t=8$  et  $b=5$ , l'aplatissement de  $r_{jk}$  est le double de celui de  $\psi_{jj,kk}$  puisque

$$\mathcal{K}(r_{jk}) = \frac{\gamma((t-1)^3+1)((b-1)^3+1)}{t^2 b^2 (t-1)(b-1)}$$

et donc les contrastes  $2 \times 2$  sont plus appropriés que les résidus pour mener une analyse classique du potentiel d'interaction décrit par les données  $H_{JK}$ .

Dans le cas où  $J \neq K$ , on obtient une forme algébrique de la correspondance entre observations  $(h_{jk})$  et contrastes  $2 \times 2$ . Lorsque  $J=K=I$ , la matrice  $\psi = \psi_{JK}$  présente de nombreuses propriétés particulières. Tout d'abord le lemme 2 se particularise comme suit

**Corollaire :**

i) la matrice  $\psi = {}^t W T$  déduite de la matrice  $W = -\frac{1}{2} (Q_1 H {}^t Q_1)$  vérifie

$$\psi = -\frac{1}{2} {}^t T H T$$

Elle a donc pour terme général

$$\psi_{jj,kk} = -\frac{1}{2} (h_{jk} - h_{j'k} - h_{jk'} + h_{j'k'})$$

ii) Si  $W \in S_n^+$  alors  $\psi \in S_n^+$

**Preuve :** Seule ii) nécessite une démonstration. Celle-ci résulte de l'expression de  $\psi$  en fonction de  $H$  et  $T$ . En effet pour tout  $\underline{v}$  de  $(F \wedge F)^*$

$${}^t \underline{v} \psi \underline{v} = -\frac{1}{2} {}^t (T \underline{v}) H T \underline{v}$$

où  ${}^t (T \underline{v}) = \underline{u} \in F_1^*$  et donc  $\sum u_i = 0$ . Par suite comme  $W$  est un élément de  $S_n^+$ ,  $H$  est conditionnellement négative et donc  ${}^t \underline{v} \psi \underline{v} \geq 0$ .  $\square$

Considérons maintenant le cas où  $J$  et  $K$  forment une partition de  $I$ . Soit  $Z$  une représentation Euclidienne de  $I$  dans un espace affine  $E$ . La matrice  $Z$  peut être décomposée en  $(X : Y)$  où  $X \in \mathcal{M}_{p,t}(\mathbb{R})$  et  $Y \in \mathcal{M}_{p,b}(\mathbb{R})$ . Nous considérons alors la décomposition en somme directe de l'espace vectoriel libre  $F$  :

$$F = F_J \oplus F_K$$

avec  $\dim(F_J) = t$  et  $\dim(F_K) = b$  et l'on note  $\mathcal{F}_J$  et  $\mathcal{F}_K$  les bases respectives de  $F_J$  et  $F_K$  choisies de sorte que

$$\mathcal{F} = \mathcal{F}_J \cup \mathcal{F}_K$$

Remarquons qu'alors,  $W_Z$  est décomposable en blocs

$$W_Z = {}^t Z M Z = \begin{pmatrix} {}^t X M X & {}^t X M Y \\ {}^t Y M X & {}^t Y M Y \end{pmatrix} = \begin{pmatrix} W_{JJ} & W_{JK} \\ W_{KJ} & W_{KK} \end{pmatrix}$$

et les relations suivantes sont satisfaites pour toute dyade  $((j,j'), (k,k'))$  de  $J^2 \times K^2$

**Proposition 5:**

$$\begin{aligned} \Psi_{jj',kk'} &= {}^t (f_j^* - f_{j'}^*) W_{12} (f_k^* - f_{k'}^*) \\ &= {}^t (x_j - x_{j'}) M (y_k - y_{k'}) \end{aligned}$$

**Preuve :** Cela résulte de la relation

$$\Psi_{JK} = {}^t T_J W_{JK} T_K = -\frac{1}{2} {}^t T_J Q_J H_{JK} Q_K T_K$$

dans laquelle nous avons noté  $W_{JK}$  la matrice des produits scalaires  ${}^t x_j M y_k$  où  $x_j$  et  $y_k$  sont choisis centrés.  $\square$

Les contrastes  $2 \times 2$  sont en fait aisés à modéliser. Si  $j \neq j'$  deux éléments de  $J$ , notons en effet  $a_{jj'}$  le vecteur de  $F_J^*$  défini par

$$a_{jj'} = \frac{1}{\sqrt{2}} (f_j^* - f_{j'}^*)$$

$a_{jj'}$  est normé à 1 et définissons de même le vecteur  $b_{kk'}$  de  $F_K^*$  pour  $k \neq k'$  deux éléments de  $K$ . Alors

$$\Psi_{jj',kk'} = {}^t (a_{jj'} \otimes b_{kk'}) h$$

Cette relation entre  $\psi$  et  $h$  permet de vérifier directement la relation suivante

$$\text{Cov}(\Psi_{jj',kk'}, \Psi_{mm',nn'}) = \sigma^2 ({}^t a_{jj'} a_{mm'}) ({}^t b_{kk'} b_{nn'})$$

Par suite la corrélations entre contrastes  $2 \times 2$  est nulle pour de nombreuses paires de dyades. On montre d'ailleurs que le sous-espace vectoriel de  $(F_J \wedge F_J) \otimes (F_K \wedge F_K)$  engendré par les contrastes  $2 \times 2$  est de dimension  $(J-1)(K-1)$ . Elle est donc  $JK/4$  fois plus petit que le nombre de contrastes distincts. Nous donnons de plus sans démonstration le résultat suivant que l'on vérifie par substitution

**Propriété :** *Il existe une bijection entre contrastes  $2 \times 2$  et résidus, définie par :*

$$\begin{aligned} \Psi_{jj',kk'} &= r_{jk} - r_{j'k} - r_{jk'} + r_{j'k'} \\ r_{jk} &= \frac{1}{JK} \sum_{j'} \sum_{k'} \Psi_{jj',kk'} \end{aligned}$$

Par suite, pour la structure statistique gaussienne définie par MANDEL (1961), les éléments de  $\psi$  forment une statistique exhaustive de l'interaction.

Enfin on vérifie la relation suivante

**Propriété :** La somme des carrés des écarts dûs à l'interaction est donnée par

$$SCE_{Int} = \sum_{j < j'} \sum_{k < k'} (\psi_{jj',kk'})^2$$

et

$$\hat{\sigma}^2 = \frac{SCE_{Int}}{b^* t^*} \quad \text{où} \quad b^* = \frac{b(b-1)}{2} \quad \text{et} \quad t^* = \frac{t(t-1)}{2}$$

**Preuve :** Le résidu  $r_{jk}$  est le terme général de la matrice  $R_{JK} = -\frac{1}{2} Q_J H_{JK} {}^t Q_K$ . Par suite puisque  ${}^t T_J T_J = Q_J$  et que  ${}^t T_K T_K = Q_K$

$$\begin{aligned} SCE_{Int} &= \sum_j \sum_{k'} (r_{jk})^2 = \text{Trace}(R_{JK} {}^t R_{JK}) = \text{Trace}\left(\frac{1}{4} Q_J H_{JK} {}^t Q_K H_{JK}\right) \\ &= \text{Trace}\left(\frac{1}{4} T_J {}^t T_J H_{JK} T_K {}^t T_K H_{JK}\right) = \text{Trace}\left(\left(-\frac{1}{2} {}^t T_J H_{JK} T_K\right) \left(-\frac{1}{2} {}^t T_J H_{JK} T_K\right)\right) \\ &= \text{Trace}(\psi_{JK} {}^t \psi_{JK}) \quad \parallel \end{aligned}$$

**Remarque :** Du fait de la définition du produit scalaire,  $SCE_{Int}$  s'écrit aussi :

$$SCE_{Int} = \frac{1}{4} \| \underline{h} \|_{Q_K \otimes Q_J}^2$$

En pratique les contrastes 2x2 représentent la fonction la plus simple des observations faisant intervenir uniquement les paramètres d'interaction  $\gamma_{jk}$ . D'autre part, l'espérance mathématique de certains contrastes 2x2 peut être nulle même si aucun  $\gamma_{jk}$  n'est nul. Dans ce cas, JOHNSON et GRAYBILL (1972) construisent à partir de ces contrastes 2x2 d'espérance nulle un estimateur sans biais de la variance résiduelle. De plus, on montre par application des résultats classiques à la décomposition de  $F \otimes F$  que

**Lemme 3:** Toute fonction des contrastes 2x2 est indépendante

i) des estimateurs des effets des traitements  $\hat{\alpha}_j = h_{j.} - h_{..}$

ii) des estimateurs des effets des blocs  $\hat{\beta}_k = h_{.k} - h_{..}$

et donc indépendante

iii) de la somme des carrés des écarts dûs aux traitements  $\sum_j (\hat{\alpha}_j)^2$

iv) de la somme des carrés des écarts dûs aux blocs  $\sum_k (\hat{\beta}_k)^2$

**Remarque :** Bien que la notion de contraste n'apparaisse pas dans la littérature consacrée à l'analyse des données, ces propriétés des contrastes 2x2 justifient en fait un grand nombre de modèles développés en Analyse des Données de Préférence (cf. d'AUBIGNY (1981), BECHTEL (1974), BOCK et JONES (1966)). En effet considérons la mesure de l'utilité comparée des objets  $k$  et  $k'$  par le juge  $j$  évaluée par

$$u_{kk'}^j = \mu_{jk} - \mu_{jk'}$$

alors

i)  $u_{kk'}^j = \beta_k - \beta_{k'}$  en l'absence d'interaction entre  $J$  et  $K$

ii)  $u_{kk'}^j = (\beta_k - \beta_{k'}) + (\gamma_{jk} - \gamma_{jk'})$  sinon

Ce deuxième modèle est étroitement lié aux modèles traités par BECHTEL (1974) et analysés ensuite par cet auteur en termes de Déploiement Multidimensionnel (cf. aussi d'AUBIGNY (1981)). De plus, en Analyse des Préférences, le contraste 2x2  $\psi_{jj',kk'}$  est nul si et seulement si

$$h_{jk} - h_{jk'} = h_{j'k} - h_{j'k'}$$

Les juges  $j$  et  $j'$  expriment alors les mêmes jugements de préférence sur  $k$  et  $k'$ .

De nombreux tests d'absence d'interaction ont été proposés dans la littérature statistique, qui utilisent la notion de contrastes 2x2. En particulier JAMES (1961) teste l'hypothèse  $H_0$  d'additivité contre l'hypothèse ordinale

$$(E(\psi_{jj',kk'}) > 0) \Leftrightarrow (j > j', k > k')$$

Ces modèles sont étudiés plus en avant par SCHAAFSMA (1966), BARLOW et al. (1972), HIROTSU (1978) et DENIS (1983) qui donne des références bibliographiques complémentaires.

La relation entre contrastes 2x2 et résidus explicitée dans ce paragraphe, conduit à rechercher les bases de  $E$  qui réalisent une décomposition additive de

$$SCE_{Int} = \text{Trace}(R_{JK} {}^t R_{JK}) = \text{Trace}(\Psi_{JK} {}^t \Psi_{JK})$$

De plus puisque  $R_{JK}$  est un tableau centré en lignes et en colonnes, on peut mener une analyse spectrale de  $R_{JK}$  et rechercher ses éléments singuliers sous les contraintes (C)

$$\begin{aligned} \text{i) } \sum_j \alpha_j &= 0 & \sum_j \alpha_j^2 &= 1 \\ \text{ii) } \sum_k \beta_k &= 0 & \sum_k \beta_k^2 &= 1 \end{aligned}$$

$\underline{\alpha}$  (resp.  $\underline{\beta}$ ) est alors un contraste de longueur 1 de  $F_J$  (resp.  $F_K$ ) et par application directe du théorème d'ECKART et YOUNG (1936) on fait le lien avec les pratiques d'Analyse en Coordonnées Principales en posant

$$V_R = {}^t R_{JK} R_{JK} \text{ et } W_R = R_{JK} {}^t R_{JK}$$

**Lemme 4** (JOHNSON et GRAYBILL (1972)) :  $\text{Max}\{({}^t \underline{u} H_{JK} \underline{v})^2; \underline{u}, \underline{v} \text{ vérifiant (C)}\}$  est atteint lorsque

- i)  $\hat{\lambda}_1$  est valeur propre maximale de  $V_R$  (et donc de  $W_R$ )
- ii)  $\underline{u}_1$  (resp.  $\underline{v}_1$ ) est un vecteur propre normé à 1 de  $V_R$  (resp.  $W_R$ )

De plus compte tenu des contraintes (C), on a

$$({}^t \underline{u} H_{JK} \underline{v})^2 = ({}^t \underline{u} Q_J H_{JK} {}^t Q_K \underline{v})^2 = 4 ({}^t \underline{u} R_{JK} \underline{v})^2$$

et donc  $\underline{u}$  et  $\underline{v}$  sont donc bien des vecteurs propres communs au spectre de  $H_{JK}$  et  $R_{JK}$ .

Enfin ces deux auteurs montrent le résultat final, valable dans le cadre du modèle de MANDEL :

**Théorème 1:** Le modèle de JOHNSON et GRAYBILL admet pour estimateurs du maximum de vraisemblance sous les contraintes (C)

- i)  $\hat{\mu} = h_{..}$
- ii)  $\hat{\alpha}_j = h_{.j} - h_{..}$
- iii)  $\hat{\beta}_k = h_{.k} - h_{..}$
- iv)  $\hat{\lambda}_1, \hat{u}$  et  $\hat{v}$  fixés par le lemme 4 estiment la partie multiplicative du modèle et alors

$$\hat{\sigma} = \frac{\sum_{s=2}^b \hat{\lambda}_s}{bt}$$

**Remarque :** L'introduction de pondérations sur J et K conduit à un résultat analogue lorsque  $F_J$  et  $F_K$  sont munis respectivement de géométries fixées par les matrices de poids  $D_{P_J}$  et  $D_{P_K}$  et donc les contraintes (C) sont rempacées par celles de  $D_P$ -centrage et  $D_P$ -normalisation. De plus on a alors

$$V_R = {}^t R_{JK} D_{P_J} R_{JK} \text{ et } W_R = R_{JK} D_{P_K} {}^t R_{JK}$$

et l'on mène l'analyse spectrale de  $V_R D_{P_J}$  et  $W_R D_{P_K}$  respectivement.

Les résultats précédents justifient une interprétation des axes principaux de l'analyse du triplet  $(R_{JK}, D_{PJ}, D_{PK})$  en termes de contrastes optimaux orthogonaux, puisque rechercher plus d'une solution équivaut à ajuster les paramètres du modèle de MANDEL et nous avons montré que ces contrastes décrivent au mieux les contrastes  $2 \times 2$ .

Lorsque  $J \neq K$ , nous retrouvons la solution du problème de Déploiement Multidimensionnel, proposé par SCHÖNEMAN (1972) puisque l'on reconnaît en  $R_{JK}$  la matrice doublement centrée que cet auteur propose de soumettre à une Analyse en Composantes Principales. Par suite, on peut interpréter l'approche de SCHÖNEMAN comme la double recherche de contrastes optimaux permettant de comparer les modalités de  $J$  (resp.  $K$ ) entre elles d'une part et à celles de  $K$  (resp.  $J$ ) d'autre part, à condition d'adopter la normalisation fixée par les contraintes (C). Notons de plus qu'alors les relations de dualité entre lignes et colonnes du tableau justifient une interprétation angulaire des valeurs  $r_{jk}$ . GREENACRE (1978) observe à juste titre qu'alors aucun critère d'approximation forte des éléments de  $H_{ij}$  n'est fourni par cette méthode.

Lorsque  $J=K=I$ , la recherche de contrastes optimaux éclaire aussi les méthodes usuelles d'analyse des tableaux de dissimilarité. En Analyse en Coordonnées Principales, on considère la matrice produit d'Hadamard d'une matrice de dissimilarité  $\Delta(Z)$  par elle-même

$$H = (\delta_{ij}^2(Z))$$

et  $R$  est alors la matrice de produits scalaires  $W = \alpha_1^+(H)$  dont l'analyse spectrale fournit les coordonnées principales. Celles-ci sont centrées et si donc on choisit de les normer à 1, les résultats précédents leur donnent le statut statistique de contrastes optimaux. La méthode peut alors être interprétée comme une étude du potentiel d'interaction défini par le tableau de données  $\Delta$ .

De plus DENIS et SPOJTVOLL (1981) et DENIS (1983) utilisent l'information apportée par les contrastes  $2 \times 2$ , pour repérer les valeurs suspectes, qui rendent délicat le recours aux résidus  $r_{jk}$  pour l'analyse du tableau  $H$ . Ils utilisent les résultats présentés plus haut sur les moments des contrastes  $2 \times 2$  pour repérer graphiquement les valeurs suspectes à l'aide d'une technique de "Half normal plots". Interprétée dans le cadre de l'Analyse en Coordonnées Principales cette méthode semble particulièrement adaptée à l'évaluation des résultats.

D'autre part lorsque  $\delta(Z)$  est une dissimilarité Euclidienne, on vérifie les propriétés suivantes qui sont de démonstration immédiate puisque  $\psi = {}^t W_Z T$  :

**Propriété :** La matrice  $\psi$  définit une forme bilinéaire symétrique sur  $(F \wedge F)^*$  de terme général

$$i) \psi_{ij,kl} = \begin{cases} \delta_{ij}^2(Z) & \text{si } (i,j) = (k,l) \\ -\frac{1}{2}(\delta_{ik}^2(Z) - \delta_{if}^2(Z) - \delta_{ji}^2(Z) + \delta_{jf}^2(Z)) & \text{sinon} \end{cases}$$

$$ii) \psi_{ij,kj} = {}^t(\underline{f}_i^* - \underline{f}_j^*) W_Z (\underline{f}_k^* - \underline{f}_l^*) \\ = {}^t(\underline{z}_i - \underline{z}_j) M (\underline{z}_k - \underline{z}_l)$$

Par suite la dimension de l'espace de représentation de  $I$  est directement liée au rang de  $\psi$  puisque  $\psi_{ij,kj}$  apparaît dans ii) comme le produit scalaire mesurant l'angle formé par les deux vecteurs  $\overrightarrow{P_i P_j}$  et  $\overrightarrow{P_k P_l}$ . On doit noter de plus que cette propriété des quadruplets de  $I$  est liée aux propriétés retenues dans la définition axiomatique d'une géométrie Euclidienne.

De façon complémentaire, toute axiomatique des espaces Euclidiens fondée sur l'existence de triangles rectangles, s'exprime elle aussi en termes de contrastes 2x2. En effet, si  $i=k$ , on obtient

$$\psi_{ij,il} = -\frac{1}{2}(-\delta_{il}^2(Z) - \delta_{ji}^2(Z) + \delta_{jl}^2(Z))$$

et donc

$$\psi_{ij,il} = 0 \quad \Leftrightarrow \quad \delta_{jl}^2(Z) = \delta_{il}^2(Z) + \delta_{ji}^2(Z)$$

Il est remarquable qu'une telle identité est vérifiée aussi par les contrastes 2x2 définis par  $\delta$  et non par  $\delta^{*2}$ . Lorsqu'on définit  $H$  comme étant égale à  $\Delta$  et non  $\Delta^{*2}$ , on retrouve la condition d'intermédiarité dont nous avons noté le rôle important dans l'élaboration d'axiomatics des espaces Euclidiens.

De plus les propriétés d'existence d'un plongement Euclidien d'un espace quasi-métrique  $(I, \delta)$  s'exprime en termes de contrastes 2x2. En effet, nous pouvons compléter les résultats de Schoenberg à l'aide du théorème suivant :

**Théorème 2:** Soit  $\psi$  la matrice des contrastes 2x2 associée à la matrice  $H = \Delta * \Delta$  où  $(I, \delta)$  est un espace quasi-métrique de cardinal fini et  $\delta$  est une dissimilarité propre. Alors  $(I, \delta)$  admet un plongement Euclidien si et seulement si  $\psi$  est une matrice à termes positifs.

**Preuve :** Nous avons montré qu'un tel plongement Euclidien existe si et seulement si la fonction

$$t \rightarrow \exp\{-\lambda t^2\}$$

appartient à RPD(I) pour tout  $\lambda$  réel positif, c'est à dire si et seulement si la matrice W de terme général

$$w_{ij} = \exp\{-\lambda \delta_{ij}^2\}$$

est définie positive. Ceci est réalisé si et seulement si toute sous-table associée à quatre indices  $i, j, k$  et  $l$  distincts vérifie

$$0 \leq \det \begin{pmatrix} w_{ik} & w_{il} \\ w_{jk} & w_{jl} \end{pmatrix} = w_{ik} w_{jl} - w_{ij} w_{jk}$$

Puisque  $\delta$  est une dissimilarité propre, ceci est équivalent à

$$\frac{w_{ik} w_{jl}}{w_{il} w_{jk}} \geq 1$$

On reconnaît ainsi une condition classique en Statistique et portant sur le birapport  $(i,j,k,l)$  qui s'écrit après passage au logarithme

$$\text{Log}(w_{ik}) + \text{Log}(w_{jl}) - \text{Log}(w_{il}) - \text{Log}(w_{jk}) \geq 0$$

Par suite le déterminant est positif si et seulement si

$$-\lambda(\delta_{ik}^2 + \delta_{jl}^2 - \delta_{il}^2 - \delta_{jk}^2) \geq 0$$

et donc, puisque  $\lambda \geq 0$  cela revient à dire que la matrice  $\psi$  vérifie la condition  $\psi_{ij,kl} \geq 0$ .  $\square$

Cette propriété nous semble intéressante non seulement par la simplicité du résultat, mais par le fait qu'elle utilise une relation étroite entre contrastes 2x2 et birapports qui ne semble pas avoir été notée auparavant, malgré l'importance de la notion de birapport en géométrie projective d'une part, et en statistique mathématique d'autre part. Les birapports interviennent en particulier dans l'étude des tables de contingence à marges ordonnées (ou non) sous le vocable anglais de odds ratios et ont été proposés par YULE (1912), cf. GOODMAN (1979 a, b, 1981, 1983), GOODMAN et KRUSKAL (1979), HABERMAN (1974, 1978, 1979), PLACKETT (1981) et AGRESTI (1984). Elle montre à nouveau les relations étroites qui lient les méthodes d'analyse de tableaux de dissimilarité et les modélisations Log-linéaires des tables de contingence symétriques (ici en effet W vérifie  $w_{jk} = w_{kj} \forall j,k \in [n]$ ).

Du point de vue de l'Analyse Linéaire des Données, ce théorème donne une condition simple d'existence de la solution de la méthode d'Analyse Logarithmique, proposée par AITCHISON (1983, 1986) pour l'analyse des données de compositions, et par KAZMIERCZAK (1985a, b) pour l'analyse des tables de contingence binaires. Il est de ce point de vue nécessaire de noter que le problème de semi-définie positivité de la matrice analysée par GOODMAN (1983) se pose aussi pour les modèles de décomposition des tables de contingence à marges ordonnées proposés par cet auteur, lorsque la construction d'une représentation multidimensionnelle est nécessaire. De façon étonnante, la discussion de ce problème est absente de la comparaison de ces modèles avec les solutions issues d'une Analyse Canonique.

#### V.1.2.2. Choix d'une origine de E et contrastes 2x2

La démarche précédente traite le problème dans un cadre vectoriel. Etant donnée une pondération  $P_1$ , l'opérateur  $Q_P$  a pour rôle de centrer le tableau H par rapport au barycentre des points massiques  $\{(P_i, p_i) ; i \in [n]\}$ . Ainsi donc la structure affine de l'espace de représentation E intervient dans ces calculs. La pondération positive  $P_1$  ou plus généralement réelle  $\gamma$  intervient dans le calcul de  $\alpha_\gamma^+(H)$  puisque

$$W = -\frac{1}{2} (Q_\gamma H {}^t Q_\gamma) \quad \text{où} \quad Q_\gamma = I - \mathbb{1} {}^t \mathbb{1} D_\gamma$$

Par suite si  $\gamma^*$  est une deuxième pondération de I, correspondant à une autre origine  $P_0^*$  de E on obtient :

$$W^* = -\frac{1}{2} (Q_{\gamma^*} H {}^t Q_{\gamma^*})$$

Mais puisque  $Q_\gamma$  est le projecteur  $D_\gamma$ -orthogonal sur le sous-espace vectoriel  $F_\gamma$  de F supplémentaire  $D_\gamma$ -orthogonal de  $\Delta_1$  dans F, on vérifie

$$\begin{aligned} Q_\gamma Q_{\gamma^*} &= Q_\gamma (I - \mathbb{1} {}^t \mathbb{1} D_{\gamma^*}) \\ &= Q_\gamma - (Q_\gamma \mathbb{1}) {}^t \mathbb{1} D_{\gamma^*} = Q_\gamma \end{aligned}$$

et donc

$$W = Q_\gamma W^* {}^t Q_\gamma = -\frac{1}{2} (Q_\gamma H {}^t Q_\gamma)$$

De sorte que tout changement d'origine dans E se traduit de façon simple dans  $\mathcal{Y}_n^+$  à l'aide d'une opération de projection. Dans  $F \otimes F$  on note que

$$(W = Q_\gamma W^* {}^t Q_\gamma) \Leftrightarrow (\underline{w} = Q_\gamma \otimes Q_\gamma \underline{w}^*)$$

de sorte que le changement d'origine se traduit dans  $F \otimes F$  par une projection sur un sous-espace  $D_\gamma \otimes D_\gamma$ -orthogonal à la droite des constantes engendrée par le vecteur  $\underline{1} \otimes \underline{1}$ .

Choisissons pour origine de  $E$  le point  $P_n$ . Alors comme plus haut on regarde  $F$  comme un espace affine de direction  $\vec{F}$ . Le référentiel  $(O, (\vec{f}_i; i \in [n]))$  permet de représenter les éléments de  $I$  par des points  $N_i = O + \vec{f}_i$  de  $F$ . Puisque  $\dim(F) = n = \text{card}(I)$ , les points  $N_i$  engendrent un sous-espace affine  $F_1$  de  $F$  de direction  $\vec{F}_1$ . Se donner une pondération telle que  $\sum \gamma_i = 1$  équivaut à se donner un point de  $F_1$ , c'est à dire un barycentre  $G$  des points  $N_i$ , défini par ses coordonnées  $\gamma_i$  dans le référentiel. Dans ce qui précède, nous nous sommes placés dans  $F$ . Alors nous avons associé à tout élément  $H$  de  $\mathcal{Y}_H$  l'élément  $W = \alpha_\gamma(H)$  de  $\mathcal{Y}_\gamma^+$ .  $W$  et  $H$  appartiennent chacun à  $\mathcal{Y}_n$  dans cette construction. Lorsqu'on se place dans  $\vec{F}_1$ , repéré par rapport à la base  $(\vec{N}_n \vec{N}_i; i \in [n-1])$ , on est amené à construire l'application

$$\begin{aligned} \alpha_\gamma^n : \mathcal{Y}_H &\rightarrow \mathcal{Y}_{n-1}^+ \\ H &\mapsto \alpha_\gamma^n(H) = W^n \end{aligned}$$

où  $W^n$  est la matrice d'ordre  $(n-1)$  de terme général

$$w_{ij}^n = -\frac{1}{2}(h_{ij} - h_{in} - h_{jn})$$

Lorsque  $\delta$  est Euclidienne, le choix de la matrice  $H = \Delta(Z) * \Delta(Z)$  permet ainsi de construire une matrice  $W^n$  dont l'analyse spectrale fournit un plongement Euclidien de  $I$ . Si le choix d'un origine n'influe pas sur la détermination du support de  $(P_i; i \in [n])$ , cette propriété d'invariance ne signifie pas que l'analyse de  $W$  et celle de  $W^n$  fournissent la même base de description de  $E$ . En général cela est faux. De plus observons que

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n w_{ij}^k &= -\frac{1}{2n} \sum_{k=1}^n (h_{ij} - h_{ik} - h_{jk}) \\ &= -\frac{1}{2}(h_{ij} - h_{i.} - h_{.j}) \end{aligned}$$

et donc  $W$  peut s'écrire

$$W = \bar{W} - \frac{1}{2} h_{..} \underline{1} \underline{1} \quad \text{où} \quad \bar{W} = \frac{1}{n} \sum_{k=1}^n W^k$$

L'analyse de  $W$  et celle de  $\bar{W}$  réfèrent donc à la même 2-forme affine, puisque pour tout  $i$  et  $j$  de  $[n]$

$$\begin{aligned} w(\vec{f}_i, \vec{f}_j) &= w(\overrightarrow{ON}_i, \overrightarrow{ON}_j) = w(\overrightarrow{OG}, \overrightarrow{OG}) + w(\overrightarrow{OG}, \overrightarrow{GN}_i) + w(\overrightarrow{OG}, \overrightarrow{GN}_j) + w(\overrightarrow{GN}_i, \overrightarrow{GN}_j) \\ &= w(\overrightarrow{OG}, \overrightarrow{OG}) + w(\overrightarrow{GN}_i, \overrightarrow{GN}_j) \end{aligned}$$

Par suite si l'on pose

$$w(\vec{OG}, \vec{OG}) = \frac{1}{2} h_{..}$$

on obtient

$$w(\vec{f}_i, \vec{f}_j) = \bar{w}(\vec{GN}_i, \vec{GN}_j)$$

et donc  $w$  et  $\bar{w}$  sont associés à la même forme bilinéaire sur  $\vec{F}_1$ . De plus  $W$  vérifie par construction

$$\underline{Q} = W\underline{\gamma} = W D_{\gamma} \underline{1}$$

et donc

$$\underline{Q} = \bar{W} D_{\gamma} \underline{1} - \frac{1}{2} h_{..} \underline{1}' D_{\gamma} \underline{1} \Leftrightarrow \bar{W} D_{\gamma} \underline{1} = \frac{1}{2} h_{..} \underline{1}$$

$\underline{1}$  est vecteur propre de  $\bar{W} D_{\gamma}$  associé à la valeur propre  $\frac{1}{2} h_{..}$ .

**Proposition 6:** Les contrastes 2x2 de la forme  $\psi_{ij, kn}$  sont déduits de la forme bilinéaire  $w^n$  par la relation

$$\psi_{ij, kn} = w_{ik}^n - w_{jk}^n$$

**Preuve :**

$$\begin{aligned} w_{ik}^n - w_{jk}^n &= -\frac{1}{2} (h_{ik} - h_{in} - h_{kn}) + \frac{1}{2} (h_{jk} - h_{jn} - h_{kn}) \\ &= -\frac{1}{2} (h_{ik} - h_{in} - h_{jk} + h_{jn}) = \psi_{ij, kn} \quad \square \end{aligned}$$

Par suite, on peut regarder  $\Psi$  comme une matrice qui intègre l'information contenue dans l'ensemble des opérateurs  $w^n$ .

### V.1.3. Les méthodes concurrentes de l'Analyse en Coordonnées Principales

La méthodologie classique d'Analyse du Triple  $(I, \delta, P_I)$  où  $\delta$  est une dissimilarité Euclidienne est limitée aux techniques de décomposition spectrale de l'opérateur  $W_Z D_p$ . Pourtant l'analogie avec l'Analyse en Composantes Principales montre que ce parti pris est très limitatif. La recherche d'axes de coordonnées peut en effet suivre un grand nombre de voies alternatives. Ainsi peut on adapter à l'étude de  $(I, \delta, P_I)$  la méthode proposée par OTTESTAD (1975) qui propose de construire une décomposition de Gram-

Schmidt de  $W_Z$ . Le choix d'une direction initiale introduit cependant une part d'arbitraire supplémentaire dans l'étude, qu'il est difficile de justifier en analyse des tableaux de dissimilarités. Cependant l'analyse des tableaux  $W^n$  relève d'un tel arbitraire. L'interprétation de l'Analyse en Composantes Principales comme une méthode de projection en drapeau (ou Projection Pursuit) particulière ouvre la voie à de nombreuses propositions alternatives, concurrentes de la méthode de OTTESTAD. Notre étude est ici marquée par le souci d'explicitier l'analogie forte entre les méthodes de l'Analyse Euclidienne du Triple  $(I, \delta, P_I)$  et les méthodes d'Analyse Factorielle du Triple  $(J, R, P_J)$  où  $R$  est la matrice de covariance ou de corrélation calculée sur les variables éléments de  $J$ . Appelons Q-Analyse une méthode d'analyse portant sur les unités statistiques, par opposition à une R-Analyse qui porte sur les variables. La relation entre  $\mathcal{D}_E^{*2}(I)$  et  $\mathcal{Y}_\gamma^*$  établie par les applications  $\alpha_\gamma^*$  et  $\beta_\gamma^*$  montre une communauté de problèmes entre les méthodes d'Analyse du Triple  $(I, \delta(Z), P_I)$  et de Q-Analyse factorielle du triple  $(I, w_Z, P_I)$  qui de façon étonnante, n'a pas été exploitée par les spécialistes de chacun de ces domaines de recherche.

McDONALD (1968, 1975) ordonne les méthodes d'analyse linéaire des données en deux classes. La première est formée par les méthodes qui construisent des composantes. Elles sont fondées sur la recherche d'une décomposition d'un vecteur aléatoire  $p$ -dimensionnel  $Z$  en deux composantes

$$\underline{Z} = \underline{Y} + \underline{\epsilon}$$

satisfaisant aux deux contraintes

$$\begin{aligned} \underline{Y} &= \underline{P}\underline{Z} \\ E(\underline{Y}'\underline{\epsilon}) &= 0 \end{aligned}$$

Il résulte de cette décomposition que  $P$  est un projecteur et pour toute factorisation  $U^tV$  du projecteur  $P$ , on peut écrire

$$\begin{aligned} \underline{Z} &= \underline{U}\underline{c} + \underline{r} \\ &= \underline{U}^t\underline{V}\underline{Z} + (\underline{I} - \underline{U}^t\underline{V})\underline{Z} \end{aligned}$$

La deuxième classe regroupe ce qu'il est convenu d'appeler les méthodes d'Analyse Factorielle. Celles-ci supposent l'existence d'un vecteur  $r$ -dimensionnel ( $r \leq p$ ) non observable ou latent, noté  $\underline{X}$  dont les composantes sont appelées les facteurs communs.  $\underline{X}$  est supposé tel que la régression  $\hat{\underline{Z}} = \underline{F}\underline{X}$  de  $\underline{Z}$  en  $\underline{X}$  a pour résidu un vecteur  $\underline{\epsilon} = \underline{Z} - \hat{\underline{Z}}$  dont la matrice de covariance respecte la contrainte de structure

$$E(\underline{\epsilon}'\underline{\epsilon}) = D_U^2$$

La matrice diagonale positive  $D_U^2$  permet alors d'écrire le modèle de structure de covariance suivant, où nous supposons pour plus de simplicité les facteurs communs orthogonaux :

$$R_Z = E(\underline{Z}^t \underline{Z}) = F^t F + D_U^2 \quad (54)$$

$D_U^2$  s'interprète comme la matrice de covariance des facteurs spécifiques attachés à chaque variable initiale. Les termes de  $D_U^2$  sont alors appelés les unicités des variables initiales formant le tableau Z.

La distinction entre ces deux classes de modèles descriptifs est fondamentale pour la compréhension des méthodes d'analyse linéaire d'un tableau de données de profil. De plus elle montre une analogie formelle forte avec la distinction faite dans notre exposé entre les méthodes d'Analyse en Coordonnées (Principales ou autres, cf. infra) et les méthodes de Codage Multidimensionnel. Nous illustrons cette distinction en développant tout d'abord une méthode d'Analyse Factorielle du Triple  $(I, \delta(Z), P_I)$ . Lorsqu'on tient compte de cette distinction, l'originalité du Codage Multidimensionnel du Triple  $(I, \delta(Z), P_I)$  demeure cependant, du fait de la pratique usuelle consistant à traiter des dissimilarités non nécessairement Euclidiennes, alors que la littérature consacrée à l'Analyse Factorielle Classique suppose toujours que la matrice analysée R est gramienne. De plus nous montrons plus bas que les méthodes de Codage Multidimensionnel traitent un problème plus général pour une deuxième raison. En effet, on analyse alors le triple  $(I, \delta, Q_{I \times I})$  où  $Q_{I \times I}$  est une pondération positive de forme générale, définie sur  $I \times I$ . Cette caractéristique change singulièrement la nature mathématique des problèmes rencontrés.

L'équation (54) se traduit dans  $\mathcal{Y}_H$  par la relation suivante, exprimée en fonction de la matrice  $H_Z = \beta_1^*(R_Z)$  de terme général  $\delta_{ij}^2(Z)$  et la matrice  $H_F = \beta_1(F^t F)$  de terme général  $d_{ij}^2(F)$ .

**Proposition 7:** L'équation structurelle (54) s'écrit de façon équivalente

$$h_{ij}(Z) = \begin{cases} h_{ij}(F) + u_i^2 + u_j^2 & \text{si } i \neq j \\ h_{ii}(F) & \text{si } i = j \end{cases}$$

**Preuve :**

$$\begin{aligned} H_Z &= \beta_1^*(R_Z) = (R_Z^* I) \underline{1} \underline{1} + \underline{1} \underline{1} (R_Z^* I) - 2R_Z \\ &= [(F^t F^* I) \underline{1} \underline{1} + \underline{1} \underline{1} (F^t F^* I) - 2F^t F] + [D_U^2 \underline{1} \underline{1} + \underline{1} \underline{1} D_U^2 - 2D_U^2] \\ &= H_F + D_U^2 \underline{1} \underline{1} + \underline{1} \underline{1} D_U^2 - 2D_U^2 \end{aligned}$$

qui est l'expression matricielle du résultat cherché. . ¶

Cette décomposition de  $H_Z$  comme somme de deux matrices

$$H_Z = H_F + H_U$$

où

$$H_U = \underline{u}^t \underline{1} + \underline{1}^t \underline{u} - 2D_U^2$$

s'interprète à l'aide de la décomposition de  $\mathcal{Y}_H$  étudiée au paragraphe §IV.5 en

$$\mathcal{Y}_H = \mathcal{Y}_H \cap \mathcal{Y}_1 \oplus \mathcal{R}$$

puisque par construction,  $H(U)$  est un élément de  $\mathcal{R}$ .

Lorsque  $\delta$  est une fonction de dissimilarité Euclidienne, le choix  $h_{ij}(Z) = \delta_{ij}^2(Z)$  donne :

$$\delta_{ij}^2(Z) = \begin{cases} d_{ij}^2(F) + u_i^2 + u_j^2 & \text{si } i \neq j \\ d_{ii}^2(F) & \text{si } i = j \end{cases} \quad (55)$$

**Definition:** Nous appelons méthode d'Analyse Factorielle du Triple  $(I, \delta(Z), P_1)$  tout ajustement du modèle structurel (55) issu d'une Q-Analyse Factorielle en Facteurs Communs.

**Remarque :** HARRIS (1962) observe que lorsque le modèle structurel (54) est décrit par la décomposition

$$W_Z = W + D_U^2$$

la contrainte de positivité de la matrice diagonale  $D_U^2$  peut être incompatible avec la semi-positivité de  $W$ . Lorsque  $W$  n'est pas gramienne, la décomposition

$$W = F^t F$$

peut résulter en des facteurs imaginaires (on parle alors du cas de Heywood) de sorte que  $\delta(F)$  n'est pas Euclidienne. Ce cas est étudié plus bas. Sa prise en compte impose l'utilisation de l'opérateur  $\beta_n$  défini sur  $\mathcal{Y}_n$  et non de sa restriction  $\beta_n^*$  à  $\mathcal{Y}_n^*$  pour construire  $H_F$ .

### V.1.3.1. Analyse Factorielle du Triple $(I, \delta(Z), P_1)$

Nous traitons ici le problème posé par l'ajustement d'un modèle d'Analyse Factorielle en termes de Q-Analyse Canonique attachée à l'étude des matrices de connexion  $W$ . Cette

présentation est due à HARRIS (1962) dans le cas plus classique d'une R-Analyse. Notre présentation est limitée ici au cadre empirique afin de garder une homogénéité de notations avec le reste de notre exposé. Les méthodes d'Analyse Factorielle du Triple dans le cas d'une pondération uniforme des éléments de  $I$  ont fait l'objet de travaux récents de CARROLL (1984) et de WINSBERG et CARROLL (1984), où elles sont présentées comme un prolongement des méthodes de Codage Multidimensionnel.

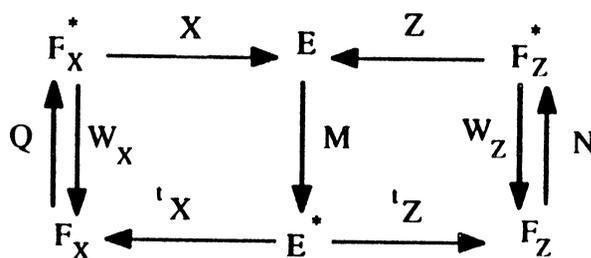
Considérons un tableau de données  $Z$  appartenant à  $\mathcal{M}_{p,n}(\mathbb{R})$  et un tableau de variables latentes  $X$  appartenant à  $\mathcal{M}_{p,r}(\mathbb{R})$  formant les inconnues du problème.  $Z$  (resp.  $X$ ) a pour colonnes les coordonnées des  $n$  unités statistiques observées (resp. les  $r$  unités statistiques latentes ou facteurs communs) dans  $E = (\mathbb{R}^p, M)$ . Bien que  $X$  soit une inconnue du problème, on peut réaliser l'Analyse Canonique formelle des deux tableaux  $Z$  et  $X$  de façon à déterminer les combinaisons linéaires de  $Z$  et de  $X$  les plus fortement corrélées. Notons qu'alors la matrice de connexion de  $[Z, X]$  vérifie

$$W = \begin{pmatrix} {}^t Z \\ {}^t X \end{pmatrix} M(Z, X) = \begin{pmatrix} W_Z & W_{ZX} \\ {}^t W_{ZX} & W_X \end{pmatrix}$$

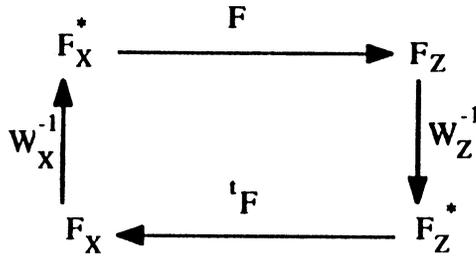
où

$$W_Z = {}^t Z M Z, \quad W_{ZX} = {}^t Z M X, \quad W_X = {}^t X M X$$

et de façon conventionnelle adoptons la notation  $F = W_{ZX}$  pour désigner la matrice de structure factorielle commune. De façon classique (cf. CAILLEZ et PAGES (1976)) l'Analyse Canonique du triplet  $(Z, X, M)$  est décrite par le schéma de dualité



avec  $\dim(F_Z) = n$  et  $\dim(F_X) = r$ , et son analyse est menée de façon équivalente par l'Analyse en Composantes Principales du triplet  $(F, W_Z^{-1}, W_X^{-1})$  résumée par le schéma de dualité



c'est à dire caractérisée par le choix de matrices  $N = W_Z^{-1}$  et  $Q = W_X^{-1}$  pour définir une géométrie euclidienne sur  $F_Z$  et  $F_X$  respectivement.

**Remarque :** Cette analogie est formelle, puisque  $X$  et  $Q$  sont inconnues. On ne peut donc espérer prouver aucunes propriétés d'emboîtement des solutions en Analyse Factorielle en Facteurs Communs.

Cette équivalence connue sous le terme de  $\Pi$ -stratégie permet de déterminer les coefficients  $\underline{b}$  et  $\underline{c}$  des combinaisons linéaires

$$Y = Z\underline{b} \text{ et } T = X\underline{c}$$

où  $\underline{b}$  et  $\underline{c}$  sont par construction des éléments de  $F_X$  et de  $F_Z$  respectivement, solutions du problème d'optimisation

$$\text{Min} \{ \|\underline{X}\underline{b} - Z\underline{c}\|_M^2 ; \underline{b} \perp \underline{c} = 0, \|\underline{b}\|_N^2 = \|\underline{c}\|_Q^2 = 1 \} \quad (56)$$

**Remarque :** De façon équivalente cette formulation du problème s'exprime sous la forme matricielle

$$X = ZC + E_1 \quad C \in \mathcal{M}_{n,r}(\mathbb{R}) \quad (57)$$

$$Z = XB + E_2 \quad B \in \mathcal{M}_{r,n}(\mathbb{R}) \quad (58)$$

où  $C$  et  $D$  sont supposées de rang  $r < n$  et  $E_1$  et  $E_2$  sont interprétées comme des matrices de résidus. Ces deux équations sont à la base des nombreuses comparaisons des méthodes en composantes qui vérifient (57) seulement avec les méthodes factorielles qui satisfont (57) et (58) (cf. MCDONALD (1968, 1970, 1975), OBENCHAIN (1973)). L'équation (58) est clairement plus générale que (57) puisqu'elle englobe les écarts d'ajustements dus à l'existence de relations non linéaires (comme (57)) mais aussi la réduction de dimension de  $n$  à  $r$ .

**Proposition 8:** Les solutions  $\underline{c}_s$  du problème (56) sont les solutions de problème aux valeurs propres

$$D_U^{-2} W_Z \underline{c}_s = \lambda_s \underline{c}_s$$

$$\langle \underline{c}_s, \underline{c}_t \rangle_{W_Z^{-1}} = \lambda_s \delta_s^t$$

où

$$\lambda_s = \frac{1}{1 - \rho_s}$$

est la valeur propre associée à la corrélation canonique  $\rho_s$ .

**Preuve :** (56) admet en particulier pour solution le vecteur  $\underline{c}$  tel que (cf. CAILLEZ et PAGES (1976))

$$(W_{ZX} W_X^{-1} W_{XZ} - \rho W_Z) \underline{c} = \underline{0}$$

c'est à dire

$$(F W_X^{-1} F - \rho W_Z) \underline{c} = \underline{0}$$

Mais puisque le modèle d'Analyse Factorielle Classique impose les contraintes de structure

$$i) W_X = I$$

$$ii) F^t F = W_Z - D_U^2$$

on réalise l'Analyse en Composantes Principales formelle du triplet  $(F, (F^t F + D_U^2)^{-1}, I)$ .

Par i) on a

$$(F^t F - \rho W_Z) \underline{c} = \underline{0}$$

et par ii)

$$[(1 - \rho) W_Z - D_U^2] \underline{c} = \underline{0}$$

On peut toujours supposer  $r < 1$  sans perte de généralité et l'on pose alors

$$\lambda_s = \frac{1}{1 - \rho_s}$$

de sorte que les vecteurs propres  $\underline{c}_s$  sont les solutions d'un problème aux valeurs propres associés aux valeurs propres  $\lambda_s$  classées par ordre décroissant. ¶

Notons de plus  $\underline{x}_s = D_U^2 \underline{c}_s$ .  $\underline{x}_s$  est un élément de  $F_Z^*$  tel que

$$W_Z D_U^{-2} \underline{x}_s = \lambda_s \underline{x}_s$$

$$\langle \underline{x}_s, \underline{x}_t \rangle_Q = \lambda_s \delta_s^t \quad \text{avec } Q = D_U^2 W_Z^{-1} D_U^2$$

**Remarque 1 :** L'étude de ce problème de Q-Analyse Canonique peut être menée dans le cadre d'une  $\Sigma$ -stratégie, en considérant l'Analyse en Composantes Principales du triple  $([Z, X], M, N)$  où  $[Z, X]$  est la représentation matricielle d'un élément de  $F \otimes E$ ,  $F$  est décomposé en

$$F = F_Z \oplus F_X$$

et la métrique  $N$  s'écrit

$$N = \begin{pmatrix} W_Z^{-1} & 0 \\ 0 & W_X^{-1} \end{pmatrix}$$

On écrit alors

$$[Z, X] N^t [Z, X] = Z W_Z^{-1} {}^t Z + X W_X^{-1} {}^t X$$

Tout vecteur propre de l'opérateur  $WN$  vérifie

$$\begin{aligned} WN \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} &= \mu \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \\ \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} &\in F \text{ et } \left\| \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \right\|_N^2 = 2 \end{aligned}$$

c'est à dire

$$\begin{aligned} W_{ZX} W_X^{-1} \mathbf{b} &= (\mu - 1) \mathbf{a} \\ W_{XZ} W_Z^{-1} \mathbf{a} &= (\mu - 1) \mathbf{b} \end{aligned}$$

et donc en particulier

$$W_{ZX} W_X^{-1} W_{XZ} W_Z^{-1} \mathbf{a} = (\mu - 1)^2 \mathbf{a}$$

Appliquons alors les contraintes i) et ii) de structure du modèle factoriel. On obtient alors

$$(W_Z - D_U^2) W_Z^{-1} \mathbf{a} = (\mu - 1)^2 \mathbf{a}$$

c'est à dire

$$D_U^2 W_Z^{-1} \mathbf{a} = \mu(2 - \mu) \mathbf{a}$$

Par suite  $\mathbf{a}_s = \mathbf{x}_s$  et

$$\mu_s (2 - \mu_s) = \frac{1}{\lambda_s} = \rho_s \quad \Leftrightarrow \quad \rho_s = 1 - \mu_s (2 - \mu_s)$$

On symétrise le problème en notant que l'équation

$$W_Z \underline{x}_s = \lambda_s D_U^2 \underline{x}_s$$

donne en posant maintenant  $\underline{x}_s = D_U \underline{c}_s$

$$D_U^{-1} W_Z D_U^{-1} \underline{x}_s = \lambda_s \underline{x}_s$$

$$\| \underline{c}_s \|_{W_Z} = \| \underline{x}_s \|_Q = 1 \quad \text{avec } Q = D_U^{-1} W_Z D_U^{-1}$$

Les composantes canoniques s'écrivent alors sous forme matricielle

$$Y = Z D_U^{-1} X D_\lambda^{-1/2}$$

et la matrice de structure factorielle  $F$ , donnant les corrélations entre les  $z_i$  et les composantes canoniques normalisées  $y_j$  est (cf. MULAİK (1972))

$$F = D_U X D_\lambda^{-1/2}$$

La relation liant  $\rho$  et  $\lambda$  montre que

$$(\lambda_s \leq 1) \Leftrightarrow (\rho_s \leq 0)$$

Puisque  $\rho_s$  s'exprime comme le carré d'un cosinus, on est amené à définir la dimension  $r$  de l'espace de représentation, comme le nombre de valeurs propres  $\lambda_s$  supérieures à 1. Les axes suivants possèdent en effet des coordonnées complexes. On pose en conséquence

$$F_r = D_U X_r (D_\lambda^{-1/2})_r$$

HARRIS (1962) remarque de plus que

$$D_U^{-1} W_Z D_U^{-1} = X D_\lambda^{-1} X \quad \Leftrightarrow \quad W_Z = D_U X D_\lambda^{-1} X D_U$$

et donc  $W_Z$  a pour inverse de Moore Penrose

$$W_Z^- = D_U^{-1} X D_\lambda^{-1} X D_U^{-1}$$

où  $D_\lambda$  est la matrice diagonale de terme général

$$l_s = \begin{cases} \frac{1}{\lambda_s} & \text{si } \lambda_s \neq 0 \\ 0 & \text{sinon} \end{cases}$$

D'autre part, puisque  $X'X = I$  par construction de  $X$

$$D_U^2 = D_U X^t X D_U$$

$$D_U^{-1} (W_Z - D_U^2) D_U^{-1} = D_U^{-1} W_Z D_U^{-1} - I = X D_\lambda^t X - X^t X = X (D_\lambda - I)^t X$$

Les calculs sont totalement achevés lorsqu'on applique le résultat suivant

**Proposition 9 (HARRIS (1962)) :** *La partie commune de Z, notée  $Z_c$ , obtenue à partir de Z en retranchant les unicités vérifie, en identifiant Z à un élément de  $\mathcal{M}_{n,n}(\mathbb{R})$*

$$Z_c = Z - D_U W_Z$$

et la matrice de structure factorielle d'ordre r s'écrit

$$F_c = D_U X_r (D_\lambda - I)_r^{1/2}$$

**Preuve :**

$$F_c = Z M^t Y_r = Z M^t Z_r D_U^{-1} X = (W_Z - D_U^2) D_U^{-1} X$$

et donc

$$F_c = D_U X_r (D_\lambda - I)_r^{1/2} \quad \parallel$$

En résumé, nous avons montré que les méthodes factorielles utilisent comme seules informations

- i) la matrice  $W_Z$  déduite de  $\delta(Z)$  par l'application de  $\alpha_1^+$ ,
- ii) la matrice des unicités qu'il convient d'estimer.

Le point ii) ne présente aucune spécificité caractéristique de l'Analyse du Triple  $(I, \delta(Z), P_I)$ . On peut donc adopter l'une quelconque des nombreuses propositions rapportées dans le cadre de l'Analyse Factorielle Classique par MULAİK (1972) ou HARMAN (1976). Ainsi RAO (1955) propose à partir d'une configuration initiale d'itérer les deux phases suivantes

- (1) estimer  $D_U$ ,
- (2) factoriser  $D_U^{-1} W_Z D_U^{-1}$  et en déduire une nouvelle estimation de  $D_U$ .

A chaque étape  $F_c$  est de dimension r fixée par la contrainte  $\lambda_s \geq 1$ . On prend alors pour initialisation

$$D_U = S^2 = (W_Z^{-1} * I)^{-1}$$

et l'actualisation de  $D_U$  proposée par RAO s'écrit

$$(u_i^2)^{-1} = \sum_{s=1}^r (x_i^s)^2 (\lambda_s - 1) + 1$$

où  $\underline{x}_s$  est le vecteur propre de  $D_U^{-1}W_Z D_U^{-1}$ . La méthode ainsi construite porte le nom d'Analyse en Composantes Principales Itératives. Son application à  $(I, \delta(Z), P_1)$  définit une méthode d'Analyse en Coordonnées Principales Itératives. De nombreuses améliorations de l'approche de RAO ont été proposées, en particulier JÖRESKOG (1969) montre que la procédure d'actualisation de RAO ne permet pas de démontrer la convergence de l'algorithme et propose une approche efficace fondée sur l'algorithme de programmation de Fletcher et Powell. Nous ne poursuivons pas cette étude ici et nous renvoyons le lecteur intéressé à la littérature citée en référence.

### V.1.3.2. L'Analyse en Coordonnées d'Images

Nous traitons dans ce paragraphe de l'application à l'étude du Triple  $(I, \delta(Z), P_1)$  de l'Analyse en Composantes d'Images, développée par GUTTMAN (1953) afin de proposer une alternative aux méthodes factorielles, plus satisfaisante selon nous que l'Analyse en Composantes Principales.

La méthode proposée par Guttman consiste à mener l'analyse spectrale de la matrice de "corrélation réduite"  $G$  calculée à partir du tableau  $\hat{Z}$  déduit de  $Z$  par une méthodologie de régression. En effet, la  $j$ -ème colonne de  $\hat{Z}$  est l'image  $\hat{Z}^j$  de la variable  $Z^j$  dans le sous-espace  $\mathcal{L} \text{ in } \langle Z^k ; k \in [p], k \neq j \rangle$  de  $F$ , c'est à dire la régression linéaire de  $Z^j$  sur ce sous-espace. Nous reprenons ici en l'adaptant au cadre d'une Q-Analyse métrique (c'est à dire en supposant l'existence d'une géométrie Euclidienne sur  $E$ ) la présentation de BENTLER (1969). Réécrivons le modèle factoriel sous sa forme additive

$$Z = C + U$$

où  $U$  est la matrice des unicités et  $C$  la matrice des composantes communes. La contrainte d'orthogonalité s'écrit dans le cadre d'une Q-Analyse

$${}^1\text{CMU} = 0$$

de sorte que

$$W_Z = {}^1\text{ZMZ} = W_C + W_U$$

La tradition factorialiste impose alors  $W_U = D_U^2$  et  $u_i^2$  est appelée unicité. Lorsqu'on suppose  $Z$  normalisée de sorte que  $W$  soit une matrice de cosinus (homogènes à des corrélations, de sorte que nous utiliserons abusivement ce terme dans la suite),  $W_C$  s'interprète alors comme la matrice des corrélations réduites et contient les communautés sur la diagonale.  $W_Z$  est un élément de  $\mathcal{Y}_n^+$  et  $W_C$  est un élément de  $\mathcal{Y}_n$  (cf. la remarque de HARRIS (1962) citée plus haut). Décomposons  $Z$  en

$$Z = PD_{\lambda}'Q \quad \text{où } 'PMP = I \text{ et } 'QM = I$$

**Lemme 5 :** *Le modèle factoriel s'écrit de façon équivalente en fonction de P et Q*

$$\text{où } W_Z = \widehat{W}_C + \widehat{W}_U + 'C\widehat{M}U + 'U\widehat{M}C$$

$$\widehat{M} = MP'PM$$

$$\widehat{W}_A = 'A\widehat{M}A$$

**Preuve :** Prémultiplions Z par P'PM, nous obtenons

$$\begin{aligned} P'PMZ &= P'PMPD_{\lambda}'Q = PD_{\lambda}'Q = Z \\ &= P'PMC + P'PMU \end{aligned}$$

Si nous appliquons l'expression  $W_Z = 'ZMZ$  à cette écriture de Z on obtient le résultat souhaité. ¶

**Proposition 10:** *La matrice de notes en facteurs P'PMC est complètement déterminée par (Z, M) et*

$$\widehat{W}_C = W_C W_Z^+ W_C \quad (60)$$

**Preuve :** Notons  $Z^+$  l'inverse généralisée de Moore Penrose de Z

$$Z^+ = QD_{\lambda}^+'P$$

où  $D_{\lambda}^+$  est la matrice diagonale de terme général

$$l_s = \begin{cases} \frac{1}{\lambda_s} & \text{si } \lambda_s \neq 0 \\ 0 & \text{sinon} \end{cases}$$

Alors puisque

$$W_{CZ} = 'CMZ = 'CM(C+U) = 'CMC = W_C$$

et donc

$$W_C Z^+ = W_{CZ} Z^+ = 'CM(PD_{\lambda}'Q)(QD_{\lambda}^+'P) = 'CMP'P$$

Lorsque C est connue, les notes en facteurs P'PMC le sont aussi. Mais

$$'(CMP'P)M(P'P'CM) = 'CMP'PMC = 'C\widehat{M}C = \widehat{W}_C$$

qui s'écrit aussi

$$\begin{aligned}\widehat{W}_C &= W_C Z^+ M {}^t Z^+ W_C = W_C ({}^t Z M Z)^+ W_C \\ &= W_C W_Z^+ W_C \quad \parallel\end{aligned}$$

$\widehat{W}_C$  s'interprète comme la matrice de corrélation de la matrice des notes en facteurs P<sup>1</sup>PMC et permet de les analyser par une méthode quelconque d'Analyse Factorielle ou d'Analyse en Composantes.

**Remarque :** Seule la contrainte d'orthogonalité  ${}^t U M C = 0$  a été utilisée pour établir ces résultats qui sont donc valables pour un grand nombre de méthodes d'analyse.

D'autre part, si  $W_C$  est définie positive, alors  $\widehat{W}_C$  l'est aussi, mais cette condition n'est pas nécessaire.

Dans la théorie de GUTTMAN (1953),  $C = \widehat{Z}$  est appelée la matrice d'images (ou d'images partielles) et la matrice des résidus  $U = Z - \widehat{Z}$  est appelée la matrice des anti-images (ou anti-images partielles). BENTLER (1969) montre que la matrice de covariance des anti-images  $\widehat{W}_C$  est en général non diagonale. Posons alors comme GUTTMAN (1953) le modèle

$$W_C = W_Z - D_U^2$$

on vérifie alors le résultat suivant

**Proposition 11:** *La matrice de covariance des images (resp. des anti-images) est*

$$\begin{aligned}\widehat{W}_C &= W_Z + D_U^2 W_Z^+ D_U^2 - 2D_U^2 \\ (\text{resp. } \widehat{W}_U &= D_U^2 W_Z^+ D_U^2)\end{aligned}$$

**Preuve :** d'après (60) la matrice de corrélation réduite s'écrit

$$\begin{aligned}\widehat{W}_C &= W_C W_Z^+ W_C = (W_Z - D_U^2) W_Z^+ (W_Z - D_U^2) \\ &= W_Z + D_U^2 W_Z^+ D_U^2 - 2D_U^2\end{aligned}$$

et clairement

$$\widehat{W}_U = D_U^2 W_Z^+ D_U^2 \quad \parallel$$

GUTTMAN propose alors de construire une représentation de  $I$  à l'aide d'une analyse spectrale de  $\widehat{W}_C$ . Lorsque  $C$  est la matrice des images définies plus haut, GUTTMAN (1953) montre que  $D_U^2$  a pour  $i$ -ème terme diagonal le carré de la corrélation multiple de  $z_i$  avec les  $z_k$  ( $k \in [n] \setminus \{i\}$ ). C'est à dire

$$D_U^2 = (W_Z^+ * I)^+ = S^2$$

On trouve alors l'expression traitée par GUTTMAN, qui suppose de plus que  $W_Z$  est inversible

$$G = W_Z + S^2 W_Z^{-1} S^2 - 2S^2$$

Les relations entre Analyse d'Images et Analyse Factorielle Canonique furent tout d'abord mises en lumière par HARIS (1962). On vérifie la relation suivante :

**Proposition 12:** Soit  $(X, \Lambda)$  le couple des éléments propres de  $Q_Z = D_U^{-1} W_Z D_U^{-1}$  solutions du problème d'Analyse Factorielle Canonique. Si  $D_U^2$  est la matrice  $S^2$  des corrélations multiples retenue par GUTTMAN et  $G^*$  la matrice définie par  $G^* = S^- G S^-$ , alors

$$G^* = X D_\gamma {}^t X$$

où

$$D_\gamma = D_\lambda + D_\lambda^{-1} - 2I$$

**Preuve :**

$$\begin{aligned} G^* &= S^- G S^- = S^- W_Z S^- + S W_Z S - 2I \\ &= (S^- W_Z S^- - I) + (S W_Z S - I) \end{aligned}$$

Posons  $Q = S^- W_Z S^-$  et notons  $D_\lambda$  la matrice diagonale des valeurs propres de  $Q$  et  $X$  la matrice des vecteurs propres associés. Nous avons montré plus haut que  $(X, D_\lambda)$  est la solution du problème d'Analyse Factorielle Canonique

$$G^* = X D_\lambda {}^t X + X D_\lambda^+ {}^t X - 2X {}^t X = X D_\gamma {}^t X$$

où

$$D_\gamma = D_\lambda + D_\lambda^+ - 2I$$

est la matrice diagonale de terme générale

$$\gamma_s = \lambda_s + \frac{1}{\lambda_s} - 2 = \frac{(\lambda_s - 1)^2}{\lambda_s} \quad \parallel$$

Remarquons de plus que les vecteurs propres de  $G^*$  sont ceux de  $Q$ , de  $Q^+$  et de la matrice

$$R = S^- (W_Z - S^2) S^- = Q - I$$

Le nombre de dimensions réelles est fixé par la contrainte

$$\lambda_s(Q) \geq 1 \Leftrightarrow \lambda_s(R) \geq 0 \Leftrightarrow \gamma_s \geq 0$$

Lorsqu'une valeur propre de  $Q$  est plus petite que 1 et proche de zéro, alors la valeur propre correspondante de  $G^*$  est positive mais très grande. Une borne supérieure pour  $\gamma_s$  permet donc de fixer le nombre de facteurs communs. De plus

$$\frac{\partial \gamma}{\partial \lambda} = \frac{\lambda^2 - 1}{\lambda^2} \quad \left\{ \begin{array}{l} \geq 0 \text{ si } \lambda \geq 1 \\ \leq 0 \text{ si } \lambda \leq 1 \end{array} \right.$$

Par suite  $\gamma$  est une fonction croissante de  $\lambda$  si  $\lambda \geq 1$  et décroissante de  $\lambda$  si  $\lambda \leq 1$ , c'est à dire pour des solutions non admissibles.

En pratique KAISER (1963) montre que la diagonalisation de  $Q$  est, parmi les problèmes équivalents donnés ci-dessus, celle qui fournit les calculs les plus simples. ANASTASSAKOS et d'AUBIGNY (1984) préconisent l'emploi de l'Analyse d'Images de préférence à l'Analyse en Composantes Principales en termes de fiabilité des critères de détermination de l'espace de représentation résumée de  $I$ . L'argumentation de ces auteurs s'applique de façon immédiate aux méthodes d'Analyse en Coordonnées d'Images et d'Analyse en Coordonnées Principales. D'autre part, les relations établies par HARRIS (1962) (cf. proposition 12) montrent que l'on peut regarder l'Analyse d'Images comme une phase d'initialisation d'une procédure d'Analyse Factorielle itérative.

Nous présentons de plus deux propriétés utiles pour l'Analyse d'Images.

**Propriété :** *La solution  $(X, D_\lambda)$  de l'analyse spectrale de  $Q$  est inchangée par toute transformation de  $Z$  de la forme  $Z \mapsto Y = ZD$  où  $D$  est une matrice diagonale à termes non nuls. Les notes en facteurs sont alors transformées par la même transformation*

**Preuve :**

$$W_Y = {}^t Y M Y = D {}^t Z M Z D = D W_Z D$$

$$W_Y^+ = D^{-1} W_Z^+ D^{-1}$$

$$S_Y^2 = (W_Y^+ * I)^+ = (D^{-1} W_Z^+ D^{-1} * I)^+ = D (W_Z^+ * I)^+ D$$

$$S_Z^2 = D S_Z^2 D = D^2 S_Z^2$$

Par suite

$$S_Y^+ = D^{-1} S_Z^+ = S_Z^+ D^{-1}$$

$$Q(Y) = S_Y^+ W_Y S_Y^+ = S_Z^+ D^{-1} D W_Z D D^{-1} S_Z^+ = S_Z^+ W_Z S_Z^+ = Q(Z)$$

D'autre part les facteurs de l'analyse de  $Q(Y)$  s'écrivent

$$F_r(Y) = S_Y X \Lambda^{1/2} = DF_r(Z) \quad (61)$$

Ils se correspondent donc par le recodage des lignes de  $F_r$ . ¶

Une conséquence directe de la propriété précédente est

**Corollaire :** L'analyse spectrale de l'opérateur  $S^{1/2}Q^+S^{1/2}$  a pour solution  $(X, D_\lambda)$  les coordonnées d'images de  $(I, \delta(Z), P_1)$ .

On applique ce corollaire à l'explicitation des liens existant entre l'Analyse en Coordonnées d'Images de  $(I, \delta(Z), P_1)$  et le choix pour origine de l'espace affine  $E$  du centre de l'hypersphère inscrite dans le simplexe  $\{P_i ; i \in [n]\}$  représentant  $I$  dans  $E$ , présenté au chapitre II. Rappelons que l'étude de GOWER (1982a) montre que  $P_0$  a pour coordonnées dans la base  $\mathfrak{X}$

$$\underline{u}_0 = r \Lambda^{-1} X S^{-1} \underline{1} \quad r = \|\underline{1}\|_S^{-2}$$

où  $X$  vérifie

$$W_Z^+ = {}^t X \Lambda^{-1} X$$

et donc en posant  $N = S^{-1} W_Z^+ S^{-1}$

$$\|\underline{u}_0\|_1^2 = r^2 {}^t \underline{1} S^{-1} W_Z^+ S^{-1} \underline{1} = r^2 \|\underline{1}\|_N^2$$

Remarquons alors que tout vecteur  $\underline{u}$  réalisant le maximum de

$$q(\underline{x}) = \frac{\|\underline{x}\|_N^2}{\|\underline{x}\|_S^{-1}}$$

optimise le Lagrangien

$$L(\underline{u}, \lambda) = \|\underline{u}\|_N^2 - \lambda(\|\underline{u}\|_S^{-1} - 1)$$

si et seulement si

$$N\underline{u} = \lambda S^{-1} \underline{u}$$

c'est à dire

$$S^{-1} W_Z^+ S^{-1} \underline{u} = \lambda S^{-1} \underline{u}$$

$$\Leftrightarrow (S^{-1/2} W_Z^+ S^{-1/2}) S^{-1/2} \underline{u} = \lambda S^{-1/2} \underline{u}$$

Posons donc  $\underline{v} = S^{-1/2} \underline{u}$ . On recherche les éléments propres solutions de

$$S^{-1/2} W_Z^+ S^{-1/2} \underline{v} = \lambda \underline{v}$$

Enfin

$$Q = S^+ W_Z S^+ \Leftrightarrow Q = S W_Z^+ S \Leftrightarrow S^{1/2} Q S^{1/2} = S^{1/2} W_Z^+ S^{1/2}$$

et donc le corollaire précédent assure que les éléments propres de  $S^{1/2} W_Z^+ S^{1/2}$  sont  $\underline{v}_s = \underline{x}_s$  et le couple  $(X, D_\lambda)$  fournit les coordonnées d'images de  $(I, \delta(Z), P_1)$ . Ainsi le choix du centre de l'hypersphère inscrite dans le simplexe  $\{P_i ; i \in [n]\}$  comme origine  $P_0$  de  $E$  apparaît cohérent lorsqu'on analyse  $(I, \delta(Z), P_1)$  en termes d'Analyse en Coordonnées d'Images.

**Remarque :** Notre présentation de l'Analyse d'Images considère  $W_Z$  comme une matrice de corrélation  $R_Z$ , alors que le formalisme de l'analyse de  $(I, \delta(Z), P_1)$  interprète  $W_Z$  comme une matrice de covariance. Mais puisque ces deux matrices sont reliées par la relation

$$R_Z = S^+ W_Z S^+$$

l'application de la propriété précédente assure de l'identité de la solution  $(X, D_\lambda)$  obtenue pour  $W_Z$  et pour  $R_Z$ . Nous ne perdons donc nullement en généralité en supposant que  $W_Z$  est une matrice de corrélation. Cependant nous avons montré que la matrice des notes en facteurs doit elle être corrigée à l'aide de la formule (61). Cette propriété d'invariance montre la nature distincte de l'Analyse en Coordonnées Principales et de l'Analyse en Coordonnées d'Images.

## V.2. ANALYSE DU TRIPLE $(I, \delta, P_1)$ DÉFINI PAR UNE DISSIMILARITÉ QUELCONQUE

Les méthodes d'Analyse en Coordonnées (Principales ou d'Images) décrites ci-dessus reposent explicitement sur le caractère gramien de la matrice  $W_Z$  associée aux données  $H(Z) = (\delta_{ij}^2(Z))$ . Les théorèmes d'existence d'un plongement Euclidien, étudiés au chapitre II montrent en particulier que cette propriété résulte directement du caractère Euclidien de la fonction de dissimilarité  $\delta$ . L'espace quasi-métrique  $(I, \delta)$  résultant d'une expérience n'est généralement pas Euclidien et la validité de l'application des méthodes d'Analyse en Coordonnées ne peut être vérifiée qu'a posteriori. Certes GLAZMAN et

LIUBITCH (1972) définissent un indice de non euclidéanité des espaces vectoriels normés qui permet d'évaluer à priori la pertinence de l'application de ce type d'analyses mais son caractère synthétique ne permet pas de détecter les causes de non euclidéanité. Posons en effet (cf. GLAZMAN et LIUBITCH (1972)) pour tout espace vectoriel normé

$$\mu((I, \delta)) = \text{Sup} \left\{ \frac{\|z_i - z_j\|^2 + \|z_i + z_j\|^2}{2(\|z_i\|^2 + \|z_j\|^2)} ; i, j \in [n] \right\}$$

où par définition

$$\delta_{ij} = \|z_i - z_j\|$$

$\delta$  est Euclidienne si et seulement si le théorème du parallélogramme est vérifié, c'est à dire si et seulement si  $\mu((I, \delta)) = 1$ . Dans le cas général  $\mu((I, \delta)) \in [1, 2]$ . Mais cette expression n'est valable que pour un espace vectoriel normé. Pour un espace quasi-métrique, elle n'est pas exprimable en fonction des dissimilarité  $\delta_{ij}$ . Cette opportunité repose en effet sur les propriétés spécifiques des normes invariantes par translation.

Dans cette section, nous montrons tout d'abord les difficultés rencontrées lors de l'application des méthodes précédentes au cas de matrices de dissimilarité non Euclidienne. Dans un deuxième temps nous remarquons que l'on doit introduire une méthodologie plus globale, consistant à associer à toute dissimilarité  $\delta$ , une dissimilarité Euclidienne  $\delta(Z)$ , que l'on analyse classiquement à l'aide d'une des méthodes précédentes. La validité des méthodes proposées repose alors sur la qualité de l'approximation de  $\delta$  par  $\delta(Z)$  et sur le critère d'approximation retenu.

Dans une telle démarche  $\delta(Z)$  apparaît comme un intermédiaire de calcul, dont le statut est le plus souvent heuristique. Nous étudions au chapitre suivant les méthodes de Codage Multidimensionnel, qui se proposent de construire la configuration X directement dans une logique d'approximation de la dissimilarité observée. On obtient ainsi des méthodes plus satisfaisantes du point de vue de la cohérence des critères d'approximation utilisés à chaque étape de l'analyse.

### V.2.1. L'Analyse en Coordonnées Principales

Dans le cas d'une fonction de dissimilarité Euclidienne, l'Analyse en Coordonnées Principales du Triple  $(I, \delta, P_1)$  est menée de façon équivalente à l'aide d'une décomposition spectrale des opérateurs  $W_Z D_P$  et  $W_Z D_P W_Z D_P$ .

Lorsque  $\delta$  n'est pas une dissimilarité Euclidienne,  $W$  n'est pas nécessairement semi-définie positive et peut conduire à l'obtention de valeurs propres négatives, traduisant l'existence de coordonnées imaginaires pures pour certains points  $P_i$ . Une telle situation peut être analysée à l'aide de la technique d'ajout de points supplémentaires. Reprenons en effet la formulation du problème originellement proposée par GOWER (1968).

Supposons qu'il existe une bijection de  $I$  dans  $[n]$  telle que les  $(k-1)$  premiers éléments  $P_i$  ( $k \leq n$ ) admettent un plongement Euclidien irréductible dans le sous-espace affine  $E_r$  de  $E$  tel que  $\dim(E_r) = r$  et introduisons un  $k$ -ième point dans l'analyse. Si le point  $P_k$  est représentable dans  $E_r$  on obtient un plongement Euclidien irréductible de  $[k]$  dans  $E_r$ . Sinon on complète la base de  $E_r$  de telle façon que pour  $i \leq k-1$ , le point  $P_i$  ait pour coordonnées dans  $E_{r+1}$

$$\underline{x}_i = \begin{pmatrix} z_i \\ 0 \end{pmatrix}$$

et  $P_k$  a pour coordonnées

$$\underline{x}_k = \begin{pmatrix} z_k \\ x_k \end{pmatrix}$$

Alors pour tout élément  $i$  de  $[k-1]$

$$\begin{aligned} \delta_{ik}^2 &= \sum_{s=1}^{r+1} (x_i^s - x_k^s)^2 = \sum_{s=1}^r (z_i^s - z_k^s)^2 + (x_k)^2 \\ &= d_{ik}^2(Z) + (x_k)^2 \end{aligned}$$

Par suite, connaissant les coordonnées  $z_i$  pour  $i = 1, 2, \dots, k$ , le nombre réel  $(x_k)^2$  vérifie

$$(x_k)^2 = \delta_{ik}^2 - d_{ik}^2(Z)$$

$x_k$  est donc réel si  $\delta_{ik} \geq d_{ik}(Z)$ , c'est à dire si la dissimilarité  $\delta_{ik}$  est estimée par défaut par  $d_{ik}(Z)$  et alors l'axe  $\Delta_{\underline{e}_k}$  est un axe réel.  $x_k$  est imaginaire pur dans le cas contraire de sorte que l'on complète  $E_r$  par un axe complexe. Cette formulation du problème regarde  $E_r$  comme un sous-espace de  $E$  et est donc fondé sur les propriétés du projecteur de  $E$  sur  $E_r$ . Une approximation bilatérale des dissimilarités  $\delta_{ik}$  par les distances reconstruites  $d_{ik}(Z)$  est caractérisée par l'existence de coordonnées imaginaires.

La matrice  $WD_P W$  est elle semi-définie positive par construction et dans la situation présente, sa décomposition spectrale conduit à l'obtention de valeurs propres positives puisque

$$\lambda_s(WD_P W) = \lambda_s^2(WD_P)$$

$W$  et  $WD_pW$  ont le même système de vecteurs propres mais le classement des valeurs propres par ordre décroissant conduit à des solutions résumées d'ordre  $r < p$  distinctes. Les deux méthodes ne coïncident plus. L'analyse de  $WD_pWD_p$  est alors la seule justifiée de notre point de vue, en particulier lorsque l'application de la méthode fait intervenir une phase de réduction des données. De plus, lorsqu'on retient les vecteurs propres de  $W$  associés aux  $r$  plus grandes valeurs propres de  $WD_pW$ , les équations données ci-dessus montrent que cette variante de la méthode conduit à une approximation bilatérale des distances interpoints. Par analogie avec la situation traitée en Analyse des Correspondances appliquée aux tableaux de contingence symétriques, les vecteurs propres associés à des valeurs propres  $\lambda_s^2(WD_p)$  associées à  $\lambda_s(WD_p) \geq 0$  (resp.  $\leq 0$ ) sont appelés facteurs directs (resp. inverses). Leur étude détaillée dans le cadre de l'Analyse des Correspondances est menée par BENZECRI et al. (1973).

**Remarque :** Lorsque  $\delta$  n'est pas une dissimilarité Euclidienne, on peut s'interroger sur la pertinence de la transformation

$$W = \alpha_p(H) \quad \text{où} \quad H = \Delta * \Delta$$

qui fonde la méthodologie de l'Analyse en Coordonnées Principales. C'est pourquoi, l'analyse spectrale de l'opérateur  $WD_pWD_p$  obtenu à partir de

$$W = \alpha_p(H) \quad \text{où} \quad H = \Delta$$

fournit une alternative toute aussi heuristique que la précédente dans le cas général, mais qui trouve sa justification dans le caractère Euclidien de  $\sqrt{\delta}$  lorsque  $\delta$  est de type  $L_1$  (cf. FICHET (1986), LE CALVE (1986)). Remarquons que l'on optimise alors la fonction critère

$$\begin{aligned} \text{Trace}(WD_pWD_p) &= \frac{1}{4} \text{Trace}(Q_p \Delta^t Q_p D_p Q_p \Delta^t Q_p D_p) \\ &= \frac{1}{4} \text{Trace}(\Delta N_p \Delta N_p) = \frac{1}{4} \|\delta\|_{N_p \otimes N_p}^2 \end{aligned}$$

Nous reprenons cette fonction critère au chapitre suivant, où nous la comparons à la fonction STRESS optimisée par certaines méthodes de Codage Multidimensionnel.

Les propriétés d'optimalité des coordonnées principales exprimées en fonction de  $WD_pWD_p$  restent inchangées et ne sont pas discutées ici. L'étude de  $WD_p$  nécessite de déterminer des propriétés d'optimalité de l'Analyse en Coordonnées Principales dans ce cas. Nous présentons ici le résultat le plus général à notre connaissance, dû à MATHAR (1985) :

**Lemme 6 :** Soit  $\|\cdot\|$  une norme arbitraire définie sur  $\mathcal{M}_{n,n}(\mathbb{R})$  et  $\mathcal{F}_H$  défini comme plus haut, alors

$$H \in \mathcal{S}_H \mapsto \|H\|_\gamma = \|Q_\gamma H' Q_\gamma\|$$

définit une norme sur  $\mathcal{F}_H$ .

**Preuve :** Seule la propriété suivante nécessite une démonstration

$$(\|H\|_\gamma = 0) \Rightarrow (H = 0)$$

Mais

$$(Q_\gamma H' Q_\gamma = 0) \Leftrightarrow (H = \underline{h}' \underline{1} + \underline{1}' h + h \underline{1}' \underline{1})$$

où

$$\underline{h} = H \underline{\gamma} \text{ et } h = \underline{\gamma}' H \underline{\gamma}$$

Puisque  $H$  appartient à  $\mathcal{F}_H$ , on doit avoir  $\underline{h} = -\frac{h}{2} \underline{1}$  et donc  $H = 0$ .  $\quad \parallel$

Notons de plus  $x_{[i]}$  les coordonnées de tout vecteur  $\underline{x}$  de  $F$ , classées par ordre décroissant et rappelons les propriétés de majoration du spectre des éléments de  $\mathcal{F}_H$

**Définition 3 (MARSCHALL et OLKIN (1979)) :** Soient  $\underline{x}$  et  $\underline{y}$  deux éléments de  $F$  tels que

$$\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

Alors  $\underline{x}$  est majoré par  $\underline{y}$  noté  $\underline{x} < \underline{y}$  si

$$\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]} \quad k = 1, 2, \dots, (n-1)$$

Associons au vecteur  $\underline{\lambda}(A)$  les valeurs propres de  $A$ , les vecteurs  $\underline{\lambda}_c(A)$  et  $\underline{\lambda}_d(A)$  des valeurs propres de  $A$  rangées par ordre croissant et décroissant respectivement. Alors MATHAR (1985) démontre le résultat suivant, établi indépendamment par CROQUETTE (1980)

**Lemme 7 :** Si  $A$  et  $B$  appartiennent à  $\mathcal{F}_H$ , alors

- i)  $\underline{\lambda}_d(A) + \underline{\lambda}_c(B) < \underline{\lambda}(A+B) < \underline{\lambda}_d(A) + \underline{\lambda}_d(B)$
- ii)  $\underline{\lambda}_d(A) - \underline{\lambda}_d(B) < \underline{\lambda}(A-B) < \underline{\lambda}_d(A) - \underline{\lambda}_c(B)$

Alors pour tout réel  $p, p \geq 1$ , notons  $f_p$  l'élément de  $F^*$  défini par

$$\underline{x} \in F \mapsto f_p(\underline{x}) = \sum_{i=1}^n |x_i|^p = \sum_{i=1}^n y_i$$

$f_p$  est la puissance  $p$ -ième de la norme  $L_p$  de  $x$  et  $f_p$  et  $(f_p)^{1/p}$  sont deux fonctions Schur convexes (cf. BERGE (1959), CROQUETTE (1980)). De plus  $(f_p)^{1/p}$  est une jauge symétrique et donc

$$\|A\|_p = [f_p(\lambda(A))]^{1/p} \quad p \geq 1$$

est une norme unitairement invariante sur  $\mathcal{Y}_H$  (cf. MARSHALL et OLKIN (1959), RAO (1980), CLARET (1987)). Notons alors

$$x^+ = \text{Max}\{x, 0\} \quad \text{et} \quad x^- = \text{Min}\{x, 0\}$$

on vérifie alors le résultat suivant :

**Théorème 3 :** Soit  $S$  un élément de  $\mathcal{Y}_n$  admettant une décomposition spectrale

$$S = VD_\lambda^t V \quad \text{où} \quad {}^tVMV = I \quad \text{et} \quad D_\lambda = \text{diag}(\lambda_{[1]}(S), \dots, \lambda_{[n]}(S))$$

Soit  $k$  un entier,  $k \leq n$ . Alors pour tout réel  $p, p \geq 1$ , on a

$$i) \quad \text{Min} \{ \|S - A\|_p ; A \in \mathcal{Y}_n^+, \text{rang}(A) \leq k \} = \left( \sum_{i=1}^k |\lambda_{[i]}^-(S)|^p + \sum_{i=k+1}^n |\lambda_{[i]}(S)|^p \right)^{1/p}$$

ii) le minimum est atteint par

$$\hat{S} = VD_\lambda^+(k)^t V$$

$$D_\lambda^+(k) = \text{diag}(\lambda_{[1]}^+(S), \dots, \lambda_{[k]}^+(S), 0, \dots, 0)$$

**Preuve :**  $f_p$  est Schur convexe et donc le lemme 7 implique

$$\begin{aligned} \|S - A\|_p^p &= \left( \sum_{i=1}^n |\lambda_{[i]}(S - A)|^p \right) \geq \sum_{i=k+1}^n |\lambda_{[i]}(S) - \lambda_{[i]}(A)|^p \\ &\geq \sum_{i=1}^k |\lambda_{[i]}^-(S)|^p + \sum_{i=k+1}^n |\lambda_{[i]}(S)|^p \quad \square \end{aligned}$$

Par suite, puisque  $\|\cdot\|_{p,Y}$  est une norme sur  $\mathcal{Y}_H$  et que  $\mathcal{D}_E^{*2}(I)$  est inclu dans  $\mathcal{Y}_H$ , on vérifie le résultat suivant dû à MATHAR (1985) :

**Théorème 4 :** Soit  $H$  un élément de  $\mathcal{Y}_H$  et  $\gamma$  un élément de  $F^*$  tel que  $\gamma \underline{1} = 1$ . Supposons de plus que

$$Q_\gamma H' Q_\gamma = V D_\lambda' V \quad \text{avec} \quad V M V = I$$

Soient  $\widehat{W} = V D_\lambda^+(k) V$  et  $\widehat{H} = \beta_\gamma^+(\widehat{W})$ , alors

$$i) \widehat{H} \in \mathcal{D}_E^{*2}(I)$$

$$ii) \|H - \widehat{H}\|_{p,\gamma} \leq \|H - A\|_{p,\gamma} \quad \forall p \geq 1, \forall A \in \mathcal{D}_E^{*2}(I), \text{ telle que } \text{rang}(Q_1 A' Q_1) \leq k < n.$$

**Preuve :** Le théorème 1 implique que pour tout élément  $A$  de  $\mathcal{Y}_n$  tel que  $Q_\gamma A' Q_\gamma$  soit semi-définie positive et de rang au plus égal à  $k$

$$\|Q_\gamma H' Q_\gamma - \widehat{W}\|_p \leq \|Q_\gamma H' Q_\gamma - Q_\gamma A' Q_\gamma\|_p$$

Mais

$$\widehat{W} = Q_\gamma \widehat{W}' Q_\gamma = Q_\gamma \widehat{H}' Q_\gamma$$

Par suite pour tout élément  $A$  de  $\mathcal{D}_E^{*2}(I)$ , telle que  $\text{rang}(Q_1 A' Q_1) \leq k < n$

$$\|H - \widehat{H}\|_{p,\gamma} \leq \|H - A\|_{p,\gamma}$$

et clairement la définition de  $\widehat{H}$  assure que  $\widehat{H} \in \mathcal{D}_E^{*2}(I)$ .  $\quad \square$

**Remarque 1 :** Dans le cas particulier où  $p = 2$  et  $\gamma = \frac{1}{n} \underline{1}$ , le résultat de Mathar est classique et présenté en particulier par MARDIA (1978b).

Alors, en ignorant l'indice  $p (= 2)$ , on obtient pour la norme de Hilbert Schmidt définie par

$$\|A\|^2 = \text{Trace}(A D_\gamma A D_\gamma)$$

$$\|H - \widehat{H}\|_\gamma^2 = \|Q_\gamma (H - \widehat{H})' Q_\gamma\|^2 = \text{Trace}(Q_\gamma (H - \widehat{H})' Q_\gamma D_\gamma Q_\gamma (H - \widehat{H})' Q_\gamma D_\gamma)$$

$$= \text{Trace}((H - \widehat{H}) D_\gamma Q_\gamma (H - \widehat{H}) D_\gamma Q_\gamma) = \|H - \widehat{H}\|_{N_\gamma, N_\gamma}^2 = \|\mathbf{h} - \widehat{\mathbf{h}}\|_{N_\gamma \otimes N_\gamma}^2$$

où

$$N_\gamma = D_\gamma Q_\gamma = Q_\gamma D_\gamma = D_\gamma - D_\gamma \underline{1} \underline{1}' D_\gamma$$

Ce résultat a pour conséquence de justifier la pratique courante consistant à négliger les valeurs propres négatives en Analyse en Coordonnées Principales, en termes de moindres carrés pondérés. Cependant, le caractère peu intuitif de la pondération  $N_\gamma \otimes N_\gamma$  utilisée limite la portée de ce résultat. Remarquons de plus que

$$\begin{aligned}
\| H - \hat{H} \|_{N_\gamma, N_\gamma}^2 &= \text{Trace} ( {}^t Q_\gamma (H - \hat{H}) D_\gamma Q_\gamma (H - \hat{H}) {}^t Q_\gamma D_\gamma ) \\
&= 4 \text{Trace} ( (W - \hat{W}) D_\gamma (W - \hat{W}) D_\gamma ) \\
&= 4 \| W - \hat{W} \|_{D_\gamma, D_\gamma}^2 = 4 \| \underline{w} - \hat{\underline{w}} \|_{D_\gamma \otimes D_\gamma}^2
\end{aligned}$$

Cette identité montre que cette optimalité porte fondamentalement sur l'approximation de  $W$  en termes de moindres carrés pondérés par une matrice gramienne  $\hat{W}$ .

**Remarque 2 :** Lorsqu'on applique ce résultat au tableau  $W = \alpha(\Delta)$  (resp.  $\alpha(\Delta * \Delta)$ ), on obtient un critère de reconstitution des dissimilarités  $\Delta$  qu'il est intéressant de rapprocher de la fonction STRESS (resp. SSTRESS) étudiée au chapitre suivant.

### V.2.2. Les Problèmes de Constante Additive et leurs Généralisations

Une deuxième forme d'approximation, destinée à assurer le caractère semi-défini de  $W$  est fondée sur les principes de perturbation d'une matrice (cf. YOSHIDA (1965) et KATO (1976)) et reprend dans le domaine de l'Analyse du Triple  $(I, \delta, P_I)$ , les principes utilisés en Statistique Classique, pour la construction d'estimateurs rétrécis (cf. STEIN (1956)) et en particulier de la régression Ridge (cf. HEARL et KENNARD (1970), DEMPSTER et al. (1977), VINOD (1978)) ou bayésienne (cf. ZELLNER (1971), LEAMER (1973, 1978), BELSEY et al. (1980)) pour se prémunir des problèmes de colinéarité. Plus spécifiquement, l'objectif établi par les promoteurs de ces méthodes est de protéger les estimations des paramètres d'un modèle linéaire des instabilités liées à l'existence de multicollinéarité entre prédicteurs. Celles-ci ont pour conséquence l'existence de valeurs propres nulles ou négligeables de la matrice de covariance des prédicteurs, qui est alors non inversible ou mal conditionnée.

L'analogie entre les méthodes de régression Ridge et les problèmes dits de constante additive ne semble pas avoir été notée jusqu'ici. Ceci peut être dû au changement de perspective nécessaire à ce constat, puisqu'ici on s'intéresse au conditionnement de la matrice de connexion  $W_X$  et non de la matrice de covariance  $V_X$ . De plus les données étant fournies par une matrice de dissimilarité,  $W_X$  ne se calcule pas sous la forme  ${}^t X M X$  et peut donc ne pas être semi-définie positive. Le problème posé par des multicollinéarités éventuelles est aggravé par cette possibilité.

Le problème de la constante additive trouve son origine dans les travaux initiaux en psychométrie (cf. TORGERSON (1952)). Il est alors présenté en relation avec les contraintes d'échelle de mesure des dissimilarités observées. On définit alors la matrice

$$H(c) = H + c (\mathbf{1}\mathbf{1} - I)$$

de terme général

$$h_{ij}(c) = h_{ij} + c (1 - \delta_i^j)$$

où  $\delta_i^j$  représente le symbole de Kronecker. Deux cas distincts sont en général retenus dans la littérature

### V.2.2.1. Perturbation de type 1

Posons

$$\begin{aligned} h_{ij} &= \delta_{ij} \text{ , } h_{ij}(c) = \delta_{ij}(c) \\ \Delta(c) &= \Delta + c (\mathbf{1}\mathbf{1} - I) \end{aligned}$$

On se propose alors de déterminer la valeur minimale de  $c$  de sorte que  $\delta(Z)$  soit une dissimilarité Euclidienne. La solution initiale du problème (cf. TORGERSON (1952)) consiste à choisir la valeur minimale  $c^*$  pour laquelle  $\delta_{ij}(Z)$  vérifie l'inégalité triangulaire.

**Lemme 8 :**

$$c^* = -2 \text{ Min} \{ w_{ij}^k \text{ ; } (i, j, k) \in [n]^3 \text{ , } i \neq j \neq k \}$$

**Preuve :** Soient  $i, j$  et  $k$  trois éléments de  $[n]$  tels que  $i \neq j \neq k$ , alors

$$\begin{aligned} w_{ij}^k(c) &= -\frac{1}{2} [\delta_{ij}(c) - \delta_{ik}(c) - \delta_{kj}(c)] = \psi_{ik, jk}(c) \\ &= w_{ij}^k + \frac{c}{2} = \psi_{ik, jk} + \frac{c}{2} \end{aligned}$$

Par suite  $\delta(Z)$  vérifie l'inégalité triangulaire si et seulement si

$$\text{Min} \{ w_{ij}^k(Z) \text{ ; } (i, j, k) \in [n]^3 \text{ , } i \neq j \neq k \} \geq 0$$

c'est à dire

$$\text{Min} \{ w_{ij}^k \text{ ; } (i, j, k) \in [n]^3 \text{ , } i \neq j \neq k \} \geq -\frac{c}{2}$$

La solution est donc obtenue pour

$$c^* = -2 \text{ Min} \{ w_{ij}^k \text{ ; } (i, j, k) \in [n]^3 \text{ , } i \neq j \neq k \} \quad \blacksquare$$

SCHÖNEMAN (1971) montre avec raison que cette solution assure que  $\delta(Z)$  est une distance si  $\delta$  est une dissimilarité propre mais n'assure pas son caractère Euclidien. SAITO (1978) (voir aussi de LEEUW et HEISER (1982)) montre le résultat suivant :

**Proposition 13:** Soit  $(I, \delta)$  un espace quasi-métrique, où  $I$  est de cardinal fini égal à  $n$ . Alors il existe un réel  $c$  tel que  $(I, \delta(c))$  admet un plongement Euclidien dans un espace affine de dimension  $(n-1)$ .

**Preuve :** Notons  $H(c)$  la matrice  $\Delta(c) * \Delta(c)$ . Elle a pour terme général

$$h_{ij}(c) = \delta_{ij}^2(c) + 2c\delta_{ij} + c^2(1 - \delta_i^j)$$

et la matrice  $W(c) = \alpha_1(H(c))$  s'écrit

$$\begin{aligned} W(c) &= -\frac{1}{2} Q_1 H(c) {}^t Q_1 \\ &= -\frac{1}{2} [Q_1 H {}^t Q_1 + 2c Q_1 \Delta {}^t Q_1 + c^2 Q_1] \\ &= W_2 + 2c W + \frac{c^2}{2} Q_1 \end{aligned}$$

où nous avons posé

$$\begin{aligned} W_2 &= -\frac{1}{2} Q_1 H {}^t Q_1 = -\frac{1}{2} Q_1 \Delta * \Delta {}^t Q_1 \\ W &= -\frac{1}{2} Q_1 \Delta {}^t Q_1 \end{aligned}$$

Par suite il existe un seul  $c^*$  tel que pour tout  $c > c^*$ ,  $w_{ij}(c) < 0$  pour tout  $i$  et  $j$  de  $[n]$ . Mais  $W(c)$  appartient à  $\mathcal{P}_1$  et donc le théorème de TAUSSKI (1949) permet d'affirmer que  $W(c)$  est positive semi-définie pour  $c > c^*$  et le rang de  $W(c)$  est alors égal à  $(n-1)$ , ce qui assure le résultat. ¶

de LEEUW et HEISER(1982) montrent de plus le résultat corollaire suivant :

**Proposition 14:** Soit  $(I, \delta)$  un espace semi-métrique, où  $I$  est de cardinal fini égal à  $n$ . Si  $(I, \delta)$  admet un plongement affine Euclidien dans  $E$  de dimension  $(n-1)$ , alors il existe un réel  $c$  tel que  $(I, \delta(c))$  admet un plongement Euclidien dans un espace affine de dimension au plus égale à  $(n-2)$ .

**Preuve** (de LEEUW et HEISER (1982)) : Soit

$$q(c) = \text{Min} \{ {}^t \underline{u} W(c) \underline{u} ; \underline{u} \text{ tel que } {}^t \underline{u} \underline{u} = 1, {}^t \underline{u} \underline{1} = 0 \}$$

Alors  $q(c)$  est une fonction continue de  $c$ . Puisque  $q(0) < 0$  par hypothèse, la proposition 13 montre qu'il existe un réel  $c^*$  tel que  $q(c) > 0$  pour tout  $c > c^*$ . Par suite, il existe un réel  $c$  appartenant à  $[0, c^*]$  et tel que  $q(c) = 0$ . ¶

L'étude des variations des valeurs propres de  $W(c)$  en fonction de  $c$  fait appel aux travaux de LANCASTER (1964). Elle ne fournit pas d'expression explicite de  $c^*$ . SAITO (1978)

propose un algorithme itératif qui utilise une méthode de gradient, initialisée par le choix de la constante déterminée par le lemme 8. Une construction de  $c^*$  plus satisfaisante, bien que manquant d'une interprétation géométrique, est préconisée par CAILLEZ (1983):

**Théorème 5:** *La constante additive  $c^*$  intervenant dans la proposition 14 est la plus grande valeur propre de la matrice*

$$B = \begin{pmatrix} 0 & 2W_2 \\ -I & -4W \end{pmatrix}$$

**Preuve (CAILLEZ (1983)) :** On cherche  $c^*$  tel que pour tout  $c > c^*$  et tout  $\underline{x}$  de  $F^*$

$${}^t \underline{x} W(c) \underline{x} \geq 0$$

Remarquons tout d'abord que  $c^*$  est strictement positif. En effet, pour tout  $\underline{x}$  il existe un réel  $c(\underline{x})$  tel que  ${}^t \underline{x} W(c) \underline{x} \geq 0$  pour tout  $c \geq c(\underline{x})$  et puisque  $\delta$  n'est pas Euclidienne, il existe un élément  $\underline{x}$  de  $F^*$  tel que  ${}^t \underline{x} W(c) \underline{x} \geq 0$  et pour lequel  $c(\underline{x}) > 0$  et donc la solution  $c^* = \text{Sup} \{ c(\underline{x}) ; \underline{x} \in F \}$  est positive. Par suite, la solution  $(c^*, \underline{x}^*)$  vérifie  $\underline{x}^* \in \text{Ker}(W(c^*))$ . C'est à dire

$$(W_2 + 2c^* W + \frac{(c^*)^2}{2} I) {}^t Q_1 \underline{x}^* = \underline{0}$$

Posons alors

$$\begin{aligned} 2W_2 \underline{x}^* &= c^* \underline{y} \\ \underline{y} + 4W \underline{x}^* + c^* Q_1 \underline{x}^* &= \underline{0} \end{aligned}$$

Ecrivons ce système d'équations sous forme matricielle

$$\begin{pmatrix} 0 & 2W_2 \\ -I & -4W \end{pmatrix} \begin{pmatrix} \underline{y} \\ Q_1 \underline{x}^* \end{pmatrix} = c^* \begin{pmatrix} \underline{y} \\ Q_1 \underline{x}^* \end{pmatrix}$$

$c^*$  est donc valeur propre de B. D'autre part, pour tout couple  $(\lambda, \begin{pmatrix} \underline{u} \\ \underline{v} \end{pmatrix})$  d'éléments propres de B

$$\begin{aligned} \text{(i)} \quad 2W_2 \underline{u} &= \lambda \underline{u} \\ \text{(ii)} \quad -\underline{u} - 4W \underline{v} &= \lambda \underline{v} \end{aligned}$$

alors

$$\text{(i)} \Rightarrow (W_2 + 2\lambda W + \frac{\lambda^2}{2} I) \underline{u} = \underline{0}$$

$$\text{(ii)} \Rightarrow (W_2 + 2\lambda W + \frac{\lambda^2}{2} Q_1) \underline{v} = \underline{0} \Leftrightarrow {}^t Q_1 \underline{v} = \underline{0}$$

Par suite

$$c^* \geq c(t) \geq \lambda \quad \parallel$$

**Remarque :** CAILLEZ (1983) note que la matrice

$$W_2 + 2cW + \frac{c^2}{2} Q_1$$

est aussi la matrice  $W$  associée par  $\alpha_1$  à la dissimilarité  $\delta(c)$  définie par

$$\delta_{ij}(c) = \begin{cases} |\delta_{ij} + c| & \text{si } i \neq j \\ 0 & \text{sinon} \end{cases}$$

par suite, la proposition 14 admet un équivalent direct pour cette dissimilarité en prenant  $c^* \leq 0$  et  $c < c^*$ . Alors  $c^*$  est la plus petite valeur propre de  $B$ .

D'autre part, LE CALVE (1988) montre que si  $(I, \delta)$  est un espace quasi-métrique où  $I$  est de cardinal fini, les quantités  $w^k$  définies pour tout  $k$  de  $[n]$  par

$$w^k = \text{Min} \{ w_{ij}^k ; i, j \in [n], i \neq j \neq k \} = \text{Min} \{ \psi_{ik, kj} ; i, j \in [n], i \neq j \neq k \} \quad (62)$$

jouent un rôle privilégié dans l'analyse de dissimilarités quelconques. Remarquons tout d'abord que  $w^k$  est positif lorsque  $\delta$  est un indice de semi-distance. Posons alors pour tout couple  $(i, j)$  d'éléments de  $[n]$

$$\delta_{ij}^* = \begin{cases} \delta_{ij} - (w^i + w^j) & \text{si } i \neq j \\ 0 & \text{sinon} \end{cases} \quad (63)$$

On vérifie alors que la relation binaire  $\mathcal{R}$  définie sur  $[n]$  par

$$(i\mathcal{R}j) \Leftrightarrow (\delta_{ij}^* = 0) \quad (\Leftrightarrow (\delta_{ij} = w^i + w^j))$$

vérifie les propriétés suivantes pour toute dissimilarité propre :

**Proposition 15 (LE CALVE (1988)) :** Soit  $(I, \delta)$  un espace quasi-métrique défini sur l'ensemble  $I$  de cardinal fini par la donnée d'une dissimilarité propre  $\delta$ . Alors si  $\underline{w}$  et  $\underline{\delta}^*$  sont définies par (62) et (63) respectivement, on a :

- i)  $\delta^*$  est une distance,
- ii)  $\mathcal{R}$  est une relation d'équivalence,
- iii)  $(i\mathcal{R}j) \Leftrightarrow (w_{il}^k = w_{jl}^k \quad \forall k, l \in [n])$ .

**Preuve :** i) résulte de l'inégalité

$$w^k \leq w_{ij}^k = -\frac{1}{2}(\delta_{ij} - \delta_{ik} - \delta_{kj})$$

et de sa conséquence directe

$$\begin{aligned} \delta_{ij} - w^i - w^j &\leq (-2w^k + \delta_{ik} + \delta_{kj}) - w^i - w^j \\ \Leftrightarrow \delta_{ij}^* &\leq \delta_{ik}^* + \delta_{kj}^* \end{aligned}$$

Si  $\delta^*$  est une distance, alors ii) est vérifiée. D'autre part le contraste  $2 \times 2$

$$\Psi_{ij,kl} = w_{il}^k - w_{jl}^k$$

s'écrit

$$\Psi_{ij,kl} = -\frac{1}{2}(\delta_{il} - \delta_{ik} - \delta_{jl} + \delta_{jk})$$

$$\Psi_{ij,kl} = -\frac{1}{2}(\delta_{il}^* - \delta_{ik}^* - \delta_{jl}^* + \delta_{jk}^*)$$

De plus l'inégalité triangulaire assure que si  $\delta_{ij}^* = 0$  alors  $\delta_{ik}^* = \delta_{jk}^*$  pour tout  $k$  de  $[n]$  et donc  $\delta_{ij}^* = 0$  si et seulement si  $\Psi_{ij,kl}^* = 0$ . D'où les identités cherchées.  $\square$

**Remarque :**

$$\Psi_{ij,kl} = w_{il}^k - w_{jl}^k = -\frac{1}{2}(\delta_{jl}^* - \delta_{jk}^* - \delta_{kl}^*)$$

puisque  $\delta_{ij}^* = 0$ , et donc ce contraste  $2 \times 2$  n'est pas nul en général, de sorte que l'on n'a pas

$$w_{il}^k = w_{jl}^k$$

L'utilisation de ces propriétés en Analyse des Données est liée à la notion d'arbres additifs introduite par BUNEMAN (1971, 1974) (cf. aussi SATTAH et TVERSKI (1977), de SOETE et al. (1984, 1986), de SOETE (1983, 1984), CARROLL et PRUZANSKI (1980), MIRKIN (1987), BROSSIER (1986) et BARTHELEMY et GUENOCHÉ (1988)). Les arbres additifs fournissent un mode de représentation arborée de  $I$ , qui est une extension simple de la notion d'arbre hiérarchique. On note en effet que si  $\eta$  est une bipartition de  $I$  et  $\delta^\eta$  l'indice de dissimilarité défini par

$$\delta_{ij}^\eta = \begin{cases} 1 & \text{si } i \text{ et } j \text{ appartiennent à des classes distinctes de } \eta \\ 0 & \text{sinon} \end{cases}$$

alors  $\delta^\eta$  est une pseudo-métrique, puisqu'elle vérifie l'inégalité triangulaire. Par suite toute combinaison linéaire à coefficients réels positifs

$$\delta = \sum_{\eta \in T} \alpha_\eta \delta^\eta$$

est une pseudo-métrie. BUNEMAN (1971) appelle une telle pseudo-métrie, une métrique d'arbre additif. En effet cet auteur observe que  $\delta$  est décrite par un arbre  $T$  qu'elle définit de façon unique. Il montre de plus qu'une condition nécessaire et suffisante pour qu'une dissimilarité définisse une métrique d'arbre additif est qu'elle vérifie l'inégalité quadrangulaire:

**Définition 4:** Soit  $(I, \delta)$  un espace quasi-métrique où  $I$  est de cardinal fini.  $\delta$  vérifie l'inégalité quadrangulaire si et seulement si pour tout quadruplet  $(i,j,k,l)$  de  $[n]$ , est vérifiée l'une des deux inéquations équivalentes suivantes :

$$i) \delta_{ij} + \delta_{kl} \leq \text{Max} \{ \delta_{ik} + \delta_{jl}, \delta_{il} + \delta_{jk} \}$$

$$ii) \text{Min} \{ \Psi_{ij,kl}, \Psi_{ik,jl} \} \leq 0$$

L'expression ii) montre en particulier, en posant  $k=l$  que  $\delta$  vérifie l'inégalité triangulaire. Plus généralement l'inégalité quadrangulaire joue vis à vis de l'inégalité ultramétrique un rôle analogue à celui de l'inégalité de Ptolémée rappelée au chapitre II, vis à vis de l'inégalité triangulaire.

Rappelons que  $\delta$  est une dissimilarité ultramétrique si et seulement si pour tout triplet  $(i,j,k)$  de  $[n]$

$$\delta_{ij} \leq \text{Max} \{ \delta_{ik}, \delta_{kj} \}$$

**Remarque :** Cette inégalité s'écrit de façon équivalente

$$0 \leq \text{Max} \{ \delta_{ik} - \delta_{ij}, \delta_{kj} - \delta_{ij} \}$$

c'est à dire, en fonction de l'inégalité triangulaire

$$\delta_{ij} - \delta_{ik} - \delta_{kj} + \text{Min} \{ \delta_{ik}, \delta_{kj} \} \leq 0$$

et donc l'inégalité ultramétrique implique l'inégalité triangulaire, de sorte que  $(I, \delta)$  est un espace semi-métrique si  $\delta$  est une fonction de dissimilarité ultramétrique.

$$w_{ij}^k \geq \frac{1}{2} \text{Min} \{ \delta_{ik}, \delta_{kj} \}$$

$$\Leftrightarrow \Psi_{ik,kj} \geq \frac{1}{2} \text{Min} \{ \delta_{ik}, \delta_{kj} \}$$

LE CALVE (1988) montre de plus que  $\delta$  est ultramétrique si et seulement si

$$w_{ij}^k = w_{ik}^j = \frac{1}{2} \delta_{ik}$$

JOHNSON (1967) (cf. aussi BENZECRI et collaborateurs (1973 tome 2)) montre qu'il existe un isomorphisme de structure ordonnée entre les espaces semi-métriques ultramétriques de cardinal fini et les hiérarchies indicées représentées par des dendogrammes ou arbres hiérarchiques.

LE CALVE (1988) montre alors le théorème suivant, qui repose sur l'interprétation de  $w^k$  comme la dissimilarité de  $k$  à son "pied"  $k^*$  :  $\delta_{kk^*} = w^k$ .

**Théorème 6:** Soit  $(I, \delta)$  un espace quasi-métrique où  $I$  est de cardinal fini. Alors  $\delta$  vérifie l'inégalité quadrangulaire si et seulement si pour tout  $k$  de  $[n]$ , il existe une partition de  $I$  en  $k$  classes, telle que

$$(w_{ij}^k = w^k) \Leftrightarrow (i \text{ et } j \text{ appartiennent à deux classes distinctes})$$

On vérifie en effet que si  $k^*$  est le pied de  $k$  alors

$$w_{ij}^k = w_{ij}^{k^*} + w^k$$

et  $k^*$  est le nœud de l'arbre additif représentatif de  $I$ , au niveau duquel le point  $k$  se rattache à l'arbre.  $\delta$  est donc représentée par un arbre additif si et seulement si  $w_{ij}^{k^*}$  est strictement positif lorsque  $i$  et  $j$  appartiennent à deux arêtes distinctes issues de  $k^*$ .

La notion d'arbre additif permet de donner une interprétation concrète aux quantités  $w^k$ , à l'aide d'outils issus de la théorie des graphes. Les termes  $w^k$  sont d'autre part utiles dans la généralisation suivante du problème de la constante additive.

Proposons nous d'associer à la dissimilarité  $\delta$ , la fonction  $\delta(\underline{u})$  définie par

$$\delta_{ij}(\underline{u}) = \begin{cases} \delta_{ij} - u_i - u_j & \text{si } i \neq j \\ \delta_{ij} & \text{sinon} \end{cases} \quad (63\text{bis})$$

soit sous forme matricielle

$$\Delta(\underline{u}) = \Delta - \underline{u} \mathbf{1} \mathbf{1} - \mathbf{1} \mathbf{1} \underline{u} + 2D_{\underline{u}}$$

On remarque ainsi que la matrice  $\underline{u} \mathbf{1} \mathbf{1} + \mathbf{1} \mathbf{1} \underline{u} - 2D_{\underline{u}}$  n'est associée à aucun élément de  $F \vee F$ . C'est un élément de l'ensemble  $\mathcal{R}$  défini au paragraphe IV.5. ¶

LE CALVE (1988) montre alors le résultat suivant:

**Théorème 7:** Soit  $(I, \delta)$  un espace quasi-métrique où  $I$  est de cardinal fini et  $\delta$  est une dissimilarité propre. Alors il existe un vecteur  $\underline{u}$  de  $F$  tel que la dissimilarité  $\delta(\underline{u})$  associée à  $\delta$  par (63-bis) est Euclidienne.

**Preuve :** Puisque  $\delta$  peut être décomposée pour  $i \neq j$  en

$$\delta_{ij} = \delta_{ij}^* + w^i + w^j$$

notons  $c$  la constante additive telle que  $\delta^*(c)$  soit Euclidienne. Alors posons pour tout  $i \in [n]$

$$u_i = w^i - \frac{c}{2}$$

de sorte que

$$\delta_{ij}^*(c) = \delta_{ij}^* + c = \delta_{ij} - w^i - w^j - c = \delta_{ij} - u_i - u_j = \delta_{ij}(\underline{u})$$

$\delta(\underline{u})$  est la dissimilarité Euclidienne cherchée.  $\parallel$

Le Calvé interprète les  $u_i$  en termes d'Analyse Factorielle en Facteurs Communs et Spécifiques. Cette interprétation ne correspond cependant pas à un modèle structurel exprimable en termes de matrice de configuration. Il nous paraît donc plus satisfaisant de retenir l'interprétation proposée par SAITO (1983, 1985, 1986) qui étudie un tel modèle comme un cas particulier de la classe des modèles structurels

$$\delta_{ij} = \begin{cases} d_{ij}(Z) - c_{ij} & \text{si } i \neq j \\ d_{ij}(Z) & \text{sinon} \end{cases}$$

spécifié par la structure simple

$$c_{ij} = \begin{cases} u_i + u_j & u_i, u_j \in \mathbb{R} \text{ si } i \neq j \\ 0 & \text{sinon} \end{cases}$$

Lorsque  $u_i$  est positif (resp. négatif), la distance Euclidienne  $d_{ij}(Z)$  est diminuée (resp. accrue) de la quantité  $|u_i|$ . Saito interprète  $u_i$  comme un effet de contexte qui affecte l'évaluation subjective de la dissimilarité entre  $i$  et  $j$ . Il reprend ainsi et formalise une idée étudiée auparavant par TORGERSON (1965), GREGSON (1975) et TVERSKY (1977). SJOBERG (1972) démontre l'existence d'effets de contexte lors de l'expérience de jugements de dissimilarité portant sur des unités statistiques hétérogènes.

Ce modèle peut de plus recevoir une interprétation en termes de densité spatiale. En effet KRUMHANSL (1978) décrit un modèle qui s'écrit dans le cas de données numériques

$$\delta_{ij} = \begin{cases} d_{ij}(Z) + \alpha\delta_i + \beta\delta_j & \text{si } i \neq j \\ d_{ij}(Z) & \text{sinon} \end{cases}$$

où  $\delta_i$  est une mesure de la densité spatiale et où  $\alpha$  et  $\beta$  sont des constantes positives. L'interprétation probabiliste de ce modèle conduit alors à imposer des contraintes de positivité à  $\alpha$ ,  $\beta$  et  $\delta_i$ , c'est à dire de négativité aux paramètres  $u_i$ . Cet auteur ne propose cependant pas d'algorithme résolutif.

De plus Saito donne une troisième interprétation à ce modèle dans la tradition des Modèles de Contenu. Chaque unité statistique, interprétée comme un stimulus est associée à une région de l'espace de représentation. La dissimilarité  $\delta_{ij}$  mesure alors l'écart entre les régions  $i$  et  $j$ . Le modèle de Saito identifie ces régions à des hypersphères et lorsque le rayon de celles-ci tend vers 0,  $\delta_{ij}$  tend vers  $d_{ij}(Z)$  qui est égale à la distance Euclidienne entre les centres des hypersphères  $i$  et  $j$ .

C'est un modèle concurrent du modèle proposé par EKMAN (1963) (resp. MICKO (1970) et WAERN (1970)) dans lequel, toute région est représentée par un vecteur (resp. un vecteur et son halo, et un rectangle). Pour Saito,  $\delta_{ij}$  ne dépend que de la distance Euclidienne entre les centres des hypersphères  $i$  et  $j$  et de leurs rayons. SAITO (1986) produit un algorithme résolutif. A partir d'une configuration initiale, il ajuste au sens des moindres carrés, le modèle (63) aux données à l'aide d'une application d'une méthode d'optimisation non linéaire de type quasi-Newton, dont la justification est limitée aux cas où  $d_{ij}(Z) \neq 0$  pour tout  $i \neq j$  et donc lorsque le gradient de  $d(Z)$  en fonction de  $Z$  existe. Cette méthode d'approximation justifierait en fait le classement de la méthode de Saito dans la classe des problèmes de Codage Multidimensionnel sous contraintes de structure (cf. d'AUBIGNY(1987)) et non dans la classe des modèles d'Analyse en Composantes. Observons que l'on a de façon formelle

$$\Delta_u = \Delta - \underline{u} \underline{1}' - \underline{1} \underline{u}' + 2D_u = \Delta + A_u \quad (64)$$

et donc on peut tenter de généraliser l'approche de Caillez à cette classe plus générale de problèmes. On réécrit pour cela (63) sous la forme

$$H_u = \Delta_u * \Delta_u = H + 2\Delta * A_u + A_u * A_u$$

où  $A_u * A_u$  (resp.  $\Delta * A_u$ ) est la matrice de  $\mathcal{Y}_H$  de terme général  $(u_i + u_j)^2$ , (resp.  $\delta_{ij}(u_i + u_j)$ ). Par suite

$$\begin{aligned} W_u &= W_2 - Q_1 (\Delta * A_u)' Q_1 - \frac{1}{2} Q_1 (A_u * A_u)' Q_1 \\ &= W_2 + W_1^u + \frac{1}{2} B_u \end{aligned}$$

où la matrice

$$W_1^u = Q_1 (\Delta^* A_u)^t Q_1$$

s'exprime en fonction des  $\delta_{ij}$  et des  $u_i$  alors que la matrice

$$B_u = Q_1 (A_u^* A_u)^t Q_1$$

s'exprime en fonction des  $u_i$  seulement. On vérifie de plus que  $B_u$  a pour terme général

$$b_{ij} = -\bar{u} (u_i + u_j - \bar{u})$$

où

$$\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i$$

Remarquons que si nous normons le vecteur  $\underline{v}$  de sorte que  $\bar{v} = 1$ , en posant  $\underline{u} = c\underline{v}$ , on obtient, mutatis mutandis, les équations précédentes sous la forme

$$W_v = W_2 + W_1^v + \frac{1}{2} c^2 B_v$$

La détermination de la constante  $c$  pour un vecteur  $\underline{v}$  donné, suit une voie analogue à CAILLEZ (1983). On pose

$$c^* B_v \underline{y} = 2W_2 \underline{x}$$

et de façon analogue à Caillez, on obtient, pour  $\underline{v}$  connu, la constante additive  $c^*$  en résolvant l'équation

$$\begin{pmatrix} 0 & 2W_2 \\ -B_v & -4B_v W_1^c \end{pmatrix} \begin{pmatrix} \underline{y} \\ \underline{x} \end{pmatrix} = c \begin{pmatrix} B_v & 0 \\ 0 & B_v \end{pmatrix} \begin{pmatrix} \underline{y} \\ \underline{x} \end{pmatrix}$$

Comme Caillez, on déduit de cette identité que  $\begin{pmatrix} \underline{y} \\ \underline{x} \end{pmatrix}$  est la solution d'un problème aux valeurs propres associée à la plus grande valeur propre  $c^*$ . De plus, pour  $\underline{x}$  et  $\underline{y}$  connus, on déduit  $\underline{v}$  comme solution du problème linéaire

$$\begin{cases} c^* B_v \underline{y} = 2W_2 \underline{x} \\ \bar{v} = 1 \end{cases}$$

Ce calcul formel montre qu'une solution itérative est nécessaire à la construction du vecteur  $\underline{u}^* = c^* \underline{v}^*$ , solution du problème posé par Saito. Le théorème de Le Calvé fournit de plus une initialisation de l'algorithme. Nous n'avons pas réalisé de comparaisons numériques entre cette approche et la méthode de Saito, à ce jour. Il doit cependant être noté que notre formulation s'inscrit dans une logique d'Analyse en Coordonnées

Principales et recourt à la transformation  $\alpha_1$ , alors que Saito décrit un algorithme d'approximation directe des dissimilarités dans la tradition du Codage Multidimensionnel.

#### V.2.2.2. Perturbation de type 2

Posons

$$h_{ij}(c) = \delta_{ij}^2(c) \quad , \quad h_{ij} = \delta_{ij}^2$$

c'est à dire

$$\begin{aligned} \delta_{ij}^2(c) &= \delta_{ij}^2 + c(1 - \delta_i^j) \\ \Delta^{*2}(c) &= \Delta^{*2} + c(\mathbf{1}\mathbf{1}^t - \mathbf{I}) \end{aligned}$$

Ce deuxième type de perturbation est attribué par LINGOES (1971) et par de LEEUW et HEISER (1982) à Guttman. Ces auteurs montrent de plus le résultat suivant (cf. aussi SCHÖNEMAN (1971) :

**Théorème 8:** Soit  $(I, \delta)$  un espace quasi-métrique, où  $I$  est de cardinal fini égal à  $n$ . Alors il existe un réel  $c$  tel que  $(I, \delta(c))$  admet un plongement isométrique Euclidien dans un espace affine Euclidien de dimension au plus égale à  $(n-2)$  où pour tout couple  $(i, j)$  d'éléments de  $[n]$

$$\delta_{ij}^2(c) = \delta_{ij}^2 + c(1 - \delta_i^j) \quad (65)$$

**Preuve :** L'expression (65) permet d'écrire la matrice  $W(c) = \alpha_1(H(c))$  sous la forme

$$W(c) = W + \frac{1}{2} c Q_1$$

Elle a pour valeurs propres

$$\lambda(W(c)) = \lambda(W) + \frac{1}{2} c$$

Par suite choisissons  $c = -2 \lambda_{\text{Min}}(W)$ . La matrice  $W(c)$  est alors semi-définie positive et de rang

$$\text{rang}(W(c)) \leq \text{rang}(W) - 1 \leq (n-1) - 1 \quad \parallel$$

Cette approximation réalise le minimum des valeurs de  $c$  qui rendent  $\delta(c)$  Euclidienne. Elle est très généralement employée en pratique, puisqu'elle définit en particulier une transformation monotone, c'est à dire telle que tout quadruplet  $(i,j,k,l)$  d'éléments de  $[n]$  tels que  $i \neq j$  et  $k \neq l$  vérifie

$$(\delta_{ij}(c) \leq \delta_{kl}(c)) \Leftrightarrow (\delta_{ij} \leq \delta_{kl})$$

Cette propriété, d'un intérêt central en psychométrie (cf. SCHÖNEMAN (1971)) permet de vérifier le résultat suivant (cf. de LEEUW et HEISER (1982))

**Théorème 9 :** *Soit  $(I, \delta)$  un espace quasi-métrique, où  $I$  est de cardinal fini et égal à  $n$ . Alors il existe une dissimilarité  $\delta^*$  telle que*

*i) pour tout quadruplet  $(i,j,k,l)$  d'éléments de  $[n]$  tels que  $i \neq j$  et  $k \neq l$*

$$(\delta_{ij} \leq \delta_{kl}) \Leftrightarrow (\delta_{ij}^* \leq \delta_{kl}^*)$$

*ii)  $(I, \delta^*)$  admet un plongement isométrique Euclidien dans un espace affine Euclidien de dimension au plus égale à  $(n-2)$*

de LEEUW et HEISER (1982) étendent le théorème 9 au cas de tableaux incomplets résultant soit de l'existence de données manquantes, soit d'absences structurelles de mesure, comme c'est le cas pour le problème de déploiement multidimensionnel. De plus LINGOES (1971) remarque que  $\delta(c)$  (resp.  $\delta^*$ ) construit dans le théorème (8) (resp. (9)) peut vérifier pour  $i \neq j$  la condition

$$\delta_{ij}(c) = 0 \text{ (resp. } \delta_{ij}^* = 0 \text{ )}$$

et donc la condition  $i \neq j$  et  $k \neq l$  ne peut pas être levée dans le théorème (9). Ce problème est traité par HOLMAN (1972), dont le travail est partiellement fondé, d'après de LEEUW et HEISER (1982), sur des travaux non publiés de L.M. Kelly. De plus HOLMAN (1972) démontre le résultat suivant

**Théorème 10:** *Soit  $(I, \delta)$  un espace quasi-métrique, où  $I$  est de cardinal fini égal à  $n$ .*

*i) si  $\delta$  est une fonction de dissimilarité ultramétrique sur  $I$ , alors  $(I, \delta)$  admet un plongement isométrique dans un espace affine Euclidien de dimension égale à  $(n-1)$ ,*

*ii) si  $\delta$  n'est pas ultramétrique, alors il existe une fonction de dissimilarité  $\delta^*$  définie sur  $I$  vérifiant la condition de monotonie*

$$(\delta_{ij} \leq \delta_{kl}) \Leftrightarrow (\delta_{ij}^* \leq \delta_{kl}^*) \quad \forall (i,j,k,l) \in [n]^4$$

*et telle que  $(I, \delta^*)$  admet un plongement isométrique dans un espace affine Euclidien de dimension au plus égale à  $(n-2)$ .*

Le résultat de Holman montre donc que les espaces ultramétriques fournissent une situation intermédiaire entre les dissimilarités Euclidiennes représentables dans des espaces affines de dimension au plus égale à  $(n-2)$  et les espaces quasi-métriques qui n'admettent pas de plongement Euclidien. C'est par exemple le cas des données de dissimilarité associées à des espaces de Banach de type  $L_p$  qui, comme nous l'avons vu plus haut, n'admettent pas de plongement isométrique dans un espace affine Euclidien de dimension finie.

Le modèle (63) admet un équivalent dans le cadre de la perturbation de type 2, défini par

$$\begin{aligned} \Delta^{*2}(\underline{u}) &= \Delta^{*2} - \underline{u}^2 \mathbf{1} - \mathbf{1} \underline{u}^2 + 2D_u^2 \\ \Leftrightarrow \delta_{ij}^2(\underline{u}) &= \delta_{ij}^2 - u_i^2 - u_j^2 + 2 u_i^2 \delta_i^j \end{aligned} \quad (66)$$

où  $\delta_{ij}^2(\underline{u})$  est supposée Euclidienne, de sorte que (66) s'écrit encore

$$\begin{aligned} \delta_{ij}^2 &= \delta_{ij}^2(Z) + u_i^2 + u_j^2 - 2 u_i^2 \delta_i^j \\ &= \sum_{s=1}^r (z_i^s - z_j^s)^2 + u_i^2 + u_j^2 - 2 u_i^2 \delta_i^j \\ &= \sum_{s=1}^r (z_i^s - z_j^s)^2 + \sum_{s=r+1}^{r+n} (u_i^s - u_j^s)^2 \end{aligned} \quad (66\text{bis})$$

où l'on a posé  $u_i^s = u_i \delta_i^j$  pour  $s = r+j$  ( $j \in [n]$ ). Le modèle (66) s'écrit de façon synthétique

$$\delta_{ij}^2 = \delta_{ij}^2(Y) = \sum_{s=1}^{r+1} (y_i^s - y_j^s)^2$$

où la matrice  $Y$  est un élément de  $\mathcal{M}_{n+r,n}(\mathbb{R})$  de structure particulière

$$Y = \begin{pmatrix} Z \\ D_u \end{pmatrix} \quad Z \in \mathcal{M}_{r,n}(\mathbb{R}) \quad (67)$$

Les équations (66bis) et (67) caractérisent un modèle d'Analyse Factorielle où  $Z$  décrit les facteurs communs et  $D_u$  décrit les facteurs spécifiques, de sorte que lorsque  $M$  est la matrice identité on vérifie l'identité

$$W_Y = W_Z + D_u^2 = \alpha_Y(H) \quad (68)$$

où  $H = \Delta^{*2}(\underline{u})$ . On obtient donc un modèle de Q-Analyse Factorielle en Facteurs Communs et Spécifiques, et  $D_u^2$  est une matrice diagonale à termes positifs lorsqu'on

suppose, comme ici, que  $u_i$  est réel pour tout  $i$  de  $[n]$ .  $D_u^2$  s'interprète alors comme une matrice d'unicités

**Remarque :** Le modèle s'étend de façon immédiate au cas où  $M$  est une matrice semi-définie positive du type

$$M = \begin{pmatrix} Q & 0 \\ 0 & D_\gamma^2 \end{pmatrix}$$

où  $D_\gamma^2$  est une matrice diagonale d'ordre  $n$  à termes positifs. Alors  $D_u^2$  a pour terme général  $\gamma_i^2 u_i^2$ .

Le modèle (66) a été étudié par CARROLL (1983) et par WINSBERG et CARROLL (1984) qui proposent un algorithme itératif permettant d'ajuster, au sens des moindres carrés, ce modèle à une matrice de dissimilarité observée  $\Delta$ . Comme pour le modèle de Saito, il s'agit là, à proprement parlé, d'un modèle de Codage Multidimensionnel. Cependant, dans les deux cas, l'imposition d'une contrainte supplémentaire

$$u_i = u \quad \forall i, i \in [n]$$

permet de retrouver le problème de la constante additive correspondant. COOPER (1972) propose dans ce cas un algorithme résolutif, qui s'applique directement à l'ajustement d'une matrice de dissimilarité. De plus, sous cette contrainte, l'équation (68) s'écrit

$$\alpha_\gamma(H) = W = W_Z + 2u^2 I$$

On reconnaît ainsi l'équation caractéristique de la perturbation d'une matrice de produits scalaires des prédicteurs utilisée en régression ridge, de paramètre de perturbation  $k = 2u^2$  (HOERL et KENNARD (1970), DEMPSTER et al. (1977), VINOD (1978), BELSEY et al. (1980) dans le cadre peu classique d'une Q-Analyse.

Ainsi le problème de la constante additive apparaît comme une extension des stratégies de perturbations de la matrice de covariance développées en régression linéaire multiple et destinées à protéger l'analyse des instabilités liées à l'existence de colinéarités entre les variables facteurs en recourant à des estimateurs biaisés. En Analyse en Coordonnées Principales, l'extension de cette technique permet de se prémunir contre l'apparition de coordonnées imaginaires pures.

Le cas général caractérisé par une matrice  $D_u^2$  non proportionnelle à l'identité correspond en régression linéaire multiple aux méthodes bayésiennes d'estimation des paramètres, exposées entre autres par ZELLNER (1981) et LEAMER (1973, 1978).

Un cas particulier du modèle (63) est étudié par FICHET (1983). Cet auteur propose de construire la solution du problème de la constante additive

$$\text{Min } \{c ; \delta(c) \in \mathfrak{D}_E(I)\}$$

où  $\delta(c)$  est la fonction de dissimilarité Euclidienne définie dans nos notations par

$$\delta_{ij}(c) = (\delta_{ij}^2 + c \left( \frac{1}{\gamma_i} + \frac{1}{\gamma_j} \right))^{1/2} \delta_i^j$$

où  $c$  est un réel positif et le vecteur  $\gamma$  de  $F^*$  représente une pondération réelle des unités statistiques. Fichet suppose de plus que  $\gamma$  est donné à priori, de sorte que la méthode a le statut d'une méthode d'Analyse en Coordonnées Principales particulière, une fois  $c$  déterminée.

**Remarque :** Les équations (63) et (65) ne sont pas équivalentes en général. Elles le deviennent si les paramètres  $u$  vérifient de plus

$$\frac{1}{\delta_{ij}(c)} = \frac{1}{u_i} + \frac{1}{u_j}$$

pour tout couple  $(i,j)$  d'éléments de  $I$  pour lequel cette expression est définie.

Un dernier modèle de constante additive a été proposé par HOANG M. THU (1978) qui propose un algorithme résolutif. Il s'écrit

$$\delta_{ij}(c) = (\delta_{ij}^c)^c + \delta_i^j$$

La justification usuellement accordée aux méthodes d'évaluation de la constante additive est la conservation des informations ordinales sur  $[n] \times [n]$  induites par les tableaux de dissimilarité. Il s'agit là en fait d'un argument très contestable lorsque les objectifs de l'analyse impose une phase de réduction de dimension de l'espace de représentation des données, car la translation des valeurs propres de  $W(c)D_p$  conduit à négliger les dimensions initialement associées à des valeurs propres négatives de  $WD_p$  où  $W = \alpha_p(H)$ .

De notre point de vue, le problème de la constante additive doit être considéré comme une heuristique puisqu'il n'y a pas d'homogénéité de critère d'optimisation entre les phases de construction de  $c^*$  et de construction du plongement isométrique de  $(I, \delta(c^*))$  dans  $E$ . Son utilisation est difficile à justifier lorsque  $\delta$  est mesurée sur une échelle de rapport. Lorsque

$\delta$  est mesurée sur une échelle d'intervalles, comme il est souvent affirmé en psychométrie, cette critique doit être nuancée. Néanmoins, il nous semble généralement préférable de définir l'Analyse en Coordonnées Principales comme la recherche des éléments propres de  $WD_pWD_p$ , éventuellement suivie d'une phase de réduction de dimension. L'obtention d'une approximation bilatérale des dissimilarités qui résulte de cette analyse dans le cas d'une matrice  $W$  non gramienne, nous semble de nature à conforter notre point de vue. On note qu'alors pour les perturbations de type 1 et 2

$$WD_pWD_p \neq W(c)D_pW(c)D_p$$

L'analogie avec les méthodes de ridge regression montre cependant que le calcul de  $W(c)$  n'est pas justifié puisque c'est la matrice de connexion qu'il convient de perturber, afin de protéger la régression contre les instabilités dues aux relations de colinéarité et non le tableau de pseudo-données  $W$ . On introduit en effet dans le cas d'une ridge regression une perturbation de la forme

$$W'(c) = WD_pW + cI$$

Puisque  $WD_pW$  est gramienne par construction, l'intérêt d'une telle perturbation est limitée à son rôle classique en régression linéaire multiple.

Le cadre bayésien s'étend de façon analogue en posant

$$W(u) = WD_pW + D_u^2$$

Il fournit une classe de méthodes d'Analyse Factorielle du triple  $(I, \delta, P_I)$  qui ne semble pas avoir été étudié à ce jour.



## CHAPITRE VI

### L'ANALYSE DU TRIPLE $(I, \delta, Q_{I \times I})$

L'étude qui précède concerne le triple  $(I, \delta, P_1)$  où  $P_1$  est une pondération définie sur  $I$ . Nous traitons ici le cas général d'une pondération  $Q_{I \times I}$  définie sur  $I \times I$ , c'est à dire une pondération agissant sur les cases du tableau de dissimilarité  $\Delta$  observé. Ce cadre d'étude est homogène à l'hypothèse probabiliste qui interprète  $\Delta$  comme la réalisation d'une matrice aléatoire et fait donc référence à un cadre formel plus riche que l'analyse précédente. Du point de vue mathématique, on est alors confronté à un problème plus complexe, puisque la simplicité des traitements présentés dans le chapitre précédent, résulte en fait de l'équivalence démontrée entre la donnée de  $P_1$  et celle d'une origine  $P_0$  dans l'espace affine Euclidien de représentation  $E$ .

En pratique, cette complexité accrue a pour contrepartie une flexibilité et une puissance accrues du fait de possibilités de traitements nouveaux offertes telles que la prise en compte des données manquantes. Outre les considérations d'ordre statistique, telles que la recherche des solutions stables par la protection de la méthode contre les données extrêmes neutralisées par un choix de pondération adaptée, d'AUBIGNY (1987) montre de plus que le recours à une pondération générale permet de prendre en compte des contraintes structurelles imposées aux dissimilarités entre unités statistiques, telles que les contraintes d'appartenance à une variété différentiable ou les contraintes de contiguïté, qui donnent lieu aux méthodes d'Analyse en Coordonnées Locales, analogues aux méthodes d'Analyse en Composantes Locales mises à l'honneur par LEBART (1969) et LEBART et BARET (1984).

La formulation du problème en termes d'approximation bilatérale de la fonction de dissimilarité observée par une fonction de distance Euclidienne conduit à ce que GREENACRE (1978) appelle une approximation forte de  $\delta$ . Dans un tel cadre d'analyse, la lecture des représentations de  $I$  est métrique, c'est à dire fondée sur la lecture des distances interpoints alors que nous avons vu précédemment que les méthodes d'analyse fondées sur l'utilisation explicite des transformations  $\alpha_p$  et  $\beta_p$  fournissent des approximations faibles des dissimilarités et conduisent à des lectures hilbertiennes fondées sur les notions d'angles.

Afin de marquer cette différence d'objectif, nous réservons le terme de Codage Multidimensionnel, traduction de l'expression anglaise Multidimensional Scaling, aux méthodes fondées sur l'approximation des dissimilarités. En France, BENZECRI (1965) parle d'Analyse des Proximités et ESCOUFIER (1975) parle de positionnement multidimensionnel, mais nous ne suivons pas ces conventions.

De nombreux articles en langue anglaise ont été consacrés à ces méthodes depuis les travaux de SHEPARD (1962a, b) et de KRUSKAL (1964a, b) qui fournissent les premiers résultats faisant la preuve de la possibilité de résoudre ce problème en produisant des algorithmes résolutifs. Citons en particulier les travaux de BORG (1981a), DAVISON (1983), GREEN et CARMONE (1970), GREEN et RAO (1972), GREEN et WIND (1973), GREENACRE (1978), HEISER (1981), HARTMAN (1979), KRUSKAL et WISH (1978), de LEEUW (1968), 1970a, b), MEULMAN (1986), SCHIFFMAN et al. (1981) qui ont beaucoup contribué à la dissémination de ces méthodes dans l'ensemble des disciplines en Sciences Sociales. Le lecteur intéressé peut de plus consulter GOLLEDGE et RAYNER (1982), LINGOES et al. (1977), LANTERMAN et FEGER (1980), SHEPARD et al. (1972). De plus ces méthodes font depuis peu leur entrée dans les ouvrages pédagogiques ou de recherche consacrés à l'analyse multivariée, comme par exemple MARDIA et al. (1979), DILLON et GOLDSTEIN (1984), TAKEUCHI et al. (1982), LEDERMANN (1984), KRISHNAIAH et KANAL (1982) et surtout GIFI (1981) qui présente une étude très argumentée du statut accordé aux différentes méthodes d'analyse multidimensionnelle dans les ouvrages de Statistique. Une bibliographie plus complète sur le sujet est produite dans les ouvrages cités en référence et dans NISHISATO (1978) et BICK et al. (1978).

En France, ce domaine d'étude a fait l'objet d'un nombre limité de travaux parmi lesquels on peut citer BENZECRI (1964, 1965), DAVY de VIRVILLE (1972), d'AUBIGNY (1975), IGLESIAS (1975), BONIFAS et al. (1978), BERTRANEU (1979), MAILLES (1979), BENINEL (1987) et d'AUBIGNY (1984) dont ce mémoire est une version révisée et complétée.

En Union Soviétique, les articles de TEREKHINA (1973) et de KAMENSKI (1977, 1979a, b) montrent un intérêt pour ces méthodes, alors qu'au Japon une tradition originale s'est développée autour des travaux de HAYASHI (1952, 1964, 1967, 1974, 1975) bien que ces recherches semblent avoir été peu remarquées aux Etats Unis à l'époque.

Notre objectif n'est pas de présenter ici un panorama exhaustif des méthodes mais plutôt d'explicitier les liens et les différences existant entre les méthodes d'Analyse en Coordonnées, usuellement décrites dans F, et les méthodes de Codage Multidimensionnel,

que nous décrivons dans  $F \wedge F$ , dans le cadre du formalisme algébrique présenté plus haut (§IV.7).

Nous parlons de Codage Multidimensionnel Ordinal lorsque la fonction de dissimilarité est mesurée sur une échelle ordinale et nous opposons ces méthodes au Codage Multidimensionnel Métrique qui traite des fonctions de dissimilarité mesurées sur une échelle de rapport ou d'intervalles. Bien que les méthodes étudiées ici aient été développées originellement dans un cadre ordinal (sous de nom de Nonmetric Multidimensional Scaling dans la littérature en langue anglaise), nous exposons dans ce chapitre le cas métrique, qui nous permet de fixer les outils algébriques définissant un cadre d'étude suffisamment général pour rendre compte des méthodes les plus connues et de montrer que les simplifications résultant du choix d'une pondération Produit  $Q_{|X|} = P_I \otimes P_I$  conduisent aux méthodes d'Analyse en Coordonnées Principales du Triple  $(I, \delta, P_I)$ . Le cas ordinal fait l'objet du chapitre suivant, dans lequel nous montrons les relations étroites qui unissent les méthodes de Codage Multidimensionnel Ordinal aux outils classiques de la Statistique non paramétrique.

## VI.1. RAPPELS ET DÉFINITIONS

Supposons donné le triplet  $(I, \delta, Q_{|X|})$  où  $I$  est un ensemble de cardinal fini égal à  $n$ ,  $(I, \delta)$  est un espace quasi-métrique et  $Q_{|X|}$  une mesure réelle définie sur  $|X|$  et on suppose de plus donnée une bijection de  $I$  dans  $[n]$ . La référence à une problématique statistique nous conduit à supposer de plus que  $Q_{|X|}$  est une mesure finie et positive. De plus, la fonction de dissimilarité  $\delta$  est par définition une fonction symétrique sur  $|X|$  et nous supposons donc ici que  $Q_{|X|}$  est aussi symétrique. Le cas général ne semble pas avoir été considéré dans les applications qui supposent  $\delta$  symétrique et l'intérêt des pondérations de signe quelconque semble limité au cas de l'analyse de dissimilarités non symétriques (cf. de LEEUW et HEISER (1982), GOWER (1977), CONSTANTINE et GOWER (1978), CHINO (1978), ESCOUFIER et GRORUD (1980), YOUNG (1975)).

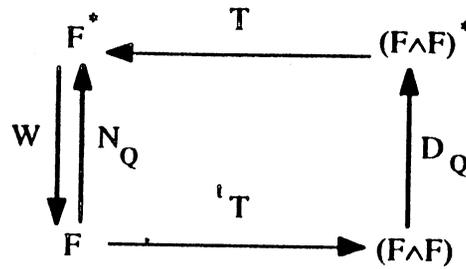
Lorsque  $Q_{|X|}$  est positive et symétrique, la forme bilinéaire symétrique  $q$  définie sur  $F \times F$  par

$$q(f_i, f_j) = q_{ij} \quad i, j \in [n]$$

permet de construire une forme quadratique diagonale définie sur  $F \wedge F$ , de matrice

$$D_Q = \text{diag} (q_{ij})$$

dans la base  $\{f_i \wedge f_j ; i < j \quad i, j \in [n]\}$  de  $F \wedge F$ . La dualité entre  $F_I$  et  $F \wedge F$  est alors résumée par le schéma algébrique suivant



$T$  définit une isométrie entre  $F_1$  et le sous-espace  $'T(F_1)$  de  $F \wedge F$ , lorsque  $F$  est muni de la métrique Euclidienne de matrice

$$N_Q = 'TD_Q T$$

de terme général

$$n_{ij} = \begin{cases} -q_{ij} & \text{si } i \neq j \\ \sum_{k \neq j} q_{ik} = q_{i+} & \text{si } i = j \end{cases}$$

Nous avons montré à l'aide du lemme de Coxeter que  $N_Q$  est une matrice de Gram de rang  $(n-1)$  telle que le vecteur  $\underline{1}_F$  est un élément du noyau  $\text{Ker}(N_Q)$ .  $N_Q$  s'écrit de façon alternative :

$$N_Q = D_q - Q$$

où  $D_q$  est la matrice diagonale de terme général  $q_{i+}$ .

Cette écriture est préférée dans la suite à l'expression

$$N_Q = \sum_{i < j} \sum q_{ij} (\underline{f}_i - \underline{f}_j) '(\underline{f}_i - \underline{f}_j)$$

qui résulte des relations de dualité induites par l'isométrie  $T$

**Remarque :** Le lemme de Coxeter semble être passé inaperçu pour les statisticiens et psychomètres, qui l'ont découvert de façon répétitive. On en trouve sans référence une démonstration analytique dans HAYASHI (1952), GUTTMAN (1968) et KAMENSKI (1976). Le lien ne semble pas avoir été fait auparavant avec les propriétés d'isométrie partielle de  $'T$  entre  $F_1$  et  $F \wedge F$ .

Il est d'autre part intéressant de noter que ce lemme intervient dans la recherche et le dénombrement des arborescences dans un graphe. BERGE (1973) appelle arborescence un arbre (c'est à dire un graphe sans cycles) muni d'une racine. Soit donc un graphe  $G = (I, U)$ . Si  $G$  est un graphe simple complet, on peut interpréter  $'T$  comme la matrice d'une application linéaire mettant en dualité l'espace  $F$  de représentation des  $n$  sommets du graphes et l'espace  $F \wedge F$  de représentation des  $n^*$  arêtes du graphe.

Considérons maintenant un graphe quelconque, non nécessairement simple. BERGE (1973) considère la matrice  $Q$  de terme général  $q_{ij}$  égal au nombre d'arêtes de  $G$  reliant  $i$  et  $j$  et la matrice diagonale  $D_q$  de terme général  $q_{i+}$  égal au demi-degré incident, c'est à dire au nombre d'arêtes ayant une extrémité égale à  $i$ . Cet auteur construit alors la matrice  $N_Q = D_q - Q$  dont le déterminant est nul puisque  $N_Q$  est de rang  $(n-1)$  et il calcule de déterminant de la matrice  $\tilde{N}_Q$  obtenue à partir de  $N_Q$  en supprimant la dernière ligne et la dernière colonne (en fait Berge choisit la première) de  $N_Q$ . BERGE (1973) montre alors le lemme suivant :

**Lemme 1:** *Soit  $G = (I,U)$  un graphe sans boucles, de matrice associée  $Q$ . Supposons de plus que  $U$  est formé de  $(n-1)$  arêtes. Alors  $G$  est une arborescence de racine  $n$  si et seulement si  $\det(\tilde{N}_Q) = 1$ . Dans tous les autres cas, le déterminant est nul.*

De plus, Berge reprend le résultat suivant dû à TUTTE (1948), dans lequel le choix de l'élément de  $I$  noté  $n$  dépend de la bijection de  $I$  dans  $[n]$  retenue

**Théorème 1:** *Soit  $G = (I,U)$  un graphe de matrice associée  $Q$ . Alors le nombre d'arborescences de racine  $n$  qui sont des graphes partiels de  $G$  est égal à  $\det(\tilde{N}_Q)$ .*

Remarquons que pour tout graphe  $G = (I,U)$ , on construit  $Q$  comme une mesure de comptage. De plus, si  $G$  est un graphe simple,  $q_{ij} \in \{0,1\}$  et  $q_{i+}$  est le demi-degré incident du sommet  $i$ . On vérifie alors

**Corollaire :** *Si  $G = (I,U)$  est un graphe simple, le nombre d'arbres maximaux de  $G$  est égal au mineur de  $N_Q$ , relatif à n'importe quel coefficient de la diagonale principale.*

Ces résultats montrent l'intérêt des outils algébriques mis en place dans le cas où  $Q$  est une mesure de comptage, susceptible donc d'être interprétée comme le nombre d'arêtes liant deux sommets quelconques d'un graphe non simple. Revenons donc maintenant à l'étude d'une pondération quelconque, et des constituants du schéma de dualité présenté ci-dessus.

Si  $W = \alpha_1(H)$  et si  $\Psi = {}^tTWT$ , alors  $\Psi$  est la matrice des contrastes  $2 \times 2$  de terme général

$$\begin{aligned} \Psi_{ij,kl} &= (w_{ik} - w_{il} - w_{jl} + w_{jk}) \\ &= -\frac{1}{2} (h_{ik} - h_{il} - h_{jl} + h_{jk}) \end{aligned}$$

**Lemme 2:** *Soit  $(I, \delta, Q_{|x|})$  un triplet tel que  $(I, \delta)$  est un espace quasi-métrique fini et  $Q_{|x|}$  une mesure positive définie sur  $I \times I$  et symétrique. Alors*

$$\sigma(H) = \sum_{i < j} \sum q_{ij} h_{ij} = \text{Trace}(\Psi D_Q) = \text{Trace}(W N_Q)$$

où  $H$  est une fonction de  $\Delta$  et

$$W = -\frac{1}{2} Q_1 H^1 Q_1 = \alpha_1(H)$$

**Preuve :**

$$\text{Trace}(\Psi D_Q) = \sum_{i < j} \sum q_{ij} h_{ij}$$

s'écrit en fonction de  $N_Q$

$$\text{Trace}(\Psi D_Q) = \text{Trace}({}^t W D_Q) = \text{Trace}(W D_Q {}^t) \text{Trace}(W N_Q) \quad \parallel$$

Cette relation s'écrit de façon analytique

$$\sum_{i < j} \sum q_{ij} h_{ij} = \sum_{i=1}^n \sum_{j=1}^n n_{ij} w_{ij} = \sum_{i=1}^n (q_{i+} w_{ij} - \sum_{j \neq i} q_{ij} w_{ij})$$

puisque

$$\text{Trace}(W N_Q) = \text{Trace}(W D_Q) - \text{Trace}(W Q)$$

En particulier si  $H = \Delta * \Delta$ , la relation s'écrit

$$\sigma(H) = \text{Trace}(\Psi D_Q) = \text{Trace}(W N_Q) = \sum_{i < j} \sum q_{ij} \delta_{ij}^2$$

d'AUBIGNY (1988) montre que les résultats de RAO (1982) s'appliquent alors à cette situation et permettent d'interpréter  $\sigma$  comme une quantité d'information quadratique (on parle aussi d'énergie en Physique). La recherche d'un plongement Euclidien isométrique approché de  $(I, \delta, Q_{|x|})$  dans un espace affine Euclidien  $E$  s'interprète alors comme la recherche d'une décomposition additive de la quantité d'information  $\sigma$ , qui fournit une définition des méthodes de Codage Multidimensionnel analogue à la présentation de l'Analyse en Composantes Principales (resp. de l'Analyse des Correspondances) comme une décomposition additive de la variance totale du tableau de données (resp. du CHI-deux calculé sur le tableau de contingence). d'AUBIGNY (1988) montre de plus que les méthodes de Codage Multidimensionnel s'interprètent alors comme des techniques de description du potentiel d'interaction décrit par le tableau  $\Delta$  dans le cadre d'un plan d'expérience croisant un facteur avec lui-même. D'autre part, lorsqu'on dispose d'une bipartition de  $I$  en  $I = J \cup K$ , telle que  $q_{ij} = 0$  si  $i$  et  $j$  appartiennent à la même classe. Il existe alors une bijection de  $I$  dans  $[n]$  induisant les décompositions suivantes de  $Q$  et de  $N_Q$

$$Q = \begin{pmatrix} 0 & Q_{JK} \\ {}^t Q_{JK} & 0 \end{pmatrix} \quad N_Q = \begin{pmatrix} D_J & -Q_{JK} \\ -{}^t Q_{JK} & D_K \end{pmatrix}$$

où  $D_J$  (resp.  $D_K$ ) est la matrice diagonale des marges en lignes (respectivement en colonnes) de  $Q_{JK}$ . On reconnaît en  $N_Q$  la matrice étudiée en Analyse des Correspondances binaires, dans le cadre d'une  $\Sigma$ -stratégie, lorsque  $Q_{I \times I}$  est une mesure de comptage. De plus, lorsque  $H = \Delta * \Delta$ ,  $\sigma$  s'écrit pour une telle partition :

$$\sigma(H) = \sum_{i < j} \sum q_{ij} \delta_{ij}^2 = \sum_{j \in J} \sum_{k \in K} q_{jk} \delta_{jk}^2$$

et fournit la quantité d'information quadratique usuellement décomposée par les méthodes de Déploiement Multidimensionnel (cf. d'AUBIGNY (1981), de LEEUW et HEISER (1981)).

Notons en effet  $D(X, Y)$  la matrice de terme général

$$d_{jk}(X, Y) = \left( \sum_{s=1}^r (x_j^s - y_k^s)^2 \right)^{1/2}$$

et  $R$  la matrice des résidus de l'ajustement du tableau rectangulaire de dissimilarité  $\Delta_{JK}$  par la matrice  $D(X, Y)$

$$R = \Delta_{JK} - D(X, Y).$$

Alors

$$\sigma(R) = \sigma(\Delta, D(X, Y)) = \sum_{j \in J} \sum_{k \in K} q_{jk} (\delta_{jk} - d_{jk}(X, Y))^2$$

de LEEUW et HEISER (1982) étendent cette méthodologie au cas où il existe une bipartition  $JUK$  de  $I \times I$  et considèrent  $J$  comme la partie de  $I \times I$  constituée par les données manquantes. Ils montrent alors l'existence d'un plongement isométrique Euclidien de  $I$  dans un espace affine de dimension au plus égale à  $(n-1)$ , sous des hypothèses analogues au cas sans données manquantes, pour une dissimilarité  $\delta^*$  définie comme la transformée monotone de  $\delta$ .

LE FOLL (1976) adopte une stratégie inverse de la nôtre pour l'étude des données spatio-temporelles. Dans ce cas, il mène une "Analyse des évolutions" en posant à partir d'une pondération  $P_I$  définie sur  $I$

$$q_{ij} = \begin{cases} p_i p_j & \text{si } i \text{ et } j \text{ appartiennent au même site} \\ 0 & \text{sinon} \end{cases}$$

de façon à ignorer la part d'inertie du nuage  $\{(N_i, p_i) ; i \in [n]\}$  d'origine temporelle, lors de la recherche du sous-espace  $E_r$  de  $E$  de dimension  $r < p$  rendant  $\sigma_r(\Delta^* \Delta)$  maximal

$$\sigma_r(\Delta^* \Delta) = \sum_{i < j} q_{ij} \delta_{ij}^2(Z_r)$$

où  $Z_r$  décrit l'ensemble des représentations de rang  $r$  de  $I$  dans  $E$ .

Soit  $(I, \delta, Q_{I \times I})$  un triplet respectant les contraintes définies ci-dessus. Le plongement isométrique de l'espace quasi-métrique  $(I, \delta)$  admet une résolution formelle proche de celle de l'Analyse en Coordonnées Principales. Cependant nous montrons ci-dessous que la présence d'une pondération  $Q_{I \times I}$  générale rend cette solution formelle inopérante en pratique. D'autre part, lorsque  $Q_{I \times I}$  n'est pas de la forme  $P_I \otimes P_I$ , où  $P_I$  est une pondération définie sur  $I$ , l'analogie avec les outils de la mécanique fondés sur la notion d'inertie disparaît. C'est pourquoi nous présentons tout d'abord dans le cas unidimensionnel une analogie avec les problèmes rencontrés dans l'étude des réseaux électriques.

## VI.2. LE CODAGE UNIDIMENSIONNEL DU TRIPLE $(I, \delta, Q_{I \times I})$

Le problème traité par les méthodes de Codage Unidimensionnel du triple  $(I, \delta, Q_{I \times I})$  est en général envisagé en termes combinatoires. DEFAYS (1978) et HUBERT et ARABIE (1985) résolvent le problème à l'aide d'outils combinatoires reposant sur les propriétés spécifiques du cas unidimensionnel telles que l'existence d'un ordre total sous-jacent sur l'ensemble des unités statistiques. L'existence d'une solution est alors assurée lorsque les hypothèses du théorème II.4.14 dû à Orth sont satisfaites. Nous adoptons ici un point de vue différent. Le problème est résolu à l'aide des outils de programmation convexe. Nous ne nous plaçons pas ici du point de vue algorithmique puisque ces outils ne sont pas exempts de faiblesses pour la résolution concrète du cas unidimensionnel, comme précisé par HUBERT et ARABIE (1985), mais leur intérêt théorique est dû au fait qu'ils sont naturellement extensibles au cas multidimensionnel. De plus, ce traitement du problème permet d'en donner une interprétation empruntée à la Physique.

Identifions les éléments de  $I$  aux nœuds d'un réseau électrique. Les arêtes  $(i, j)$  du graphe  $G = (I, U)$  décrivant le réseau représentent des liens directs de résistance  $q_{ij}$ . Un réseau électrique est donc un graphe valué par la résistance des liaisons entre nœuds. Notons de plus  $u_i$  le potentiel électrique au nœud  $i$ . Alors de façon classique, le taux de dissipation d'énergie dans le réseau est proportionnel à la quantité

$$D(\underline{u}, Q) = \sum_{i < j} \sum \frac{(u_i - u_j)^2}{q_{ij}} \quad (69)$$

Si  $\underline{u}$  n'est pas fixé par les contraintes physiques, la dissipation minimale est réalisée lorsque

$$0 = \frac{\partial D(\underline{u}, Q)}{\partial u_i} = \sum_k \frac{(u_i - u_k)}{q_{ik}}$$

$$\Leftrightarrow u_i = \frac{\sum_k (q_{ik})^{-1} u_k}{q_{i+}}$$

puisque  $q_{ii} = 0$ . Reformulons maintenant le problème et considérons  $\delta_{ij}$  le flot courant de  $i$  à  $j$  dans le réseau. La dissipation d'énergie est alors proportionnelle à

$$E(\Delta, Q) = \sum_{i < j} \sum q_{ij} \delta_{ij}^2$$

et l'on doit respecter un équilibre du courant à chaque nœud du réseau. Si  $c_i$  est le courant net fourni au réseau par le nœud  $i$ , alors le problème s'écrit (cf. WHITTLE (1971)):

$$\text{Max } \{-E(\Delta, Q) ; \sum_k \delta_{ik} = c_i \quad i \in [n]\}$$

Le lagrangien associé à ce problème est donc

$$L(\Delta, Q, \underline{u}) = -\sum_{i < j} \sum q_{ij} \delta_{ij}^2 + \sum_i u_i (c_i - \sum_k \delta_{ij})$$

dont l'optimum est atteint lorsque

$$\delta_{ik} = \frac{u_i - u_k}{q_{ik}} \quad (70)$$

L'expression (70) est l'expression de la loi d'Ohm. Le maximum de  $L(\Delta, Q, \underline{u})$  est

$$L^*(\Delta, Q, \underline{u}) = -\sum_{i < j} \sum \frac{(u_i - u_j)^2}{q_{ij}} + \sum_i u_i c_i$$

$$= -\sum_{i < j} \sum q_{ij} \delta_{ij}^2 + \sum_i u_i c_i$$

et le potentiel électrique au nœud  $i$  est proportionnel au taux de variation de la dissipation d'énergie totale résultant du courant introduit au nœud  $i$ .

Imposons maintenant de minimiser la dissipation d'énergie aux seuls nœuds qui sont privés de connexion externe au réseau. Ceux-ci sont tels que  $c_i = 0$  et doivent donc

minimiser  $D(\underline{u}, Q)$  défini en (69). On obtient ainsi deux formulations duales du problème puisque les potentiels électriques  $u_i$ , liés à  $\delta$  par la loi d'Ohm (70) s'interprètent à l'aide de  $L(\Delta, Q, \underline{u})$  comme les multiplicateurs de Lagrange, intervenant dans la minimisation de la dissipation de l'énergie  $E(\Delta, Q)$ .

### VI.3. LE CODAGE MULTIDIMENSIONNEL DU TRIPLE $(I, \delta, Q_{|x|})$

Nous abordons maintenant les méthodes de Codage Multidimensionnel. Dans une première section nous étudions à des fins de comparaison, les tentatives d'extension de l'Analyse en Coordonnées Principales à l'étude du triple  $(I, \delta, Q_{|x|})$ . Puis nous traitons le problème de Codage Multidimensionnel proprement dit.

#### VI.3.1. Outils algébriques d'Analyse du Triple $(I, \delta, Q_{|x|})$

Soit  $\underline{a}$  un vecteur de  $F^*$  de coordonnées  $a_i$  dans la base  $\mathcal{F}^*$ . Nous munissons  $F^*$  de la structure affine déterminée par le choix d'une origine  $O$  et du référentiel  $(O, \{\overrightarrow{ON}_i, i \in I\})$  ou  $N_i$  est l'extrémité du vecteur de base  $f_i^*$ . Munissons  $F^*$  d'une géométrie Euclidienne définie par la forme quadratique  $b$  de matrice  $B$ , supposée de plein rang. Notons de plus  $c_i$  la longueur de la projection du point  $N_i$  sur l'axe  $\Delta_{\underline{a}}$  support de  $\underline{a}$ . Alors

$$\begin{aligned} c_i &= b(\overrightarrow{ON}_i, \underline{a}) = b(\overrightarrow{ON}_i, \sum_{j=1}^n a_j \overrightarrow{ON}_j) = \langle f_i^*, \sum_{j=1}^n a_j f_j^* \rangle_B \\ &= \sum_{j=1}^n a_j b_{ij} \end{aligned}$$

$\underline{c} = B\underline{a}$  est donc une forme linéaire sur  $F^*$ , c'est à dire un élément de  $F$  associé à  $\underline{a}$ . Cherchons donc le vecteur  $B$ -normé  $\underline{a}_1$  tel que  $D(\underline{c}^1, Q)$  soit minimum

$$D(\underline{c}^1, Q) = \sum_{i < j} \sum q_{ij} (c_i^1 - c_j^1)^2$$

**Proposition 1:** *Le vecteur de  $F$  qui réalise*

$$\text{Min } (D(\underline{c}^1, Q); \underline{c} = B\underline{a}, \underline{a} \in F^*, \|\underline{a}\|_B = 1) \quad (71)$$

*est la solution du problème aux valeurs propres*

$$BN_Q \underline{c} = \lambda \underline{c} \quad \|\underline{c}\|_{B^{-1}} = 1$$

*Si  $\underline{c}^1 = B\underline{a}^1$ ,  $\underline{a}^1$  est la solution du problème aux valeurs propres*

$$N_Q B \underline{a} = \lambda \underline{a} \quad \|\underline{a}\|_B = 1$$

**Preuve :** Le Lagrangien associé au problème (71) s'écrit

$$\sum_{i < j} \sum q_{ij} (c_i^1 - c_j^1)^2 + 4 \lambda (1 - \sum_i \sum_j a_i^1 b_{ij} a_j^1)$$

son gradient s'annule lorsque

$$\begin{aligned} \lambda c_k^1 &= \sum_{i < j} \sum p_{ij} (c_i^1 - c_j^1) (b_{ik} - b_{jk}) \\ \Leftrightarrow \lambda c_k^1 &= \sum_i b_{ik} \left( \sum_j c_j^1 (q_{i+} \delta_i^k - q_{ij}) \right) \end{aligned}$$

c'est à dire sous forme matricielle

$$\lambda \underline{c}^1 = B (D_q - Q) \underline{c}^1 = B N_Q \underline{c}^1 \quad (72)$$

De plus, si  $\underline{c}^1 = B \underline{a}^1$ , on obtient

$$B N_Q B \underline{a}^1 = \lambda B \underline{a}^1 \quad \Leftrightarrow \lambda \underline{a}^1 = B^{-1} B N_Q B \underline{a}^1$$

qui conduit au résultat souhaité.  $\square$

Les coordonnées des points  $N_i$  sur l'axe  $\Delta_{\underline{a}^1}$  réalisant la dissipation minimum d'énergie sont donc solutions de l'équation aux valeurs propres (72) qui généralise les équations de l'Analyse en Coordonnées Principales. La métrique euclidienne définie sur  $F$  n'est pas ici une matrice diagonale et n'est donc interprétable en termes de choix d'une origine. De plus les analogies à la mécanique, fondées sur la notion d'inertie d'un nuage de points massiques disparaît aussi. Il est par suite appréciable de disposer d'un nouveau cadre d'analogie avec la physique. D'autre part, il correspond aux points  $N_i$  de  $F^*$  les points  $P_i$  de  $E$  et on pose alors la définition

**Définition 1:** On appelle *Analyse en Coordonnées Principales du triple  $(I, \delta, Q_{I \times I})$* , l'analyse spectrale de l'opérateur  $B N_Q$ . Les coordonnées principales d'un point  $P_i$  de  $E$  sont alors définies dans la base de  $E$  formée des vecteurs propres  $B$ -orthonormés de  $B N_Q$ .

**Remarque :** Etant donné un tableau de contingence  $Q_{JK}$ , interprétons  $J$  et  $K$  comme les éléments d'une bi-partition de  $I$ . Définissons  $\delta$  comme la fonction de distance définie sur  $J \times K$  par :

$$\delta_{jk}^2 = \begin{cases} \left( \frac{1}{q_{j+}} + \frac{1}{q_{k+}} \right) (1 - \delta_j^k) & \text{si } (j,k) \in J \times J \text{ ou } (j,k) \in K \times K \\ \frac{1}{q_{j+}} + \frac{1}{q_{k+}} + 2 \frac{q_{jk}}{q_{j+} q_{k+}} & \text{si } (j,k) \in I \times K \end{cases}$$

où  $q_{j+}$  est le terme général de la matrice  $D_q$  induite par la mesure de comptage définie par la matrice

$$Q = \begin{pmatrix} 0 & Q_{JK} \\ {}^t Q_{JK} & 0 \end{pmatrix}$$

Par suite

$$N_Q = D_q - Q = \begin{pmatrix} D_J & -Q_{JK} \\ -{}^t Q_{JK} & D_K \end{pmatrix}$$

où

$$D_q = \begin{pmatrix} D_J & 0 \\ 0 & D_K \end{pmatrix}, \quad D_J = \text{diag}(q_{j+}), \quad D_K = \text{diag}(q_{k+})$$

de sorte que lorsque l'on choisit pour matrice  $B$  la matrice  $D_q^{-1}$ , l'équation (72) est celle de l'Analyse des Correspondances de  $Q_{JK}$ , exprimée en termes de  $\Sigma$ -stratégie. LE FOLL (1982) montre que cette propriété est conservée dans le cas de l'Analyse des Correspondances Multiples, qui est caractérisé par une partition de  $I$  en  $p$  classes. Remarquons de plus, en posant

$$N_{JK} = D_J - Q_{JK} \quad \text{et} \quad B_J = D_J^{-1}$$

que l'équation aux valeurs singulières

$$B_J N_{JK} \underline{c} = \lambda \underline{c}$$

devient

$$D_J^{-1} Q_{JK} \underline{c} = (1-\lambda) \underline{c}$$

et fournit les équations de transition de l'Analyse Factorielle des Correspondances où la définition de  $B_J$  montre que

$$\delta_{ij}^2 = b_{jj} + b_{kk} - 2 b_{jk} = \frac{1}{q_{j+}} + \frac{1}{q_{k+}}$$

Revenons au cas général et factorisons  $B$  en  $B = {}^t X M X$ , alors l'équation (72) est équivalente aux deux expressions suivantes :

$$\begin{aligned} \lambda B \underline{a}_1 &= B N_Q B \underline{a}_1 \\ \lambda {}^t X M X \underline{a}_1 &= {}^t X M X N_Q {}^t X M X \underline{a}_1 \end{aligned}$$

$X$  est la matrice d'une application linéaire définie sur  $F^*$  à valeurs dans un espace vectoriel  $E$ , par suite, si  $\underline{u}^1 = X \underline{a}_1$  est le vecteur de  $E$  donnant les coordonnées du point  $P_i$  dans  $E$  et

$$V_X = XN_Q{}^tX$$

alors

$$\begin{aligned}\lambda_{\underline{u}}^1 &= XN_Q{}^tXM\underline{u}^1 = V_XM\underline{u}^1 \\ {}^t\underline{a}_1B\underline{a}_1 &= {}^t\underline{u}^1M\underline{u}^1 = 1\end{aligned}\quad (73)$$

On vérifie de plus la conséquence des isométries ainsi construites :

$$\text{Trace}(V_XM) = \text{Trace}(BN_Q) = \text{Trace}(\Psi D_Q)$$

L'existence d'un solution dépend donc des propriétés de semi-définie positivité de l'opérateur  $BN_Q$

**Proposition 2:**  $BN_Q$  est une matrice semi-définie positive sur  $F$

**Preuve :**

$$\begin{aligned}D(c^1, Q) &= \frac{1}{2} \sum_i \sum_j q_{ij} (c_i^1 - c_j^1)^2 \\ &= \frac{1}{2} \sum_i \sum_j q_{ij} (c_i^1 - c_j^1) \left( \sum_{k=1}^n a_1^k (b_{ik} - b_{jk}) \right) \\ &= \frac{1}{2} \sum_{k=1}^n a_1^k \left( \sum_i \sum_j q_{ij} (c_i^1 - c_j^1) (b_{ik} - b_{jk}) \right) \\ &= \sum_{k=1}^n a_1^k (\lambda_1 c_k^1) = \lambda_1\end{aligned}\quad (74)$$

Si  $Q_{|x|}$  est positive, (74) implique alors que  $\lambda_1$  est positive.  $\parallel$

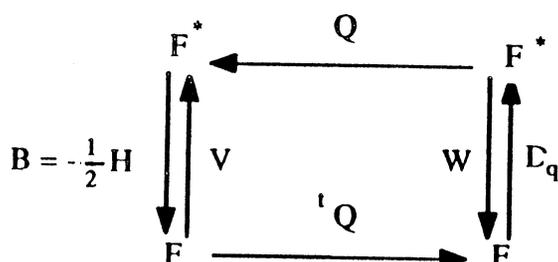
LE FOLL (1982) étudie l'équation (72) et l'interprète dans le cadre d'une Analyse en Composantes Principales particulière, en considérant alors le triplet  $({}^tQ_q, B, D_q)$  où

$$Q_q = D_q^{-1} (D_q - Q) = D_q^{-1} N_Q \quad \text{et} \quad B = -\frac{1}{2} H$$

Dans ce cas, on vérifie avec les notations habituelles que l'on a

$$\begin{aligned}W &= {}^tQ_q B Q_q \\ V &= {}^tQ_q D_q Q_q = D_q Q_q Q_q = D_q Q_q = (D_q - Q)\end{aligned}$$

et l'analyse est résumée par le schéma de dualité suivant



**Remarque :** Lorsque  $Q$  est un tableau de contingence symétrique induit par une mesure de comptage sur  $I \times I$ ,  $D_q$  contient sur la diagonale la répartition marginale induite par  $Q$  sur  $I$  et l'analyse linéaire de ce schéma permet de retrouver les équations de l'Analyse Factorielle des Correspondances.

LE FOLL (1972) factorise alors  $N_Q$  en  $N_Q = ZD_q{}^tZ$  de sorte que

$$\lambda \underline{c}^1 = BZD_q{}^tZ \underline{c}^1$$

et en posant

$$\underline{v}^1 = {}^tZ \underline{c}^1$$

$${}^t\underline{v}^1 D_q \underline{v}^1 = {}^t\underline{c}^1 N_Q \underline{c}^1 = 1$$

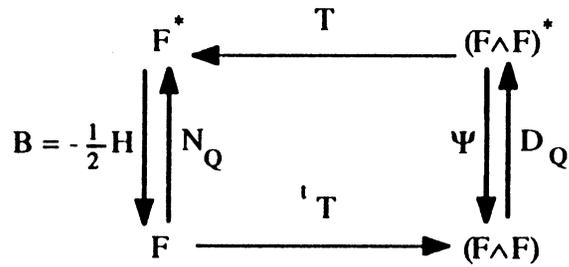
et on obtient une équation plus classique

$$\lambda \underline{v}^1 = W D_q \underline{v}^1$$

où

$$W = {}^tZ B Z = -\frac{1}{2} {}^tZ H Z$$

conduisant aux équations d'une Analyse en Coordonnées Principales formelle associée à la pondération  $P_I$  de matrice représentative  $D_q$ . La pertinence de cette interprétation est limitée au cas où la matrice  $B = -\frac{1}{2}H$  est semi-définie positive et est donc assurée d'après les résultats présentés au chapitres II et III si et seulement si  $\delta$  est une fonction de dissimilarité Euclidienne. Dans le cas contraire le choix  $H = \Delta^* \Delta$  n'a alors aucune optimalité particulière et toute autre transformation des données (par exemple  $h_{ij} = \delta_{ij}^a$ ,  $a > 0$ ) est également justifiée. Lorsque  $B$  n'est pas gramienne, on peut recourir à une perturbation de  $B$ . On retrouve ainsi posé le problème de la constante additive. Compte tenu des critiques faites plus haut sur une telle démarche, nous préférons établir les propriétés résultant de la dualité entre  $F$  et  $(F \wedge F)^*$  qui est résumée par le schéma



Les résultats établis dans ce cadre reposent alors sur la factorisation suivante de  $N_Q$

$$N_Q = {}^t D_Q T = D_Q - Q$$

et l'équation aux valeurs propres (72) admet comme contrepartie dans  $F \wedge F$  l'équation

$$\Psi D_Q \underline{d}^1 = \lambda \underline{d}^1 \quad (72\text{-bis})$$

où  $\Psi$  est la matrice des contrastes  $2 \times 2$

$$\Psi = {}^t B T = -\frac{1}{2} {}^t T H T$$

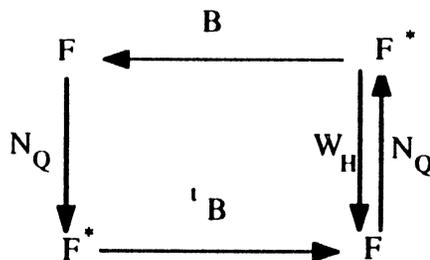
et le vecteur

$$\underline{d}^1 = {}^t T \underline{c}^1 = (c_i^1 - c_j^1)_{i < j}$$

s'interprète dans le cadre de notre problème de Physique comme le vecteur des différences de potentiel électrique aux nœuds du réseau et de plus vérifie les contraintes de normalisation

$${}^t \underline{d}^1 D_Q \underline{d}^1 = 1$$

L'équation (72-bis) montre de façon explicite que le problème traité dans ce cas est une Analyse en Coordonnées Principales menée dans  $F \wedge F$  à l'aide de la matrice des contrastes  $2 \times 2$ . Cette caractérisation est cependant d'un intérêt limité du fait de la taille des matrices soumises à l'analyse spectrale. Construisons cependant par analogie avec le cas de l'analyse du triple  $(I, \delta, P_I)$  le schéma de dualité qui décrit l'analyse en Composantes Principales du triplet  $(B, N_Q, N_Q)$



où

$$W_H = {}^t B N_Q B = \frac{1}{4} {}^t H N_Q H$$

L'analyse linéaire de ce triplet est alors clairement équivalente à l'Analyse en Coordonnées Principales du triple  $(I, \delta, Q_{I \times I})$  lorsque  $\delta$  est une dissimilarité Euclidienne. Dans le cas général, comme pour l'analyse de  $(I, \delta, P_I)$ , les deux méthodes divergent lorsqu'on fait intervenir une phase de réduction de dimension, puisque  $W_H N_Q$  est par construction une matrice gramienne opérant sur  $F$ , alors que  $BN_Q$  peut admettre des valeurs propres négatives. Notons de plus que si par construction l'analyse spectrale de  $BN_Q$  construit ici encore une approximation interne des dissimilarités, l'analyse du triple  $(B, N_Q, N_Q)$  permet de construire une approximation bilatérale des dissimilarités lorsque  $\delta$  n'est pas Euclidienne. En effet, l'analyse spectrale de  $W_H N_Q$  se traduit dans  $F \wedge F$  par l'analyse spectrale de l'opérateur  $\Psi D_Q \Psi D_Q$  et

$$\text{Trace}(W_H N_Q) = \text{Trace}(\Psi D_Q \Psi D_Q).$$

Les techniques de constante additive s'appliquent alors de façon inchangée.

La situation dans ce cas est totalement analogue à celle que nous avons décrite dans le cas du triple  $(I, \delta, P_I)$ . En particulier si  $\delta$  n'est pas Euclidienne, le recourt à la tranformation

$$H = \Delta * \Delta$$

ne présente pas de justification particulière et l'on doit comme plus haut envisager la méthode alternative menée en fonction de  $H = \Delta$ .

### VI.3.2. Le cas particulier $Q_{I \times I} = P_I \otimes P_I$

Nous étudions ici le cas particulier où tout couple  $(i, j)$  d'éléments de  $I$  vérifie

$$q_{ij} = p_i \cdot p_j$$

**Proposition 3:** Si

$$Q_{I \times I} = P_I \otimes P_I$$

alors  $D(\underline{c}^1, Q)$  est alors l'inertie du nuage  $\{(N_i, p_i) ; i \in [n]\}$  sur l'axe  $\Delta_{\underline{a}_1}$

$$D(\underline{c}^1, Q) = I_{\Delta_{\underline{a}_1}^\perp}$$

**Preuve :**

$$\begin{aligned} D(\underline{c}^1, Q) &= \frac{1}{2} \sum_i \sum_j p_i p_j (c_i^1 - c_j^1)^2 \\ &= \sum_i p_i (c_i^1)^2 - \sum_j p_j (c_j^1)^2 = I_{\Delta_{\underline{a}_1}^\perp} \quad \square \end{aligned}$$

Cela signifie que les deux interprétations du problème, électrique et mécanique, coïncident dans ce cas simple. Le taux de dissipation d'énergie entre les nœuds du réseau électrique s'exprime alors comme le moment d'inertie d'un nuage de points massiques autour de son barycentre. D'autre part

$$Q = p {}^t p = D_p \mathbb{1} {}^t \mathbb{1} D_p$$

Le projecteur  $P_p$  sur la droite des constantes de  $F$  est donc

$$P_p = \mathbb{1} {}^t \mathbb{1} D_p = D_p^{-1} Q$$

$Q$  ne dépend que de  $p$ . On note donc ici  $N_p$  la matrice associée à  $Q_{|x|}$  dans  $F$ .

$$N_p = D_q - Q = D_p - D_p \mathbb{1} {}^t \mathbb{1} D_p$$

c'est à dire que  $Q_p$  est l'opérateur  $D_p$ -symétrique

$$D_p Q_p = N_p = {}^t Q_p D_p$$

L'équation (72) se simplifie dans le cas d'une telle métrique

**Proposition 4:** *Les vecteurs propres centrés de  $BN_p$  sont les vecteurs propres de  $WD_p$  et définissent donc les coordonnées principales de  $(I, \delta, P_1)$ .*

**Preuve :**

$$\lambda \underline{c}^1 = BN_q \underline{c}^1 = H {}^t Q_p D_p \underline{c}^1$$

ce qui implique

$$\lambda Q_p \underline{c}^1 = Q_p B {}^t Q_p D_p \underline{c}^1 = (Q_p B {}^t Q_p) {}^t Q_p D_p \underline{c}^1$$

Soit en posant  $\underline{x}^1 = {}^t Q_p \underline{c}^1$  et  $W = Q_p B {}^t Q_p$

$$\lambda \underline{x}^1 = W D_p \underline{x}^1 \quad \parallel$$

#### VI.4. LES MÉTHODES DE CODAGE MULTIDIMENSIONNEL DU TRIPLE $(I, \delta, Q_{|x|})$

Nous considérons ici le problème du plongement Euclidien approché du triple  $(I, \delta, Q_{|x|})$  où  $(I, \delta)$  est un espace quasi-métrique de cardinal fini égal à  $n$ .  $Q_{|x|}$  est une pondération définie sur  $|x|$ , réelle symétrique et positive. Nous nous proposons alors d'ajuster un modèle additif du type

$$f(\delta_{ij}) = g(d_{ij}(X)) + \varepsilon_{ij} \quad (75)$$

au sens où l'on se propose de minimiser une fonction des résidus  $\epsilon_{ij}$  donnée à priori. Dans cette équation, l'étude menée dans les deux sections précédentes montre que la transformation  $f$  est en général soit l'identité, soit la fonction carré. Nous avons cependant étudié au chapitre précédent une classe plus générale de transformations. CRITCHLEY (1978) étudie en effet la classe de fonctions paramétrées

$$f(x) = x^\alpha \quad \alpha \in \mathbb{R}^+$$

et se propose de déterminer le coefficient  $\alpha$ , conduisant à l'optimisation d'un critère donné d'approximation. RAMSAY (1969, 1977, 1978, 1980, 1982) traite le cas où  $f$  et  $g$  sont des transformations logarithmiques, de façon à construire un tableau  $D(X)$  en termes d'estimation de  $\Delta$  au sens du maximum de vraisemblance dans le cadre d'une structure statistique log-normale dont la pertinence est justifiée par une analyse des processus de collecte des données de dissimilarité dans les applications de la méthode en psychométrie. Cette approche modélisante est poursuivie et complétée par TAKANE (1977a, 1978) qui s'intéresse en particulier aux données de confusion. Notre étude suppose que de telles transformations des données brutes ont été prises en compte dans la construction de  $\Delta$  et traite principalement les modèles simples caractérisés par deux fonctions  $f$  et  $g$  égales à l'identité ou à la fonction carré.

Nous présentons tout d'abord la méthode proposée par HAYASHI (1952) sous le nom de  $E_{ij}$ -quantification ou Méthode IV de Quantification. On a alors  $f = \text{Identité}$  et  $g(x) = x^2$ .

Nous traitons ensuite les approches issues des travaux de SHEPARD (1962a, b), KRUSKAL (1964a, b) et GUTTMAN (1968). Dans ce cas  $f = g = \text{Identité}$ . On parle alors de modèle d'ajustement de degré 1 et nous utilisons l'abréviation MINRES pour désigner les méthodes de minimisation des résidus au sens des moindres carrés pondérés. Puis nous traitons le cas des transformations carré. On parle alors de modèle d'ajustement de degré 2 et nous utilisons l'abréviation MINRES2 pour désigner les méthodes de minimisation des résidus au sens des moindres carrés pondérés. L'étude de ces modèles permet de montrer plus explicitement les relations entre les méthodes de Codage Multidimensionnel et d'Analyse en Coordonnées Principales. Cette comparaison de méthodes est inspirée par une réinterprétation des travaux de TAKANE (1977b) dans le formalisme algébrique présenté plus haut.

Les méthodes de Codage Multidimensionnel se singularisent alors soit par le critère d'adéquation qui est optimisé, soit par la normalisation de la solution, soit encore par l'algorithme itératif mis en œuvre pour la construction de la matrice de représentation  $X$  de rang donné  $r$ .

**Remarque :** L'étude des méthodes présentées ci-dessous respecte les conventions adoptées par leurs auteurs. Il convient cependant de marquer très clairement le caractère second du choix d'une transformation des données de dissimilarité dans la caractérisation mathématique des méthodes d'analyse. Toute transformation à priori de  $\Delta$  en une matrice  $\Delta^{*a}$  ( $a \in \mathbb{R}$ ,  $a > 0$ ) permet en effet d'adapter une méthode particulière aux besoins d'une analyse. Il n'y a donc pas de différence formelle entre les deux modèles

$$\delta_{ij}^a = d_{ij}(X) + \varepsilon_{ij} \quad (76)$$

$$\delta_{ij} = d_{ij}(X) + \varepsilon_{ij} \quad (77)$$

puisque'en posant  $\delta \Pi = \delta^{*a}$ , le modèle (76) est un cas particulier de (77). Cet argument ne doit pas cacher, bien sûr, que du point de vue de la pertinence concrète des analyses, les modèles (76) et (77) ne sont pas équivalents. Le choix des dissimilarités  $\delta$  et  $\delta^{*a}$  ne traduit pas les mêmes hypothèses de travail et ne conduit donc pas aux mêmes interprétations.

De plus FICHET (1985, 1987) et LE CALVE (1986, 1987, 1988) étudient en détail la structure d'inclusion des grandes classes de fonction de dissimilarité. Ces auteurs montrent en particulier qu'un grand nombre de fonctions de dissimilarité sont de type  $L_1$ . On sait alors que  $\sqrt{\delta}$  est une fonction de dissimilarité Euclidienne. Par suite, bien que ces résultats ne précisent pas l'influence d'une transformation des données par une fonction puissance sur la dimension du support du nuage représentatif, cette transformation semble particulièrement adaptée à la recherche d'un plongement Euclidien approché dans un espace affine de faible dimension. Enfin, l'exemple du modèle proposé par Ramsay montre que du point de vue numérique, l'utilisation de fonctions  $g$  plus complexes qu'une fonction puissance pose des problèmes d'applicabilité de certains algorithmes d'optimisation.

D'autre part, les modèles (76) et (77) ont un statut très différent lorsqu'on se propose comme CRITCHLEY (1978) d'estimer la valeur optimale du paramètre  $a$  (dans le modèle (76)) lors de l'application d'une méthode donnée à l'analyse du triple  $(I, \delta, Q_{|X|})$ .

De plus, les spécificités des méthodes notées dans l'exposé proviennent d'un choix de normalisation de la solution assurant la cohérence interne des méthodes.

#### VI.4.1. La méthode IV de quantification

HAYASHI (1952) suppose donnée une fonction de similarité propre sur  $I \times I$  et traite le cas d'une pondération  $Q_{|X|}$  triviale définie par

$$(D_Q = \frac{1}{n^*} I_{n^*}) \Leftrightarrow (Q = \frac{1}{n^*} \mathbf{1}\mathbf{1})$$

Notre présentation décrit cette méthode en termes de dissimilarité et suppose donné le triplet  $(I, \delta, Q_{|X|})$  respectant les contraintes de spécification données ci-dessus. Nous montrons plus bas que l'introduction d'une pondération non triviale conduit à la perte des certaines propriétés d'invariance de la méthode. Hayashi propose de construire la solution  $X^*$  de rang  $r$  qui réalise dans  $\mathcal{M}_{r,n}(\mathbb{R})$  le maximum de la fonction critère

$$\theta_1(X) = \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}^2(X) \quad (78)$$

$\theta_1(X)$  n'est pas le coefficient de corrélation entre  $\Delta$  et  $D^{*2}(X)$ , contrairement aux objectifs fixés par Hayashi. Par suite, l'optimisation de  $\theta_1(X)$  ne conduit pas forcément à accroître la corrélation entre  $\Delta$  et  $D^{*2}(X)$ . Il nous semble donc préférable d'adopter l'interprétation de la méthode en termes d'énergie dissipée dans un réseau électrique, de résistance

$$r_{ij} = q_{ij} \delta_{ij}$$

C'est donc une extension directe de la méthode présentée au paragraphe §IV.2 au cas multidimensionnel.  $\theta_1(X)$  s'écrit de façon équivalente

$$\theta_1(X) = \sum_{i < j} \sum r_{ij} d_{ij}^2(X) = \text{Trace}(\Psi_X D_R) = \text{Trace}(W_X N_R)$$

où  $R$  est la matrice de terme général  $r_{ij}$  et

$$H_X = D(X) * D(X) \quad W_X = \alpha_1(H_X) \quad N_R = T D_R {}^t T = D_r - R$$

Les équations précédentes peuvent être résumées sous la forme suivante

**Proposition 5:** *La méthode de quantification IV admet pour solution d'ordre  $r$ , la matrice  $X$  dont les colonnes sont les  $r$  vecteurs propres associés au  $r$  plus grandes valeurs propres obtenues par l'analyse spectrale de  $N_R$ .*

**Remarque :** Nous avons supposé dans cet énoncé que la valeur propre  $\lambda_r$  est simple. Lorsque  $\lambda_r$  est une valeur propre multiple, on doit donner un résultat exprimé sous une forme plus prudente à propos de la dimension de l'espace de plongement. Nous retenons cependant cette convention dans la suite afin de limiter la part des considérations techniques dans notre exposé.

L'emploi de la méthode de quantification IV est souvent justifié par ses propriétés d'invariance par transformation affine des données. Posons en effet

$$\delta\Pi = \alpha + \beta\delta \quad \alpha, \beta \in \mathbb{R} \quad \beta > 0 \quad (79)$$

Nous nuancions ici le résultat démontré par OKAMOTO et TODA (1973), en remarquant que l'invariance des résultats n'est pas conservée lorsque  $Q_{|x|}$  est une pondération non triviale. De plus, OKAMOTO et ISOGAI (1978) montrent que la propriété d'invariance annoncée n'est valable que si les vecteurs propres de  $N_R$  sont tels que

$$\underline{u}_s \underline{u}_s = 1 \quad s = 1, 2, \dots, r$$

**Proposition 6:** Soit  $X^*$  la solution du problème de quantification IV construite à l'aide des vecteurs propres de  $N_R$  normés à 1. Alors si  $Q_{|x|}$  est la pondération triviale  $Q_{|x|} = \underline{1}^t$ ,  $X^*$  est invariante par transformation affine de type (79) des dissimilarités.

**Preuve :** Posons pour tout couple (i,j) d'éléments de I

$$r_{ij}' = q_{ij} (\alpha + \beta \delta_{ij}) = \alpha q_{ij} + \beta r_{ij}$$

$$R' = \alpha Q + \beta R$$

$$N_{R'} = \alpha N_Q + \beta N_R$$

Soit  $\underline{u}$  un vecteur propre de  $N_R$ . Alors

$$N_{R'} \underline{u} = \alpha N_Q \underline{u} + \beta N_R \underline{u} = \alpha N_Q \underline{u} + \beta \lambda \underline{u}$$

On obtient

$$N_{R'} \underline{u} = \lambda \Pi \underline{u}$$

seulement dans le cas où  $\underline{u}$  est aussi vecteur propre de  $N_Q$  et alors

$$(N_Q \underline{u} = \mu \underline{u}) \Rightarrow (N_{R'} \underline{u} = (\alpha\mu + \beta\lambda) \underline{u})$$

Dans le cas particulier où  $Q_{|x|} = \underline{1}^t$ , nous avons vu que  $N_Q = nQ_1$ .  $Q_1$  admet (n-1) valeurs propres égales à 1 et une valeur propre nulle correspondant au vecteur constant. Par suite toute famille libre de (n-1) vecteurs orthogonaux au vecteur constant  $\underline{1}_F$  décrit  $F_1$ . Les vecteurs propres  $\underline{u}_s$  de  $N_R$  peuvent donc être retenus comme éléments de cette base et alors

$$N_Q \underline{u} = n \underline{u}$$

et l'on retrouve le résultat de OKAMOTO et TODA (1973). ¶

OKAMOTO et ISOGAI (1978) étudient les propriétés de la méthode lorsque les vecteurs propres sont normés à la valeur propre correspondante. Bien sûr, l'invariance par transformation affine de  $\delta$  est perdue. Ces auteurs montrent de plus (dans la cas  $q_{ij} = 1$ ) que les deux contraintes suivantes

$$\text{Trace} ({}^t X N_Q X) = \frac{1}{r} \sum_{s=1}^r \| \underline{u}_s \|_{N_Q}^2 = 1 \quad (80)$$

$$\text{Trace} (\Psi_X D_Q) = \sum_{i < j} r_{ij} d_{ij}^2(X) = 1 \quad (81)$$

sont équivalentes, ce qui est une conséquence directe du lemme de Coxeter. D'autre part, SAITO (1982) utilise la propriété d'invariance donnée ci-dessus dans le cas d'une pondération triviale pour comparer les solutions fournies par l'Analyse en Coordonnées Principales (non invariante par une telle transformation des données) et par la méthode de quantification IV. En effet, lorsqu'elle est satisfaite, la propriété d'invariance de la solution se traduit par une transformation des valeurs propres

$$\lambda \Pi_s = n\alpha + \beta \lambda_s$$

Le paramètre  $\alpha$  joue le rôle d'une constante additive et le choix des coefficients  $\alpha$  et  $\beta$  conduit à des stratégies distinctes de la détermination de la dimension  $r$  à retenir dans une phase de réduction de la dimension conduisant à une représentation résumée de  $I$ . L'évaluation de critères du type

$$\frac{\sum_{s=1}^r \lambda_s^a}{\sum_{s=1}^n \lambda_s^a} \quad a \in \mathbb{R}^+$$

perd en effet toute valeur intrinsèque ici. Saito montre d'ailleurs sur un exemple numérique, que  $\alpha$  et  $\beta$  influent beaucoup sur la séparation des valeurs propres, c'est à dire sur l'information usuellement jugée déterminante en analyse linéaire des données pour fixer la dimension d'une représentation résumée des données. Il montre aussi la complexité du problème posé par la détermination de la dimension d'un représentation de  $I$  à l'aide de critères utilisés par analogie avec l'Analyse en Composantes Principales.

Compte tenu de ces caractéristiques négatives de la méthode, Saito propose de déterminer la dimension du résumé à retenir, à l'aide d'une évaluation d'un indice de corrélation ordinale (le TAU de Kendall ou le RHIO de Spearman) entre  $\underline{\delta}$  et  $\underline{d}(X)$ . Cette méthode n'est pas optimale puisque ces indices n'ont pas de raison particulière de croître en fonction de la dimension retenue puisque

$$(d_{ij}^2(X_r) \leq d_{kl}^2(X_r)) \quad \Rightarrow \quad (d_{ij}^2(X_{r+1}) \leq d_{kl}^2(X_{r+1}))$$

D'autre part, le problème de normalisation de la solution du problème traité par Hayashi, conduit à s'interroger sur le rôle joué par les relations (80) et (81). Remarquons que

$$\sigma(\underline{\delta}, \underline{d}^2(X)) = \sum_{i < j} \sum q_{ij} (\delta_{ij} - d_{ij}^2(X))^2 = \sum_{i < j} \sum q_{ij} \delta_{ij}^2 + \sum_{i < j} \sum q_{ij} d_{ij}^4(X) - 2 \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}^2(X)$$

Par suite, maximiser  $\theta_1(X)$  est équivalent à minimiser le critère d'ajustement aux moindres carrés pondérés  $\sigma(\underline{\delta}, \underline{d}^2(X))$  dans le seul cas où on impose une contrainte du type

$$\sum_{i < j} \sum q_{ij} d_{ij}^4(X) = k, \quad k \text{ donné} \quad (82)$$

Mais si l'on pose

$$\text{HSTRESS}(X) = \frac{\sum_{i < j} \sum q_{ij} (\delta_{ij} - d_{ij}^2(X))^2}{\sum_{i < j} \sum q_{ij} d_{ij}^4(X)}$$

Alors on sait qu'il y a équivalence entre les trois problèmes suivants :

- (P1) Min { HSTRESS(X) ;  $X \in \mathcal{M}_{r,n}(\mathbb{R})$  }
- (P2) Min {  $\sigma(\underline{\delta}, \underline{d}^2(X))$  ;  $X \in \mathcal{M}_{r,n}(\mathbb{R})$  et X vérifie (82) }
- (P3) Max {  $\theta_1(X)$  ;  $X \in \mathcal{M}_{r,n}(\mathbb{R})$  et X vérifie (82) }

Le problème P3 est différent du problème traité par Hayashi du fait de l'adjonction de la contrainte de normalisation (82). De plus, en analyse du triple il est courant de normaliser de façon globale la solution X de sorte que

$$\sum_{i < j} \sum q_{ij} \delta_{ij}^2 = \sum_{i < j} \sum q_{ij} d_{ij}^2(X)$$

Par suite (82) s'interprète comme une contrainte portant sur la variance empirique de la variable  $d^2(X)$  définie sur  $(I \times I, Q_{I \times I})$ . Elle conduit à neutraliser les effets de variance interpoints lors de la comparaison des solutions admissibles X et doit donc être rapprochée du constat fait par SHEPARD (1962), qui préconise d'accroître la variance des distances interpoints afin de diminuer la dimension de l'espace affine dans lequel est réalisé le plongement Euclidien.

#### VI.4.2. La méthode MINRES

Les programmes de Codage Multidimensionnel les plus couramment utilisés se proposent de minimiser les résidus du modèle

$$\delta_{ij} = d_{ij}(X) + \varepsilon_{ij} \quad (83)$$

au sens d'un critère de moindres carrés pondérés. L'origine de cette formulation du problème remonte aux travaux de KRUSKAL (1964a,b) qui donne une formulation mathématique du problème en termes d'optimisation et un algorithme résolutif convainquant, destinés à rationaliser l'approche heuristique proposée par SHEPARD (1962a, b).

Cet auteur retient comme mesure de la qualité de l'approximation, la moyenne quadratique des résidus  $\epsilon_{ij}$ . Nous parlons alors de méthode MINRES pour qualifier toute méthode de minimisation des résidus en moyenne quadratique. En effet, la fonction STRESS minimisée par KRUSKAL mesure l'écart entre dissimilarités observées  $\underline{\delta}$  et distances reconstruites  $\underline{d}(X)$  au sens des moindres carrés. Cet indice est un cas particulier de la classe d'indices proposée par OBENCHAIN (1973) correspondant à  $\psi(1, 2, \Delta)$ :

$$\text{STRESS}(X) = \sum_{i < j} \sum q_{ij} (\delta_{ij} - d_{ij}(X))^2 \quad (84)$$

$$= \sum_{i < j} \sum q_{ij} \delta_{ij}^2 + \sum_{i < j} \sum q_{ij} d_{ij}^2(X) - 2 \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(X) \quad (85)$$

Originellement, KRUSKAL (1964) traite le cas d'une pondération triviale  $q_{ij} = 1$  pour tout couple  $(i,j)$  d'éléments de  $I$  et GUTTMAN (1968) utilise la pondération  $q_{ij} = 0$  si la donnée  $\delta_{ij}$  est manquante et 1 sinon.

Les programmes développés par Guttman, Lingoes et Roskam (cf. LINGOES (1973)) optimisent de façon implicite cette fonction STRESS particulière, bien qu'ils mesurent la qualité de l'ajustement à l'aide d'un critère utilisé auparavant par Hahashi. De plus KRUSKAL et CARROLL (1969) présentent une étude détaillée des fonctions critères utilisées pour le développement des méthodes de Codage Multidimensionnel.

Ecrivons maintenant la fonction STRESS dans  $F \wedge F$

$$\text{STRESS}(X) = \| \text{Veca}(\Delta) - \text{Veca}(D(X)) \|_{D_Q}^2 \quad (84 \text{ bis})$$

$$= \| \underline{\delta} - \underline{d}(X) \|_{D_Q}^2 \quad (84 \text{ ter})$$

$$= \| \underline{\delta} \|_{D_Q}^2 + \| \underline{d}(X) \|_{D_Q}^2 - 2 \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q} \quad (85 \text{ bis})$$

**Remarque :** L'équation (84-ter) permet de comparer le problème posé par Kruskal à celui qui est résolu en Analyse en Coordonnées Principales pour la variante introduite au paragraphe §V.2.1 pour  $H = \Delta$ . Le critère optimisé s'écrit alors avec des notations évidentes

$$\sigma(X) = \text{Trace}\{(W_\delta - W_X)D_p(W_\delta - W_X)D_p\}$$

On observe alors clairement que les deux méthodes diffèrent par la géométrie Euclidienne adoptée dans  $F \otimes F$  puisque la méthode proposée s'exprime en termes de dissimilarité sous la forme

$$\text{Min} \{ \sigma(X) \ ; \ X \in \mathcal{M}_{r,n}(\mathbb{R}) \}$$

où d'après l'équation de  $\sigma(X)$ ,

$$\sigma(X) = \frac{1}{4} \text{Trace}((\Delta - D(X))N_p(\Delta - D(X))N_p) = \frac{1}{4} \| \underline{\delta} - \underline{d}(X) \|_{N_p \otimes N_p}^2$$

Supposons maintenant que  $\underline{d}(X)$  est différentiable en tout point  $X$ . Le gradient de la norme Euclidienne  $\|x\|$  est

$$\nabla \| \cdot \| \Big|_x = \frac{x}{\|x\|} \quad \text{si } x \neq 0$$

et donc

$$\nabla \underline{d}(X) \Big|_{\underline{x}_i - \underline{x}_j} = \frac{\underline{x}_i - \underline{x}_j}{d_{ij}(X)} \quad \text{si } i \neq j$$

Notons alors  $\gamma$  le gradient de  $\underline{d}(X)$

$$\gamma(\underline{u}) = \nabla \underline{d}(X) \Big|_{\underline{u}}$$

Alors GLAZMAN et LIUBITCH (1971) rappellent que

i) comme tout gradient d'une norme,  $\gamma$  vérifie

$$(a) \gamma(\lambda \underline{u}) = \lambda \gamma(\underline{u}) \quad \text{pour tout } \lambda > 0$$

$$(b) \gamma(\underline{u}_1 + \underline{u}_2) \leq \gamma(\underline{u}_1) + \gamma(\underline{u}_2)$$

$$(c) \text{Sup} \{ |\gamma(\underline{u})| \ ; \ \|\underline{u}\| = 1 \} = 1$$

ii)  $\| \cdot \|$  est différentiable au point  $x \neq 0$  si et seulement si son gradient est antisymétrique en ce point, c'est à dire si  $\gamma(-\underline{u}) = -\gamma(\underline{u})$

C'est dire que  $\underline{d}(X)$  est différentiable en tout point  $\underline{u}$  tel que  $\gamma(\underline{u})$  appartient à  $F \wedge F$ , justifiant ainsi le choix de  $F \wedge F$  pour formaliser les méthodes de Codage Multidimensionnel.

De plus, l'expression de  $\gamma(\underline{u})$  montre que les équations normales déduites de (84) par différentiation dépendent des paramètres de façon non linéaire et leur résolution impose de recourir à une démarche itérative. Par suite, KRUSKAL (1964a,b) construit une configuration caractérisée par la matrice  $X$  réelle d'ordre  $r \times n$ , à l'aide d'une procédure de minimisation de l'indice de STRESS, sous des contraintes techniques de normalisation présentées plus bas. Il traite le problème numérique sous-jacent à l'aide d'une méthode de

gradient particulièrement élaborée quant au choix du pas d'itération (cf. KRUSKAL (1977)). Il étend de plus cette méthode au traitement de données de dissimilarité transformées, c'est à dire de disparités

$$t_{ij} = f(\delta_{ij}) \quad i, j \in I$$

où  $f$  est un élément d'une classe de fonctions  $\mathcal{F}$  déterminée par les contraintes de mesurage liées soit au processus de collecte des données (on suppose par exemple que  $f$  est une fonction monotone par rapport à l'ordre induit par les indices de dissimilarité observés (cf. KRUSKAL (1964a,b), d'AUBIGNY (1975))), soit au souci d'assurer la stabilité des résultats en retenant des fonctions  $f$  dépendant d'un petit nombre de paramètres (on suppose par exemple que  $f$  est une fonction polynomiale de degré fixé (cf. KRUSKAL (1964))), soit encore au cumul de ces deux objectifs (on suppose par exemple que  $f$  est une fonction spline cubique monotone (cf. WINSBERG et RAMSAY (1983), WINSBERG et CARROLL (1985)) ou un lisseur à noyau (cf. EL HADRI (1988)). Les disparités résultent alors d'un ajustement, le plus souvent optimal au sens des moindres carrés, de la fonction de dissimilarité  $\delta$  en fonction d'une fonction de distance  $d(X)$  calculée pour une configuration  $X$  donnée. On résoud alors le problème d'optimisation

$$\begin{aligned} & \text{Min}\{\text{STRESS}(X) ; \mathbf{t} \in \mathcal{F}\} \\ & \text{STRESS}(X) = \|\mathbf{t} - \underline{d}(X)\|_{D_Q}^2 \end{aligned}$$

Plus précisément, Kruskal désigne l'indice (84) sous le nom de RAW STRESS (ou STRESS non normalisé) et propose deux normalisations distinctes de cet indice

$$\begin{aligned} \text{STRESSFORM1}(X) &= \frac{\text{STRESS}(X)}{\sum_{i < j} \sum q_{ij} d_{ij}^2(X)} \\ \text{STRESSFORM2}(X) &= \frac{\text{STRESS}(X)}{\sum_{i < j} \sum q_{ij} (d_{ij}(X) - \bar{d}(X))^2} \end{aligned}$$

où

$$\bar{d}(X) = \frac{\sum_{i < j} \sum q_{ij} d_{ij}(X)}{\sum_{i < j} \sum q_{ij}}$$

KRUSKAL et CARROLL (1969) étudient les relations existant entre de nombreux indices mesurant la liaison entre  $\underline{\delta}$  et  $\underline{d}(X)$ . En particulier, ces auteurs comparent le STRESS à l'indice proposé par GUTTMAN (1968) sous le nom de coefficient d'Aliénation. d'AUBIGNY (1977, 1979) étudie les propriétés asymptotiques du STRESSFORM2, regardé comme estimateur d'un indice d'association monotone théorique. de LEEUW (1973a,b) et de LEEUW et STOOP (1984) produisent des bornes de variation des indices STRESSFORM1 et STRESSFORM2. Ces études apportent une information plus

rigoureuse sur le comportement des algorithmes de Codage Multidimensionnel que les nombreuses simulations menées entre autres par CLARK (1976), KLAHR (1969), LEVINE (1978), SPENCE (1973), SPENCE et OGILVIE (1973), STENSON et KNOLL (1969) et WAGENAAR et PADMOS (1971) dans le but de guider l'utilisateur dans le choix d'une dimension de l'espace de représentation E. Les tables qui en résultent sont supposées aider l'utilisateur à juger la qualité globale d'une représentation du point de vue de l'approximation des dissimilarités initiales par les distances reconstruites  $D(X)$ . On peut en particulier consulter ARABIE (1973, 1978), SPENCE (1974), SPENCE et YOUNG (1978) pour mesurer les difficultés rencontrées par une pratique fondée sur les résultats de telles simulations. d'AUBIGNY (1987) discute certaines méthodes plus élaborées d'évaluation de la qualité de la solution.

De nombreuses propositions d'algorithmes ont suivi les travaux de Kruskal et de Shepard. En particulier, EL HADRI (1988), KRUSKAL (1977), KRUSKAL et al. (1973), de LEEUW (1970a,b), LINGOES (1973), LINGOES et RAMSAY (1973), ROSKAM (1968, 1969b), YOUNG et TORGERSON (1967), YOUNG et al. (1980) analysent en détail les aspects algorithmiques des méthodes de Codage Multidimensionnel. Nous présentons ici l'approche de de Leeuw et Heiser (cf. de LEEUW (1977) et de LEEUW et HEISER (1980)) du fait de ses propriétés de convergence connues et de ses relations avec les travaux de Guttman publiés tardivement dans GUTTMAN (1968). de Leeuw et Heiser montrent que l'utilisation de méthodes de gradient, préconisée par de très nombreux auteurs dont KRUSKAL (1964a,b) n'est pas satisfaisante pour minimiser la fonction STRESS en fonction de X puisque cette fonction n'est pas différentiable aux points X pour lesquels il existe au moins un couple de points  $(i,j) \in I \times I$ ,  $i \neq j$  tel que  $d_{ij}^2(X) = 0$ . Les auteurs généralisent en conséquence la démarche usuelle au cadre de l'optimisation convexe non différentiable et proposent un algorithme appartenant à la classe des méthodes de sous-gradient. EL HADRI (1988) propose un algorithme de presque gradient qui semble mieux respecter les informations relatives aux petites dissimilarités. Il est néanmoins limité à des valeurs de n plus petites que l'algorithme de de Leeuw et Heiser et est trop récent pour disposer d'expériences numériques suffisantes. La justification de l'algorithme de de Leeuw et Heiser repose sur le fait que toute fonction de distance de la forme

$$d(X) : (i,j) \in I \times I \rightarrow d_{ij}(X) = \mu(\underline{x}_i - \underline{x}_j)$$

où  $(\underline{x}_i - \underline{x}_j)$  est le terme général de la matrice TX, est convexe et positivement homogène lorsque  $\mu(\cdot)$  est une jauge. C'est en particulier vrai pour les distances utilisées en Codage Multidimensionnel et donc pour la distance Euclidienne utilisée ici. Par suite, si l'on écrit le produit scalaire intervenant dans (85-bis)

$$\rho(X) = \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(X) = \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q}$$

sous la forme équivalente:

$$\rho(X) = \sum_{i < j} \sum q_{ij}(X) d_{ij}^2(X) = \|\underline{d}(X)\|_{D_{Q(X)}}^2$$

on obtient une fonction convexe positivement homogène, associée à la matrice diagonale  $D_{Q(X)}$ , définie à partir de la matrice symétrique et à termes positifs  $Q(X)$  de terme général

$$q_{ij}(X) = \begin{cases} q_{ij} \delta_{ij} \frac{1}{d_{ij}(X)} & \text{si } d_{ij}(X) \neq 0 \\ 0 & \text{sinon} \end{cases} \quad (86)$$

c'est à dire par l'expression

$$D_{Q(X)} = D_Q D_\Delta D_{D(X)} \quad (87)$$

$Q(X)$  est une matrice symétrique lorsque  $\Delta$  et  $Q$  le sont. De plus, à pondération  $q_{ij}$  constante, le terme  $q_{ij}(X)$  pondère la paire  $(i,j)$  d'autant plus fortement (resp. faiblement) que  $d_{ij}(X)$  sous-estime (resp. surestime) la dissimilarité  $\delta_{ij}$ . Enfin, si on interprète  $r_{ij} = q_{ij} \delta_{ij}$  comme une résistance du lien direct liant les nœuds  $i$  et  $j$  d'un réseau électrique, nous observons que la matrice  $D_{Q(X)}$  est une inverse généralisée de la matrice des résistances "locales" associées à la configuration  $X$  et définit une loi d'Ohm généralisée par rapport à l'équation (70) donnée au paragraphe §VI.2.

Par application du lemme de Coxeter,  $Q(X)$  induit une forme quadratique semi-définie positive sur  $F$  de matrice  $N_{Q(X)}$  notée dans la suite  $N(X)$

$$N(X) = T D_{Q(X)} {}^t T$$

On munit ainsi  $F_1$  d'une géométrie Euclidienne locale au sens où elle est indiquée par  $X$ , pour laquelle on peut écrire

$$\begin{aligned} \rho(X) &= \text{Trace}(X N(X) {}^t X M) \\ &= \|\underline{\delta}\|_{N(X) \otimes M}^2 \end{aligned} \quad (88)$$

#### Remarques:

(a) L'expression (86) met en lumière le rôle joué par les termes  $d_{ij}(X)$  nuls dans les problèmes de non différentiabilité de la fonction STRESS. La formule (87) utilisée par la méthode de de Leeuw et Heiser montre que leur méthode de sous-gradient se traduit par le choix d'une inverse généralisée particulière  $D_{D(X)}$  de  $D_{D(X)}$ . Ce choix est unique lorsque le sous-différentiel  $\partial\rho(X)$  est réduit à un point qui est alors le gradient  $\nabla$

$\rho(X)$ , c'est à dire lorsqu'aucun terme  $d_{ij}(X)$  n'est nul. On retrouve la formulation de KRUSKAL (1964) avec

$$D_{D(X)} = D_{D(X)}^{-1}$$

Les méthodes de presque gradient, utilisées par exemple par EL HADRI (1988), diffèrent principalement de la méthode de de Leeuw et Heiser par le choix d'une matrice de résistance "locale" fait dans cette phase de description de l'algorithme.

(b) Lorsque l'une des deux matrices  $\Delta$  et  $Q$  n'est pas symétrique, on doit projeter la matrice  $D_Q D_\Delta$  sur le cône des matrices symétriques opérant sur  $F \wedge F$  afin d'assurer la symétrie de  $N(X)$  et ainsi la pertinence du modèle Euclidien. La matrice résultant de cette projection est aisée à calculer puisqu'elle a pour terme général

$$\frac{1}{2} (q_{ij} \delta_{ij} + q_{ji} \delta_{ji})$$

(c) L'application du lemme de Coxeter est fondée sur le caractère diagonal de la métrique  $D_{Q(X)}$  sur  $F \wedge F$ . L'algorithme de de Leeuw et Heiser ne s'applique pas dans le cas du choix de métriques pourtant particulières  $\Phi = N_p \otimes N_p$  qui caractérise l'Analyse en Coordonnées Principales ou dans le cas d'une pondération générale  $\Phi = N_Q \otimes N_Q$ . Rappelons que nous avons montré au §IV.4 que seul existe un isomorphisme de structure d'espace vectoriel entre les formes quadratiques diagonales sur  $F \wedge F$  et les formes bilinéaires symétriques sur  $F$  et donc le statut de la matrice semi-définie positive  $N = T\Phi^t T$  n'est pas clair lorsque  $\Phi$  n'est pas diagonale. Cependant l'analyse spectrale de  $NW_\delta$  est équivalente à celle de  $\Phi\Psi$ , au sens où

$$\text{Trace}(\Phi\Psi) = \text{Trace}(NW_\delta)$$

et permet d'analyser la matrice  $\Psi$  des contrastes  $2 \times 2$  induits par  $\delta$  au prix d'un effort numérique sensiblement accru. Le rôle de la métrique  $\Phi$  reste cependant à déterminer. TAKANE (1977) montre qu'en fait l'Analyse en Coordonnées Principales consiste à retenir la matrice  $\Phi$  et non la matrice diagonale  $D_\Phi = \Phi * I$  et diffère donc du choix dans  $F_1$  de la métrique

$$N(X) = N = T D_\Phi^t T$$

La méthode consistant à utiliser la métrique  $D_{Q(X)} = D_\Phi D_\Delta D_{D(X)}$  dans la procédure de de Leeuw et Heiser ne semble pas avoir fait l'objet de travaux théoriques ou empiriques.

La fonction  $\rho(\cdot)$  définie en (88) est une fonction convexe et positivement homogène de  $X$  puisque  $d(\cdot)$  l'est. On peut donc utiliser un algorithme de programmation convexe pour

maximiser cette fonction de  $X$  et ainsi minimiser la fonction STRESS, en remarquant que le premier terme de droite de l'équation (85-bis) est fixé par les données.  $\rho(X)$  est une norme finie d'après (82) et par suite (cf. ROCKAFELLAR (1970)) son sous-différentiel  $\partial\rho(X)$  est un sous ensemble convexe non vide et compact du dual algébrique  $F_1^* \otimes E^*$  de  $F_1 \otimes E$  (ou à un isomorphisme près, de l'ensemble des applications linéaires définies sur  $F_1^*$  et à valeurs dans  $E$ ). Sous cette forme, nous observons que  $E$  est une inconnue du problème, on ne peut donc pas démontrer de théorèmes d'emboîtement des solutions de dimension distincte. Nous continuons cependant de présenter la méthode en introduisant une métrique  $M$  sur  $E$  dans le souci d'explicitier le cadre formel de description de la méthode de de Leeuw et Heiser.

Rappelons que  $\mathcal{M}_{r,n}(\mathbb{R})$  est muni d'une structure Euclidienne lorsqu'on définit le produit scalaire

$$\langle A, B \rangle = \text{Trace}(M^t A N B) = \langle \underline{a}, \underline{b} \rangle_{N \otimes M}$$

associé à tout choix de métriques  $M$  et  $N$  définies sur  $E$  et  $F_1$  respectivement. Nous supposons de plus que  $N$  est la métrique  $N(X)$ .

$Y$  est par définition un élément de  $\partial\rho(X)$  si et seulement si l'on a pour toute configuration  $\underline{z}$  dans  $F_1 \otimes E$ :

$$\rho(Z) \geq \rho(X) + \langle Y, Z - X \rangle$$

Par suite de LEEUW (1977) démontre le résultat suivant étendu ici au cas où  $E$  est muni d'une métrique  $M$ , qui permet de vérifier que  $Y = N(X)^t X M$  est un sous-gradient de  $\rho$  au point  $X$ , c'est à dire un élément de  $\partial\rho(X)$ . En effet  $Y$  est la matrice d'une application linéaire de  $E$  dans  $F_1^*$ , dans les bases  $\mathcal{E}$  et  $\mathcal{F}^*$ .

**Lemme 3:** Pour tous  $X, Z \in F_1 \otimes E$ , on a

$$\rho(X) \geq \text{Trace}(X N(X)^t Z M)$$

ou de façon équivalente

$$\rho(Z) \geq \langle \underline{x}, \underline{z} \rangle_{N(X) \otimes M}$$

**Preuve:** l'inégalité de Cauchy-Schwarz appliquée dans  $E = (\mathbb{R}^p, M)$  assure que

$$d_{ij}(Z) d_{ij}(X) \leq \sum_{s=1}^p \sum_{t=1}^p m_{st} (x_i^s - x_j^s)(z_i^s - z_j^s)$$

et donc en notant  $d_{ij}(X)$  le terme générique de  $D_{D(X)}$ :

$$\sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(X) \leq \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(Z) \left[ \sum_{s=1}^p \sum_{t=1}^p m_{st} (x_i^s - x_j^s)(z_i^s - z_j^s) \right]$$

qui conduit au résultat après arrangement des termes de droite. ¶

**Remarque :** La complexité des calculs mis en œuvre lors de la minimisation de la fonction STRESS provient du fait que la métrique  $N(X)$  dépend du point  $X$ . En effet, pour une matrice symétrique semi-définie positive  $N$  fixée, le gradient de la semi-norme

$$\eta^0(Z) = \text{Trace}(Z N {}^t Z M)$$

définie sur un espace vectoriel  $F_1 \otimes E$  est classiquement :

$$\nabla \eta^0(Z) = N {}^t Z M \quad (89)$$

Dans  $F_1 \otimes F_1$  nous pouvons réécrire la relation (84) sous la forme (84-ter) qui explicite la nature statistique du problème traité par de Leeuw et Heiser :

$$\text{STRESS}(X) = \|\underline{\delta} - \underline{d}(X)\|_{D_Q}^2$$

où  $\underline{\delta}$  est le tenseur symétrique de terme général  $\delta_{ij}$ . De plus, du fait de l'homogénéité de la fonction distance, minimiser la fonction STRESS par rapport à  $X$  équivaut à minimiser la quantité

$$S(X, \beta) = \|\underline{\delta} - \beta \underline{d}(X)\|_{D_Q}^2$$

en fonction de  $(X, \beta)$ , avec  $\beta \in \mathbb{R}$ . Pour  $X$  fixé,  $\underline{d}(X)$  est connue et l'on doit donc résoudre un problème classique de régression aux moindres carrés sans terme constant. Le minimum en  $\beta$  est atteint pour

$$b(X) = \frac{\langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q}}{\|\underline{d}(X)\|_{D_Q}^2} = \frac{\rho(X)}{\eta^2(X)} \quad (90)$$

Le minimum de  $S(X, \beta)$  s'écrit de façon classique en fonction du coefficient  $r(X)$  de corrélation linéaire entre  $\underline{\delta}$  et  $\underline{d}(X)$  défini par

$$r(X) = \frac{\langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q}}{\|\underline{\delta}\|_{D_Q} \|\underline{d}(X)\|_{D_Q}} = \frac{\rho(X)}{\|\underline{\delta}\|_{D_Q} \eta(X)} \quad (91)$$

$$S(X, b(X)) = \|\underline{d}(X)\|_{D_Q}^2 (1 - r^2(X)) \quad (92)$$

Le terme de corrélation est employé ici de façon clairement abusive puisque  $\underline{d}$  et  $\underline{d}(X)$  ne sont pas centrés. D'autre part,  $\|\underline{d}(X)\|_{D_Q}$  est une donnée du problème et donc dans cette de régression, la minimisation du STRESS est déterminée, dans le cadre général du Codage Multidimensionnel par la propriété suivante qui résulte immédiatement de (90), (91) et (92).

**Propriété :**

- (1) si  $X^*$  minimise  $S(X,b)$ , elle maximise  $r(X)$ .
- (2) si  $X$  maximise  $r(X)$ , le minimum de  $S(X,b)$  en  $(X,b)$  est atteint en:

$$X^* = \frac{\rho(X)}{\eta^2(X)} X \quad (93)$$

Par conséquent, la minimisation du STRESS est équivalente à la recherche de la configuration  $X$  qui réalise le maximum du rapport de deux normes:

$$q(X) = \frac{\rho(X)}{\eta(X)} = \frac{\text{Trace}(X N(X) {}^t X)}{\sqrt{\text{Trace}(X N_Q {}^t X)}}$$

Ce problème présente une forte analogie avec l'optimisation du rapport de Rayleigh caractéristique de l'Analyse en Composantes Principales. Cependant, on optimise alors un critère du type

$$\frac{\eta_1^2(X)}{\eta_2^2(X)}$$

où  $\eta_k^2(X)$  est défini par le choix d'une norme  $N^k$  pour  $k=1, 2$ . La complexité du problème résulte du fait que la norme  $\rho(X)$  fait intervenir une géométrie locale  $N(X)$  dépendant du point  $X$ . De plus l'Analyse en Composantes Principales optimise un rapport de Rayleigh homogène au rapport

$$\frac{\rho(X)}{\eta^2(X)} \text{ et non } \frac{\rho(X)}{\eta(X)}$$

Cependant les résultats de CROQUETTE (1980) (voir aussi CLARET (1986)) permettent de montrer l'identité des solutions de ces deux problèmes en Analyse en Composantes Principales, lorsque  $N(X)$  ne dépend pas de  $X$ .

On résoud ce problème en utilisant un algorithme proposé par ROBERT (1967) qui généralise l'algorithme de calcul des éléments propres de  $N_Q^+ N$  par la puissance itérée au

cas où la matrice  $N$  dépend de  $X$  et est alors notée  $N(X)$ . Ses propriétés de convergence sont connues (cf. ROBERT (1967), de LEEUW (1977), de LEEUW et HEISER (1980)).

**Algorithme :**

à l'itération (0) initialisation :  $X^{(0)}$  tel que  $\eta^2(X^{(0)}) = 1$

à l'itération (v)  $Y^{(v)} \in \partial\rho(X^{(v)})$

$X^{(v+1)} \in \partial\eta^0(Y^{(v)})$

normalisation  $\eta^2(X^{(v+1)}) = 1$

où  $\eta^0(\cdot)$  est la jauge polaire de  $\eta(\cdot)$ . Ici  $\eta(X)$  est une norme Euclidienne, dont la norme polaire est définie de façon classique par la métrique inverse (généralisée)  $N_Q^+$ . par suite les résultats exposés plus haut montrent que l'on peut choisir

$$\begin{aligned} Y^{(v)} &= N(X^{(v)}) {}^t X^{(v)} M \\ Z^{(v)} &= N_Q^+ Y^{(v)} = N_Q^+ N(X^{(v)}) {}^t X^{(v)} M \end{aligned}$$

et donc

$$\begin{aligned} \|Z^{(v)}\|_{N_Q, M^{-1}}^2 &= \text{Trace}(M^{-1} {}^t Z^{(v)} N_Q Z^{(v)} M) \\ &= \text{Trace}(X^{(v)} N(X^{(v)}) N_Q^+ N(X^{(v)}) {}^t X^{(v)} M) \\ &= \|X^{(v)}\|_{M, A^{(v)}}^2 \end{aligned}$$

où

$$A^{(v)} = N(X^{(v)}) N_Q^+ N(X^{(v)})$$

Soit, après normalisation

$$\begin{aligned} {}^t X^{(v+1)} &= \frac{1}{\|Z^{(v+1)}\|_{N_Q, M}} N_Q N(X^{(v)}) {}^t X^{(v)} M \\ &= \frac{1}{\|X^{(v)}\|_{M, A^{(v)}}^2} N_Q^+ N(X^{(v)}) {}^t X^{(v)} M \end{aligned}$$

**Remarque 1 :** Au voisinage du minimum,  $\delta_{ij}$  est peu différent de  $d_{ij}(X)$  et donc  $N(X)$  est proche de  $N_Q$ .

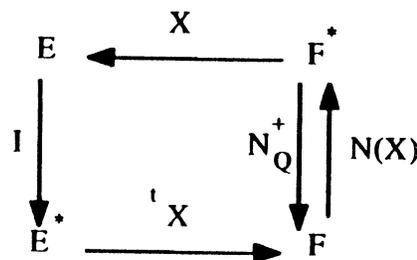
**Remarque 2 :** La solution  $X$  est de plus une solution stationnaire de l'équation

$$\begin{aligned} {}^t X &= \frac{1}{\|Z\|_{N_Q, M^{-1}}} N_Q N(X) {}^t X M \\ &= \frac{1}{\|X\|_{M, A(X)}} N_Q^+ N(X) {}^t X M \end{aligned} \quad (94)$$

où

$$A(X) = N(X)N_Q^+ N(X)$$

et  $N_Q^+ N(X)$  ne s'interprète pas, à notre connaissance comme un opérateur proportionnel au projecteur orthogonal sur le sous-espace vectoriel  $F_X$  de  $F$  engendré par  ${}^t X$ . de LEEUW et HEISER (1982) traitent le cas  $M=I$ . L'introduction d'une métrique  $M$  quelconque est en particulier utile pour rendre compte des rotations de la solution qui interviennent dans la phase d'interprétation des résultats. Lorsque nous adoptons la convention de ces auteurs, la situation à l'optimum est alors résumée par le schéma de dualité



dans lequel puisque  $X$  est fixé,  $N(X)$  l'est aussi. On résoud alors un problème classique aux valeurs propres, et les axes de la représentation s'interprètent de façon identique au cas de l'Analyse en Coordonnées Principales du triple  $(X, N(X), M)$  où  $M=I$ . Posons

$$\lambda = \| X \|_{I, A(X)}$$

alors (94) s'écrit de façon équivalente en fonction de la matrice  $W_X = N(X)^+$

$${}^t X = \lambda W_X N_Q {}^t X \quad (95)$$

Les liens existant entre les deux approches sont alors clairement établis en remarquant que les vecteurs propres de l'opérateur  $N_Q^+ N(X)$  associés à la solution optimale fournie par l'algorithme de de Leeuw et Heiser sont aussi les vecteurs propres de l'opérateur  $W_X N_Q$ . L'analyse du triple  $(I, \delta_I, Q_{|X|})$ , exprimée dans  $F$  en termes d'optimisation de la fonction STRESS fait donc intervenir des outils algébriques analogues à ceux qui interviennent en Analyse en Coordonnées Principales, la difficulté du traitement résultant en fait de l'impossibilité de fixer à priori la géométrie de  $F$ . Cependant, l'équation (94) montre aussi que ces vecteurs propres sont associés à une seule valeur propre multiple. Il n'existe donc pas de hiérarchisation possible des vecteurs propres de  $W_X$  et toute rotation des axes fournit une solution d'égale qualité.

Les expressions (85) et (85-bis) définissent la fonction STRESS dans  $F \wedge F$ . On vérifie ainsi que si on impose une contrainte supplémentaire du type

$$\| \underline{d}(X) \|_{D_Q}^2 = \sum_{i < j} \sum q_{ij} d_{ij}^2(X) = k$$

où k est par exemple égal à n ou encore à  $\sum_{i < j} \sum q_{ij} \delta_{ij}^2$

alors seulement, la minimisation de la fonction STRESS est équivalente à la maximisation de

$$\rho(X) = \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q} = \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(X)$$

Par suite de Leeuw et Heiser ont développé une méthode distincte de la méthode de quantification IV de Hayashi.

D'autre part, la construction d'un plongement approché de I dans un espace affine de faible dimension pose deux problèmes, d'une part celui de la détermination correcte de la dimension de l'espace de représentation, et d'autre part, celui du degré d'euclidéanité de la fonction  $\delta$ . On peut évaluer en pratique l'influence conjointe de ces deux facteurs. En effet supposons que  $\delta$  soit Euclidienne. Alors si la dimension fixée par l'utilisateur majore la dimension p d'un plongement irréductible de  $(I, \delta)$ , la solution  $X^*$  construite annule la fonction STRESS. Notons X la solution obtenue en r dimension avec  $r < p$ . On vérifie alors la proposition

**Proposition 7:** *Le biais d'estimation résultant du choix de la représentation X au lieu de  $X^*$  est égal à*

$$1 - b(X)$$

**Preuve :** La solution  $\underline{d}(X)$  est optimale si et seulement si elle annule les équations normales

$$\frac{\partial \text{STRESS}}{\partial \underline{d}(X)} = 2D_Q[\underline{d}(X) - \underline{\delta}] = 0$$

et donc à l'optimum

$$0 = \langle \underline{\delta}, \underline{d}(X) - \underline{\delta} \rangle_{D_Q} = \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q} - \|\underline{\delta}\|_{D_Q}^2 \quad (96)$$

c'est à dire

$$\sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(X) = \sum_{i < j} \sum q_{ij} \delta_{ij}^2$$

Par suite, à l'optimum, la fonction STRESS vérifie en conséquence de (96)

$$\text{STRESS}^*(X) = \langle \underline{d}(X^*), \underline{d}(X^*) - \underline{\delta} \rangle_{D_Q} = \|\underline{d}(X^*)\|_{D_Q}^2 - \langle \underline{\delta}, \underline{d}(X^*) \rangle_{D_Q}$$

et donc l'équation (90) permet d'écrire

$$\text{STRESSFORM1}^*(X) = \frac{\text{STRESS}^*(X)}{\|\underline{d}(X^*)\|_{D_Q}^2} = 1 - \frac{\langle \underline{\delta}, \underline{d}(X^*) \rangle_{D_Q}}{\|\underline{d}(X^*)\|_{D_Q}^2} = 1 - b(X^*)$$

A l'optimum,  $\text{STRESSFORM1}(X)$  est égal à 0 et la valeur théorique de  $b(X^*)$  est donc 1. La valeur observée du STRESS est en fait

$$\text{STRESS}(X) = \|\underline{d}(X) - \underline{\delta}\|_{D_Q}^2 = \|\underline{d}(X)\|_{D_Q}^2 + \|\underline{\delta}\|_{D_Q}^2 - 2 \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q}$$

et

$$\text{STRESSFORM1}(X) = \frac{\|\underline{\delta}\|_{D_Q}^2}{\|\underline{d}(X)\|_{D_Q}^2} + 1 - 2 \frac{\langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q}}{\|\underline{d}(X)\|_{D_Q}^2} = 1 - b(X)$$

et donc le biais d'approximation résultant du choix de la solution  $X$  au lieu de  $X^*$  est égal à

$$\text{STRESSFORM1} - \text{STRESSFORM1}^* = \text{STRESSFORM1} = 1 - b(X) \quad \parallel$$

L'intérêt pratique d'une telle propriété est lié à l'utilisation du diagramme de Shepard (cf. KRUSKAL et WISH(1970)). Cet outil d'interprétation représente chaque couple  $(i,j)$  d'éléments de  $I$  par le point de  $\mathbb{R}^2$  de coordonnées  $(\delta_{ij}, d_{ij}(X))$ . Si l'optimum de STRESS est atteint,  $b(X) = 1$  et donc les couples  $(i,j)$  sont situés sur la première bissectrice des axes, sinon  $b(X) \neq 1$ , de sorte que l'ajustement du nuage de points massiques  $\{(\delta_{ij}, d_{ij}(X)), q_{ij}\}; i,j \in [n]\}$  par une droite donne une droite de pente différente de 1, ce qui indique un biais, qui peut être dû soit à un problème de convergence de l'algorithme, soit au choix d'une dimension de l'espace de représentation  $r$  trop petite, soit encore, dans le cas d'une fonction de dissimilarité quelconque, au fait que  $\delta$  n'est pas Euclidienne.

Le rôle de  $b(X)$  comme indicateur du biais d'estimation de  $\Delta$  par  $D(X)$  se généralise à la classe des indices d'ajustement définie par OBENCHAIN (1973). Introduisons en effet un coefficient d'homotétie  $b$  dans l'expression de  $\psi(\alpha, \beta, Q)$

$$\psi(\alpha, \beta, Q) = \sum_{i < j} \sum q_{ij} (\delta_{ij}^\alpha - b d_{ij}^\alpha(X))^\beta$$

Alors le minimum de  $\psi(\alpha, \beta, Q)$  en fonction de  $b$  est obtenu lorsque  $\frac{\partial \psi}{\partial b} = 0$

La solution  $b(X)$  de cette équation est particulièrement simple lorsque  $\beta = 2$ , c'est à dire, lorsqu'on se restreint à la classe des indices d'ajustement au sens des moindres carrés. En effet on obtient alors

$$b(X) = \frac{\sum_{i < j} \sum q_{ij} \delta_{ij}^{\alpha} d_{ij}^{\alpha}(X)}{\sum_{i < j} \sum q_{ij} d_{ij}^{2\alpha}(X)}$$

et l'expression de  $\psi(\alpha, 2, Q)$  montre que le tracé du diagramme de Shepard ( $\delta^{\alpha}, \underline{d}^{\alpha}(X)$ ) fournit une aide graphique analogue au cas précédent, permettant de détecter les biais d'estimation des distances. Nous utilisons plus bas ce résultat dans le cas particulier  $\alpha=2$  qui conduit aux méthodes MINRES 2.

Si  $b(X) > 1$  (resp.  $< 1$ ), la solution tend à surestimer les grandes (resp. petites) dissimilarités et à sous-estimer les petites (resp. grandes).

**Remarque 1 :** Si  $Q = \mathbf{1}\mathbf{1}$ , alors  $N_Q = 2nQ_1$  et par suite  $N_Q^+ = \frac{1}{2n} Q_1$

On obtient pour  $M=I$ , puisque  $N(X)$  a pour support  $F_1$

$$Z = N_Q^+ N(X)X = \frac{1}{2n} Q_1 N(X)X = \frac{1}{2n} N(X)X$$

Par suite, on retrouve l'algorithme proposé par GUTTMAN (1968) et  $Z$  a pour terme général

$$z_i^s = \frac{1}{n} \sum_{j \neq i} \frac{\delta_{ij}}{d_{ij}(X)} (x_i^s - x_j^s) = \frac{1}{n} \sum_{j \neq i} \delta_{ij} \gamma_{ij}$$

où  $\gamma_{ij}$  définit le sous-gradient de  $\underline{d}(X)$  par

$$\gamma_{ij} = \begin{cases} \frac{(x_i^s - x_j^s)}{d_{ij}(X)} & \text{si } d_{ij}(X) \neq 0 \\ 0 & \text{si } d_{ij}(X) = 0 \end{cases}$$

Soit  $\text{signe}(x)$  la fonction définie par

$$\text{signe}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases}$$

alors dans le cas unidimensionnel, on vérifie que

$$z_i = \frac{1}{n} \sum_{j \neq i} \delta_{ij} \text{signe}(x_i - x_j)$$

**Remarque 2 :** Les algorithmes de sous-gradient sont caractérisés par une vitesse de convergence lente. Par suite de Leeuw et Heiser proposent d'accélérer la convergence de leur algorithme en posant

$$\tilde{Z}^{v+1} = Z^v + \alpha (X^v - Z^v) \quad \alpha \in ]-1, 1[$$

Si  $v_\alpha$  est la vitesse de convergence pour le cas  $\alpha$ , alors la limite  $X^\infty$  de la suite  $X^v$  vérifie

$$\begin{aligned} X^{v+1} - X^\infty &= (1 - \alpha)Z^v + \alpha X^v - X^\infty \\ &= (1 - \alpha)(Z^v - X^\infty) + \alpha(X^v - X^\infty) \end{aligned}$$

et donc

$$\eta(X^{v+1} - X^\infty) \leq (1 - \alpha)v_0\eta(X^v - X^\infty) + \alpha\eta(X^v - X^\infty)$$

Posons donc  $\alpha = \epsilon_0 - 1$  et  $v_0 = 1 - \epsilon_1$  pour  $\epsilon_0$  et  $\epsilon_1$  assez petits. On a alors

$$v_\alpha = 1 - \epsilon_1 + \epsilon_1(\epsilon_0 - 1) = (1 - \epsilon_1)^2 + \epsilon_1(\epsilon_0 - \epsilon_1) \cong (1 - \epsilon_1)^2 = v_0^2$$

et le nombre d'itérations est réduit de moitié lorsqu'on fait ce choix pour  $\alpha$ .

#### VI.4.3. La méthode MINRES 2

On se propose d'étudier les méthodes d'analyse du triple  $(I, \delta, Q_{|x|})$  associées au modèle

$$\delta_{ij}^2 = d_{ij}^2(X) + \epsilon_{ij} \quad i, j \in [n] \quad (97)$$

La minimisation de la moyenne quadratique des résidus  $\epsilon_{ij}$  conduit aux méthodes appelées ici MINRES 2. La fonction critère est alors  $\psi(2,2,Q)$ , appelée SSTRESS par TAKANE et al. (1977). Ces auteurs proposent en effet un algorithme efficace d'optimisation de la fonction

$$\text{SSTRESS}(X) = \sum_{i < j} \sum q_{ij} (\delta_{ij}^2 - d_{ij}^2(X))^2 \quad (98)$$

La méthode MINRES 2 est fondée sur l'optimisation de la fonction  $\text{SSTRESS}(X)$  qui présente l'intérêt d'être continuellement différentiable en tout point  $X$  et conduit donc à des algorithmes numériques plus efficaces en théorie que la méthode MINRES. La contrepartie de cette qualité est l'importance accrue des grandes dissimilarités dans l'établissement de l'optimum. Récrivons la fonction objectif sous une forme équivalente

$$\begin{aligned} \text{SSTRESS}(X) &= \sum_{i < j} \sum q_{ij} \epsilon_{ij}^2 \\ &= \text{Trace}(W_\epsilon N_Q) = \text{Trace}(\Psi_\epsilon D_Q) \end{aligned}$$

où

$$\begin{aligned}
W_\varepsilon &= -\frac{1}{2} Q_1 E^{*2} {}^t Q_1 = -\frac{1}{2} (Q_1 \Delta^{*2} {}^t Q_1 - Q_1 D^{*2}(X) {}^t Q_1) \\
&= W_\delta - W_X \\
\Psi_\varepsilon &= {}^t W_\varepsilon T = {}^t W_\delta T - {}^t T T \\
\Psi_\varepsilon &= \Psi_\delta - \Psi_X \\
\text{SSTRESS}(X) &= \text{Trace}(W_\varepsilon D_q) - \text{Trace}(W_\varepsilon Q)
\end{aligned}$$

L'expression précédente montre la complexité des problèmes rencontrés en Codage Multidimensionnel fondé sur l'optimisation de la fonction SSTRESS.

HAYASHI (1974a, b, 1975) semble être à l'origine de cette formulation du problème de Codage Multidimensionnel. Il traite une classe de problèmes plus vaste qu'il n'est discuté ici puisqu'il considère le cas de fonctions de dissimilarité mesurées sur des échelles nominales et ordinales et pas seulement d'échelle de rapport comme il est supposé en général dans ce mémoire ou d'échelle d'intervalles lorsqu'on traite le problème de la constante additive. Par suite, l'algorithme proposé par Hayashi est omnibus et d'une efficacité moindre, malgré l'intérêt que présente sa logique de pas à pas. Celle-ci s'interprète comme un exemple de méthode de projections en drapeau (ou Projection Pursuit, HUBER (1985)) et a donc pour intérêt de construire un drapeau de sous-espaces (c'est à dire tels que  $E_r$  est un sous espace vectoriel de  $E_{r+1}$ ) de représentation de  $I$  dans  $E$  alors que les méthodes présentées dans ce chapitre ne vérifient pas de propriété d'emboîtement des espaces solutions.

Dans  $F \wedge F$ , on observe que

$$\begin{aligned}
\text{SSTRESS}(X) &= \|\underline{\delta}^{*2} - \underline{d}^{*2}\|_{D_Q}^2 \\
&= \|\underline{\delta}^{*2}\|_{D_Q}^2 + \|\underline{d}^{*2}\|_{D_Q}^2 - 2 \langle \underline{\delta}^{*2}, \underline{d}^{*2} \rangle_{D_Q}
\end{aligned}$$

Par suite, la minimisation de la fonction objectif SSTRESS(X) équivaut à la maximisation de l'indice

$$\theta_1(\underline{\delta}^{*2}, \underline{d}^{*2}(X)) = \sum_{i < j} \sum q_{ij} \delta_{ij}^2 d_{ij}^2(X)$$

si et seulement si on impose de plus une contrainte de la forme

$$\|\underline{d}^{*2}\|_{D_Q}^2 = \sum_{i < j} \sum q_{ij} d_{ij}^4(X) = k$$

On obtient d'autr part, une formulation équivalent du problème sous la forme :

**Proposition 8:** *La minimisation de  $\theta_1(\delta^{*2}, d^{*2}(X))$  en fonction de  $X$  est équivalente à la minimisation de*

$$\text{SSTRESSFORM1}(X) = \frac{\text{SSTRESS1}(X)}{\|\underline{d}^*\|_{D_Q}^2}$$

La preuve de ce résultat est immédiate. Cette proposition pose la question de la normalisation de la fonction  $\text{SSTRESS}(X)$  qui est discutée par YOUNG et al. (1979, 1980). La minimisation de  $\text{SSTRESS}(X)$  fournit donc une méthode distincte de l'approche de Hayashi et de celle de Guttman-Kruskal. De plus, nous avons vu au chapitre précédent que l'Analyse en Coordonnées Principales minimise un critère de la forme

$$\text{STRAIN}(X) = \frac{1}{4} \|\underline{\delta}^* - \underline{d}^*\|_{N_Q \otimes N_Q}^2$$

et donc les deux méthodes diffèrent par la géométrie Euclidienne choisie sur  $F \wedge F$  afin de réaliser l'ajustement au sens des moindres carrés pondérés. Le cadre le plus simple de comparaison est obtenu lorsque  $q_{ij}$  est égal à 1 pour tout couple  $(i,j)$  d'éléments de  $[n]$ . OBENCHAIN (1973) établit de plus le résultat suivant, qui joue un rôle central dans la comparaison de ces deux méthodes. Rappelons que

$$(H_\epsilon = \underline{w}_\epsilon \mathbf{1} + \mathbf{1} \underline{w}_\epsilon - 2W_\epsilon) \Leftrightarrow (\underline{w}_\epsilon = \alpha_1(H_\epsilon)) \quad (99)$$

où le vecteur  $\underline{w}_\epsilon$  s'écrit avec des notations évidentes, puisque  $H_\epsilon = H_\delta - H_{d(X)}$

$$\begin{aligned} \underline{w}_\epsilon &= (W_\epsilon * \mathbf{1}) \mathbf{1} = (W_\delta * \mathbf{1}) \mathbf{1} - (W_X * \mathbf{1}) \mathbf{1} \\ \underline{w}_\epsilon &= \underline{w}_\delta - \underline{w}_X \end{aligned}$$

Par suite il est possible de comparer les fonctions objectif optimisées d'une part en Analyse en Coordonnées Principales et d'autre part en Codage Multidimensionnel, selon la méthode MINRES 2. Cette comparaison est présentée ici dans le cas d'une pondération triviale afin de favoriser l'interprétation intuitive du résultat. L'expression mathématique du cas général est obtenue par les mêmes voix.

**Proposition 9:** *Si  $Q$  est la pondération triviale et  $X$  est une représentation centrée*

$$\text{SSTRESS}(X) = 4 \text{STRAIN}(X) + 2 \langle \mathbf{1}, \underline{w}_\epsilon \rangle^2 + \|\underline{w}_\epsilon\|^2 \quad (100)$$

**Preuve :**

$$\begin{aligned} \text{SSTRESS}(X) &= \|\Delta * \Delta - D(X) * D(X)\|^2 = \|H_\epsilon\|^2 \\ \text{STRAIN}(X) &= \|W_\delta - W_X\|^2 = \|\underline{w}_\epsilon\|^2 \end{aligned}$$

et par définition de  $\alpha_1$

$$\underline{w}_\varepsilon = \underline{u}_\varepsilon - \frac{1}{2} \left( \frac{1}{n} \mathbf{1}^t \mathbf{H}_\varepsilon \mathbf{1} \right) \mathbf{1} = \underline{u}_\varepsilon - \frac{1}{2} (\delta_{..}^2 - d_{..}^2(X)) \mathbf{1}$$

où l'on a posé

$$\underline{u}_\varepsilon = \frac{1}{n} \mathbf{H}_\varepsilon \mathbf{1}$$

$\underline{u}_\varepsilon$  est le vecteur de  $i$ -ième coordonnée égale à  $\delta_{i.}^2 - d_{i.}^2(X)$

La proposition résulte alors immédiatement de (99) et du fait que  $\mathbf{W}_\varepsilon \mathbf{1} = 0$ . ¶

**Remarque :** Si  $\delta$  est une dissimilarité Euclidienne,  $\underline{w}_\varepsilon$  a pour  $i$ -ième coordonnée

$$w_i = \delta_{i.}^2 - d_{i.}^2(X) - \frac{1}{2} (\delta_{..}^2 - d_{..}^2(X)) = I_{z_i} - I_{x_i}$$

où

$$I_{x_i} = d_{i.}^2(X) - \frac{1}{2} d_{..}^2(X)$$

désigne l'inertie du point  $P_i$  de coordonnées  $x_i$  par rapport au barycentre du nuage de points  $\{(P_i, \frac{1}{n}) ; i \in [n]\}$

D'autre part, lorsqu'on impose de plus la normalisation

$$\|\underline{\delta}\|_{D_Q}^2 = \|\underline{d}(X)\|_{D_Q}^2$$

c'est à dire

$$\delta_{..}^2 = d_{..}^2(X)$$

on obtient la relation

$$\underline{w}_\varepsilon = \underline{u}_\varepsilon$$

La proposition 9 fournit une borne inférieure de SSTRESS(X) puisque l'équation (100) montre que

$$\text{SSTRESS}(X) \geq 4 \text{ STRAIN}(X)$$

L'Analyse en Coordonnées Principales est caractérisée par une moindre prise en considération des termes diagonaux de la matrice  $\mathbf{W}_R$ , puisqu'elle minimise les écarts non diagonaux. Une conséquence pratique importante de cette observation est qu'elle permet de considérer l'Analyse en Coordonnées Principales comme un outil privilégié de construction d'une configuration  $X$  permettant d'initialiser une procédure de Codage Multidimensionnel optimisant la fonction SSTRESS.

**Remarque :** Les arguments développés lors de la comparaison entre les indices SSTRESS(X) et STRAIN(X) montrent en travaillant sur la fonction de dissimilarité  $\sqrt{\delta}$  que de façon analogue

$$\text{STRESS}(X) \geq 4 \sigma(X)$$

et bien que l'interprétation mécaniste de  $w_\epsilon$  ne soit plus pertinente, on peut utiliser l'Analyse en Coordonnées Principales du triple  $(I, \sqrt{\delta}, P_1)$  où  $P_1$  est la pondération uniforme, comme initialisation de la méthode de Codage Multidimensionnel MINRES.

Remarquons maintenant que l'extension  $\alpha_1^+$  de  $\alpha_1$  à  $\mathcal{Y}_n$  associée à toute matrice diagonale  $D_u$ , une matrice symétrique définie par le lemme suivant

**Lemme 4:** Soit  $D_u$  une matrice diagonale d'ordre  $n$  de terme générale  $u_i$ , et  $\underline{w}$  un vecteur de  $F^*$  de terme général

$$w_i = \frac{1}{n} (u_i - \frac{1}{n} \sum_j u_j)$$

Alors

$$W_u = \alpha_1(D_u) = -\frac{1}{2} D_u + \frac{1}{2} (\underline{w} \underline{1} + \underline{1} \underline{w})$$

**Preuve :**

$$\begin{aligned} W_u &= -\frac{1}{2} Q_1 D_u {}^t Q_1 \\ &= -\frac{1}{2} (D_u - \frac{1}{n} \underline{1} \underline{1} D_u - \frac{1}{n} D_u \underline{1} \underline{1} + \frac{1}{n^2} \underline{1} \underline{1} D_u \underline{1} \underline{1}) \end{aligned}$$

qui conduit au résultat.  $\square$

Par suite la perturbation de  $H_\epsilon$  par une matrice diagonale ne se traduit pas par une perturbation de la diagonale de  $W_\epsilon$ . D'autre part si  $D_u$  est proportionnelle à l'identité, l'analyse de  $H_n = W_\epsilon + uI$  conduit à étudier

$$W_u = W_\epsilon - \frac{1}{2} uI$$

puisque  $\underline{w}$  est alors le vecteur nul. L'ajout d'une constante additive à  $H_\epsilon$  conduit donc à l'ajout d'une constante additive à  $W_\epsilon$ . Si donc les modèles factoriels formulés sur  $H$  et sur  $W$  ne se correspondent pas, le problème de la constante additive présente lui une symétrie entre les deux situations.

TAKANE et al. (1977) résolvent le problème MINRES 2 à l'aide d'un algorithme de moindres carrés alternés. Le principe d'une telle itération généralise au cadre non linéaire,

la procédure de Gauss Seidel dont la flexibilité et l'efficacité dans les problèmes d'estimation linéaire est fondée sur la partition de l'ensemble des paramètres à estimer, que l'on évalue de façon alternée. Dans le cadre de l'analyse des données de profil, cette démarche algorithmique est adoptée par WOLD (1966a), LYTTKENS (1966) donne une interprétation géométrique de l'itération et DAUGAVET (1968) étudie les propriétés de convergence de la méthode. WOLD (1966b) étudie son application au cadre non linéaire (cf. WOLD et LYTTKENS (1969)). Dans le cas présent, la méthode des moindres carrés alternés est décrite sommairement comme suit. Pour tout élément  $i$  de  $[n]$ , supposons que les  $x_j$  sont fixés pour tout  $j \neq i$ . On construit alors une représentation  $x_i$  de  $P_i$  dans  $E$  de façon à minimiser  $SSTRESS(X)$  et on itère le procédé en décrivant  $[n]$  tant que la valeur de la fonction critère diminue. Nous renvoyons le lecteur à TAKANE et al. (1977) pour une présentation moins succincte de la méthode et de ses extensions. Nous préférons exposer ici l'approche du problème proposée par BROWNE (1985, 1987), dont les éléments constitutifs apportent un éclairage intéressant sur la nature statistique du problème traité par la méthode MINRES 2.

Soit  $X$  un élément de  $\mathcal{M}_{r,n}(\mathbb{R})$  qui induit la matrice de distance Euclidienne  $D(X)$ . Alors

$$W_X = {}^t X M X$$

$$D^{*2}(X) = \underline{w}_X {}^t \underline{1} + {}^t \underline{1} \underline{w}_X - 2W_X$$

où

$$\underline{w}_X = (W_X * I) \underline{1}$$

Posons dans la suite

$$H_\delta = \Delta * \Delta$$

$$H_X = D(X) * D(X)$$

de sorte que

$$Q_\epsilon = Q * (H_\delta - H_X) = Q * H_\delta - Q * H_X$$

et notons  $q_\epsilon = Q_\epsilon \underline{1}$ .  $D_\epsilon$  est alors la matrice diagonale définie par  $D_\epsilon \underline{1} = q_\epsilon$ . On vérifie alors le résultat suivant

**Lemme 5:** *La fonction  $SSTRESS(X)$  est différentiable en tout point  $X$  et*

$$\frac{\partial SSTRESS(X)}{\partial {}^t X} = 8 (Q_\epsilon - D_\epsilon) {}^t X \quad (101)$$

**Preuve :** Elle résulte de l'expression de la dérivée partielle de  $SSTRESS(X)$  en fonction des coordonnées  $x_i$  :

$$\frac{\partial \text{SSTRESS}(X)}{\partial x_i^s} = 4 \sum_{k < l} q_{kl} (\delta_{kl}^2 - d_{kl}^2(X)) (\delta_i^k - \delta_i^l) (x_k^s - x_l^s)$$

où  $\delta_i^l$  est le symbole de Kronecker. Par suite

$$\begin{aligned} &= 4 q_{kl} (\delta_{kl}^2 - d_{kl}^2(X)) (\delta_i^k - \delta_i^l) (x_k^s - x_l^s) \\ &= 4 (q_{kl} \delta_{kl}^2 - q_{kl} d_{kl}^2(X)) (\delta_i^k - \delta_i^l) (x_k^s - x_l^s) \end{aligned}$$

qui donne le résultat souhaité après réorganisation.  $\quad \parallel$

**Remarque :** On peut obtenir formellement l'expression (101) par dérivation matricielle à l'aide des formules données par BALESTRA (1975) ou MAGNUS et NEUDECKER (1988).

L'expression (101) généralise le résultat de Browne qui ne considère que le cas d'une pondération triviale  $Q_{|X|} = \mathbf{1}^l \mathbf{1}$ . Elle s'exprime aussi en fonction des matrices

$$\begin{aligned} N(\delta^2) &= - (Q_\delta - D_\delta) = -T D_{Q(\delta)} {}^l T \\ N(X) &= - (Q_X - D_X) = -T D_{Q(X)} {}^l T \end{aligned}$$

On a en effet

$$N(\epsilon) = - (Q_\epsilon - D_\epsilon) = N(X) - N(\delta^2) = -T (D_{Q(X)} - D_{Q(\delta)}) {}^l T = -T D_{Q(\epsilon)} {}^l T$$

Cette relation s'interprète en remarquant que l'on munit alors  $F \wedge F$  de la pondération  $Q_\epsilon$  définie sur  $|X|$  qui est bien symétrique mais non nécessairement positive et par suite  $N(\epsilon)$  n'est pas nécessairement gramienne. D'autre part l'équation (101) montre la différence qui existe entre l'approche de LEEUW et HEISER (1977) pour résoudre le problème MINRES et l'approche de Browne pour résoudre le problème MINRES 2.

i) de Leeuw et Heiser considèrent la pondération positive

$$q_{ij}(X) = q_{ij} \frac{\delta_{ij}}{d_{ij}(X)}$$

ii) Browne considère la pondération de signe quelconque

$$q'_{ij}(\epsilon) = q_{ij} (d_{ij}^2(X) - \delta_{ij}^2)$$

Mais  $-q'_{ij}(\epsilon)$  est positive si  $\delta$  est Euclidienne et  $X$  est obtenue par projection.

L'expression (101) généralise aussi le travail de KAMENSKI (1977) au cas de pondérations quelconques, en ce qu'elle caractérise la solution  ${}^l X$  comme un élément du noyau de l'application linéaire de matrice  $N(\epsilon)$  puisque  $X$  est solution si elle vérifie

$$N(\epsilon) {}^l X = \mathbf{0} \tag{102}$$

**Remarque :** Lorsque  $F$  est muni de la géométrie (a priori non Euclidienne) induite par  $N(\epsilon)$ , on note que

$$\begin{aligned}\text{Trace}(XN(\epsilon)'X) &= \text{Trace}(W_X N(\epsilon)) \\ &= \text{Trace}({}'TW_X T D_{Q\epsilon}) = \text{Trace}(\Psi_X D_{Q\epsilon})\end{aligned}$$

Cette fonction est reliée à la fonction SSTRESS puisqu'on obtient formellement

$$\begin{aligned}\text{Trace}(\Psi_X D_{Q\epsilon}) &= -\frac{1}{2} \text{Trace}({}'TD^{*2}(X)TD_{Q\epsilon}) \\ &= -\frac{1}{2} \sum_{i < j} \sum q_{ij} (d_{ij}^2(X) - \delta_{ij}^2) d_{ij}^2(X) \\ &= -\frac{1}{2} (\text{SSTRESS}(X) + \langle \underline{\delta}^{*2}, \underline{d}^{*2}(X) \rangle_{D_Q} - \|\underline{\delta}^{*2}\|_{D_Q}^2)\end{aligned}$$

Par suite le résultat fondamental justifiant la méthode de Browne se généralise immédiatement à l'aide des produits d'Hadamard. Considérons en effet une matrice de représentation  $X$  élément de  $\mathcal{M}_{r,n}(\mathbb{R})$  qui factorise la matrice  $Q^*W_X$  sous la forme  $Q^*W_X = 'XX$ , alors on a le théorème suivant:

**Théorème 2:**  $X$  vérifie les propriétés

- i)  $N(\epsilon)'X = 0$
- ii)  $'X = Q_1'X$
- iii)  $X'X = D_\lambda$

si et seulement si

$$W(\epsilon)'X = 'XD_\lambda \tag{103}$$

où  $W(\epsilon)$  est déduite de  $N(\epsilon)$  par la relation

$$W(\epsilon) = \alpha_1(Q_\delta - D_\epsilon)$$

**Preuve :**

$$\begin{aligned}W(\epsilon)'X &= -\frac{1}{2} Q_1 (Q_\delta - D_\epsilon)'Q_1'X = -\frac{1}{2} Q_1 (N(\epsilon) + Q_X)'Q_1'X \\ &= -\frac{1}{2} Q_1 N(\epsilon)'Q_1'X - \frac{1}{2} Q_1 Q_X'Q_1'X \\ &= -\frac{1}{2} N(\epsilon)'X - \frac{1}{2} Q_1 Q_X'Q_1'X = -\frac{1}{2} Q_1 Q_X'Q_1'X \\ &= Q^*W_X'X = 'XX'X = 'XD_\lambda\end{aligned}$$

puisque

$$\begin{aligned}'Q_1'X &= Q_1'X = Q_1'X \\ Q^*W_X &= 'XX \quad \parallel\end{aligned}$$

Remarquons de plus que

$$Q_\delta - D_\varepsilon = Q_\delta - (D_\delta - D_X) = N(\delta^2) + D_X$$

Par suite, le problème est résolu à l'aide d'une perturbation des termes diagonaux de la matrice  $N(\delta^2)$  par une matrice diagonale dépendant de  $X$ . La solution du problème ne peut donc être construite que par une procédure itérative.

D'autre part, si  $Q_\delta$  est un élément de  $\mathcal{Y}_H$ , il n'en est pas de même de la perturbation  $D_\varepsilon$ . On a donc

$$W(\varepsilon) = \alpha_1^+(Q_\delta) - \alpha_1(D_\varepsilon)$$

c'est à dire

$$W(\varepsilon) = W_\delta - W_\varepsilon$$

où l'expression analytique de  $W_\varepsilon$  est donnée par le lemme 5. Celui-ci montre dans le cas d'une dissimilarité Euclidienne  $\delta(Z)$  que si les inerties de chacun des points  $P_i$  sont bien reconstruites par la configuration  $X$ , alors la perturbation  $W_\varepsilon$  est nulle.

**Remarque :** Les équations liant  $W(\varepsilon)$  à  $W_\delta$  montrent que l'on ne résoud pas un problème interprétable en termes d'Analyse Factorielle en Facteurs Communs, malgré l'analogie consistant à itérer une perturbation de la diagonale de la matrice d'intérêt. Cela tient à ce qu'ici, c'est  $Q_\delta$  et non  $W_\delta$  qui est ainsi perturbée. La forme de  $W_\varepsilon$  montre que la méthode MINRES 2 de Codage Multidimensionnel ne généralise pas à proprement parler l'Analyse Factorielle Classique puisque les termes  $w_i$  dépendent des  $u_i$ . C'est donc un problème distinct.

La généralisation de l'approche de Browne au cas du triplet  $(I, \delta, Q_{|X|})$  où la pondération  $Q_{|X|}$  est non triviale, conduit à opérer sur les matrices  $Q^*W_\delta$  et  $Q^*H_\delta$ . On observe ainsi le rôle très distinct joué par les pondérations  $Q_{|X|}$  de forme générale par comparaison avec les traitements présentés en analyse du triple  $(I, \delta, P_I)$ . Cela conduit à s'interroger sur la pertinence d'une méthode d'Analyse en Coordonnées Principales du triple  $(I, \delta^*, \Lambda_I)$  où  $\Delta^* = \Delta^*Q$  et  $\Lambda_I$  pondère chaque unité statistique par  $\frac{1}{n}$ , ou plus généralement par la masse réelle  $\lambda_i$ . Une telle méthode d'analyse conduirait cependant à une approximation faible des dissimilarités.

Cependant, l'application de la proposition de Obenchain présentée plus haut, montre que cette analyse fournit une bonne initialisation de la procédure de Browne ou de celle de Takane et al. en présence de pondérations générales. De plus, puisque

$$Q_{\delta} - D_{\varepsilon} = N(\delta^2) + D_X$$

si nous notons

$$W_{\delta}^N = \alpha_1(N(\delta^2))$$

il vient

$$W(\varepsilon) = W_{\delta}^N - \frac{1}{2} Q_1 D_X {}^t Q_1 = W_{\delta}^N + W_X^D$$

où  $W_X^D$  a pour terme général

$$(W_X^D)_{ij} = u_i \delta_i^j - \frac{1}{n} (u_i + u_j - u.)$$

calculé à partir de la solution  $\underline{u}$  du système d'équations

$$\underline{u} = H_X \underline{1}$$

BROWNE (1985) propose de déterminer  $\underline{u}$  par une procédure itérative, puisqu'à chaque étape de la procédure algorithmique, on suit les quatre étapes de calcul suivantes :

- (1) Calcul de  $X^v$  à partir des  $r$  vecteurs propres de  $Q * W^v(\varepsilon)$
- (2) Calcul de  $H_{X^v}$
- (3) Calcul de

$$\underline{u}^{v+1} = \alpha^v H_{X^v} \underline{1} + (1 - \alpha^v) \underline{u}^v$$

- (4) Calcul de  $Q * W^{v+1}(\varepsilon)$

L'expérience algorithmique d'une telle procédure est trop pauvre actuellement pour tirer des conclusions précises d'une comparaison avec l'algorithme des moindres carrés alternés décrit par TAKANE et al. (1977).

D'autre part l'intérêt pratique des méthodes exposées ci-dessus repose sur la démonstration de théorèmes d'existence et d'unicité du problème traité. Le cas du problème MINRES 2 est plus aisé que de celui de MINRES. En effet  $\mathfrak{D}_E(I)$  n'est pas un convexe fermé (cf. aussi d'AUBIGNY (1975), FICHET (1983)) et la solution ne peut pas être caractérisée en termes de projection (cf. aussi MEULMAN (1986)). Dans le cas du problème MINRES 2, l'existence et l'unicité résultent du fait que  $\mathfrak{D}_E^{*2}(I)$  est un cône convexe fermé inclus dans  $\mathcal{Y}_n^+$ . Par suite lorsqu'on n'impose pas de contraintes de dimension,  $\underline{d}(X) = \hat{\underline{d}}$  s'obtient par projection de  $\underline{\delta}$  sur  $\mathfrak{D}_E^{*2}(I)$ .

De façon analogue, en Analyse en Coordonnées Principales nous avons construit la solution  $\hat{w}$  en termes de projection. En Analyse en Coordonnées Principales, ce projecteur est linéaire et donc la projection est construite de façon analytique. Ce n'est pas le cas de la méthode MINRES 2. Plus précisément d'AUBIGNY (1975), FICHET (1983, 1985, 1987), CRITCHLEY (1978, 1983, 1985a, b) montrent que  $\mathcal{Y}_1^+$  et  $\mathcal{D}_E^{*2}(I)$  sont des cônes convexes fermés dans  $\mathcal{Y}_n^+$  et  $\mathcal{Y}_H$  respectivement. Par suite, l'existence et l'unicité de la solution résulte dans les deux cas du théorème classique (cf. LUENBERGER (1969), ROCKAFELLAR (1971))

**Théorème 3:**

- i) Si  $K$  est un cône convexe fermé de l'espace de Hilbert et  $K^*$  le cône polaire (ou dual) de  $K$ , alors  $K^*$  est un cône convexe fermé tel que  $(K^*)^* = K$
- ii) Le problème  $\text{Min} \{ \|h - x\| ; x \in K \}$  admet une solution unique  $\hat{x}$  caractérisée par
- (a)  $\hat{x} \in K$
  - (b)  $h - \hat{x} \in K$
  - (c)  $\langle \hat{x}, h - \hat{x} \rangle = 0$

Remarquons en effet que si  $Q_{|x|}$  est une pondération réelle positive et symétrique de matrice  $Q$ , alors  $Q^*W$  est semi-définie positive si et seulement si  $W$  l'est. Par suite les résultats classiques peuvent être généralisés par application directe du théorème précédent.

**Théorème 4:** Les problèmes suivants admettent une solution unique

- i)  $\text{Min} \{ \|Q^*(W_\delta - W_X)\|^2 ; W_X \in \mathcal{Y}_1^+ \}$
- ii)  $\text{Min} \{ \|Q^*(\Delta^{*2} - D^{*2}(X))\|^2 ; D^{*2}(X) \in \mathcal{D}_E^{*2}(I) \}$

CRITCHLEY (1985) remarque (proposition 12 p.9) que les cônes étudiés ont pour cônes polaires

- i)  $(\mathcal{Y}_1^+)^* = -\mathcal{Y}_1^+$
- ii)  $(D^{*2}(X))^* = \mathcal{Y}_1' = \{S' ; S \in \mathcal{Y}_1^+\}$

où  $S'$  est la matrice symétrique à diagonale nulle obtenue par projection de  $S$  sur  $\mathcal{Y}_H$ . Il utilise alors ces propriétés pour démontrer une conjecture de Gower

**Théorème 5 (CRITCHLEY (1985)) :** La dimension  $\hat{p}$  de toute représentation  $\hat{X}$  de  $I$  induite par la solution  $D(\hat{X})$  du problème MINRES 2 vérifie :

- i) Si  $n = 3$ ,  $\hat{p} = \text{rang}(\hat{W}_X)$

$$ii) \hat{p} \leq \text{rang}(\hat{W}_X)$$

où  $\hat{W}_X$  est la solution du problème d'Analyse en Coordonnées Principales.

Ce résultat justifie du point de vue de l'estimation de la dimension  $\hat{p}$  du plongement Euclidien approché de  $(I, \delta, Q_{I \times I})$ , l'utilisation de la solution de l'Analyse en Coordonnées Principales pour initialiser une procédure de Codage Multidimensionnel, fondée sur l'utilisation de la fonction critère SSTRESS.

#### VI.4.4. La méthode ELEGANT

L'algorithme de résolution du problème MINRES 2 proposé par Browne diffère de celui proposé par de LEEUW (1975) sous le nom de ELEGANT. (Voir aussi TAKANE (1976)). Le formalisme développé dans ce mémoire permet de vérifier que la présentation de l'algorithme ELEGANT dû à de LEEUW (1975) est totalement explicitable à l'aide des matrices  $\Psi$  et  $\Phi$  opérant dans  $F \wedge F$ . En effet, supposons, comme cet auteur, que la matrice  $Q$  est triviale et associions à toute matrice carrée  $N$ , la matrice obtenue en annulant les termes diagonaux

$$O(N) = N - N * I$$

Alors, puisque  $D_{D^*2(X)} = \Psi_X * I$ , on peut écrire

$$D_{\Delta^*2} - \Psi_X * I = D_{\Delta^*2} + (O(\Psi_X) - \Psi_X)$$

et donc en posant, comme de Leeuw

$$B(X) = T(D_{\Delta^*2} + O(\Psi_X))'T \quad (104)$$

on obtient pour transformée des résidus

$$N(\epsilon) = T(D_{\Delta^*2} - \Psi_X * I)'T = TB(X)'T - T\Psi_X 'T$$

et donc l'équation stationnaire

$$0 = N(\epsilon)'X = B(X)'X - N(X)'X$$

où

$$N(X) = T\Psi_X 'T$$

conduit à la résolution de l'équation stationnaire

$$N(X)'B(X)'X = 'X$$

de LEEUW (1975) résoud l'équation

$$B(X)'X = 'X\Lambda$$

et utilise donc clairement une méthode de perturbation. Bien qu'élégante (d'où son nom), cette solution pose de gros problèmes de mise en œuvre, en termes d'occupation de place en mémoire. Cependant Browne observe qu'une description de l'algorithme est possible dans F. Il écrit en effet l'itération sous la forme

$$(\alpha W(\epsilon) + (1-\alpha)'XX)'X = 'X\Lambda$$

où  $\alpha = \frac{2}{n}$ . Par suite cette réécriture de la solution de Leeuw perturbée  $N(\epsilon)$  à l'aide des équations tautologiques

$$('XX)'X = 'X\Lambda$$

équivalentes à l'équation (103). Chaque itération de l'algorithme ELEGANT comporte les trois étapes suivantes :

(1)  $X^v$  est obtenu par factorisation de  $Q^*W^v$  à partir des vecteurs propres associés aux  $r$  plus grandes valeurs propres

(2)  $W^v$  est obtenu par

$$\alpha^v W^v(\epsilon) + (1-\alpha^v)'X^{v-1}X^v$$

où  $W^v(\epsilon)$  est obtenu à partir de  $\underline{u}^v$  comme précédemment

(3)  $\underline{u}^{v+1} = H_{X^v}\underline{1}$

Les propriétés de convergence des deux procédures sont discutées par BROWNE (1985) qui les compare du point de vue de l'efficacité numérique. Cet auteur présente de plus un algorithme de type Newton-Raphson qui semble préférable aux deux premiers. Nous renvoyons cependant à BROWNE (1985) pour sa description, puisque la description d'algorithmes résolutifs sort de notre propos, lorsqu'ils ne mettent pas en lumière certaines caractéristiques d'utilisation de la dualité entre F et  $F\wedge F$ .

L'ensemble des méthodes décrites dans cette section a pour caractéristique commune d'utiliser la dualité entre F et  $F\wedge F$  dans la construction de leurs éléments constitutifs. De plus ces méthodes se sont ainsi avérées toutes distinctes comme il a été montré auparavant par TAKANE (1977). Cet auteur applique de plus les mêmes règles de double écriture, dont nous montrons qu'elle est une application directe de la dualité entre F et  $F\wedge F$ , au cas de l'Analyse en Coordonnées Principales. En effet si, pour tout  $i$ ,  $p_i = \frac{1}{n}$ , on vérifie alors puisque  $Q_1 = T^tT$  :

$$\begin{aligned}
W &= \frac{1}{n^2} T \left( -\frac{1}{2} {}^t \Delta^{*2} T \right) {}^t T \\
&= \frac{1}{n^2} T \left( D_{\Delta^{*2}} + O \left( -\frac{1}{2} {}^t \Delta^{*2} T \right) \right) {}^t T \\
&= \frac{1}{n^2} \left( T D_{\Delta^{*2}} {}^t T + T O \left( -\frac{1}{2} \Psi_{\Delta^{*2}} {}^t T \right) \right)
\end{aligned}$$

où l'on observe le rôle caractéristique joué dans cette méthode, par les termes diagonaux de la matrice des contrastes 2x2 déduite des dissimilarités observées. Sous cette forme on constate la spécificité de la méthode, exprimée dans  $F \wedge F$ , par comparaison avec des formules équivalentes données plus haut pour les autres méthodes. En particulier, par comparaison avec les termes de la matrice  $B(X)$  utilisée par la méthode ELEGANT dans l'équation (104), on observe qu'alors, ce sont les contrastes 2x2 construits à partir des distances reconstruites qui sont pris en compte par la méthode.

#### VI.4.5. Introduction d'une Constante Additive

Les méthodes de Codage Multidimensionnel peuvent être étendues simplement au traitement de dissimilarités mesurées sur une échelle d'intervalles. On introduit alors une constante additive. Nous exposons ici la généralisation du modèle MINRES caractérisée par le modèle

$$\begin{aligned}
\delta_{ij} &= d_{ij}(X) + c (1 - \delta_i^j) + e_{ij} \quad i < j \\
\Leftrightarrow \Delta &= D(X) + c ( \mathbf{1}' \mathbf{1} - I ) + E \\
\Leftrightarrow \underline{\delta} &= \underline{d}(X) - c \underline{u} + \underline{e}
\end{aligned}$$

où

$$\underline{u} = \text{Vec}( \mathbf{1}' \mathbf{1} - I )$$

et on se propose de résoudre le problème suivant qui fournit une solution  $X$  unique à une rotation près des axes de représentation dans l'espace affine  $E$  de dimension  $r$  muni de la métrique  $M$  :

$$(P) \text{ Min } \{ \| \underline{e} \|_{D_Q}^2 ; X \in \mathcal{M}_{r,n}(\mathbb{R}), X D_Q \mathbf{1} = 0, \text{diag}(X D_Q {}^t X) = I, c \in \mathbb{R} \}$$

Montrons qu'alors la solution peut être construite en décomposant le problème, de façon à utiliser une modification élémentaire de l'un des algorithmes exposés plus haut.

**Proposition 10:** *Le minimum de (P) est obtenu comme solution du problème conditionnel*

$$(Pc) \text{ Min } ( \| \hat{\delta} - d(X) - \hat{c}u \|_{D_Q}^2 ; X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_q \mathbf{1} = 0, \text{diag}(XD_q^t X) = I )$$

où  $\hat{c}$  est l'estimateur aux moindres carrés généralisés

$$\hat{c} = \frac{1}{(n-1)q_+} \text{Trace}(P_q(\Delta - D(X)))$$

Preuve : La contrainte d'échelle  $\text{diag}(XD_q^t X) = I$  assure que

$$Q^*(\mathbf{1}\mathbf{1}^t - I)^t X = -^t X$$

Puisque

$$Q^*(\mathbf{1}\mathbf{1}^t - I)^t X = D_q \mathbf{1}\mathbf{1}^t = P_q = (I - Q_q)$$

et de plus si l'on pose

$$d^{ij}(X) = \begin{cases} \frac{1}{d_{ij}(X)} & \text{si } d_{ij}(X) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

le terme général de la matrice  $D^-(X)$ , alors

$$D^-(X) * D(X) = \mathbf{1}\mathbf{1}^t - I$$

Posons enfin  $D_1(X)$  (resp.  $D_2(X)$ ) la matrice diagonale de terme général, l'élément  $(i,j)$  ( $i < j$ ) de la matrice

$$\begin{aligned} Q_1(X) &= D^-(X) * E * Q \\ (\text{resp. } Q_2(X) &= D^-(X) * (\Delta - c(\mathbf{1}\mathbf{1}^t - I)) * Q) \end{aligned}$$

Alors les équations normales permettant la résolution du problème (P) s'écrivent en fonction des matrices

$$\begin{aligned} N_1(X) &= TD_1(X)^t T \\ N_2(X) &= TD_2(X)^t T \end{aligned}$$

En effet, nous pouvons étendre les résultats de KRANE (1978) au cas d'une pondération  $Q_{|x|}$  non triviale en utilisant les règles de calcul différentiel formel (cf. MAGNUS et NEUDECKER (1988)) :

$$\begin{aligned} \frac{\partial}{\partial c} \| e \|_{D_Q}^2 &= - \text{Trace}(P_q E) \\ \frac{\partial}{\partial^t X} \| e \|_{D_Q}^2 &= 2 \text{Vec}(M^t X N_1(X)) = 2 \text{Vec}(M^t X (I + N_2(X))) \end{aligned}$$

Mais on vérifie que

$$\begin{aligned} \text{Trace}(P_q E) &= \text{Trace}(P_q(\Delta - D(X) - c(\mathbf{1}\mathbf{1}^t - I))) \\ &= \text{Trace}(P_q(\Delta - D(X))) - c \text{Trace}(D_q \mathbf{1}\mathbf{1}^t (\mathbf{1}\mathbf{1}^t - I)) \end{aligned}$$

$$= \text{Trace} (P_q(\Delta - D(X))) - (n-1)q_{.+}c$$

où

$$q_{.+} = \sum_{i=1}^n q_{i+}$$

Le minimum conditionnel de  $c$  est donc son estimation aux moindres carrés pondérés

$$\hat{c} = \frac{1}{(n-1)q_{.+}} \text{Trace} (P_q(\Delta - D(X)))$$

de sorte que l'on détermine  $\hat{X}$  et  $W_{\hat{X}}$  en résolvant le problème conditionnel (Pc). ¶

**Remarque :** Puisque  $\| \underline{u} \|_{D_Q}^2 = q_{.+}$ , on vérifie que

$$\| \underline{e} \|_{D_Q}^2 = \| \underline{\delta} - \underline{d}(X) \|_{D_Q}^2 + (\hat{c})^2 q_{.+} - \sum_{i < j} \sum q_{ij} (\delta_{ij} - d_{ij}(X))$$

dont l'expression est particulièrement simple lorsqu'on impose la normalisation globale de la solution

$$\sum_{i < j} \sum q_{ij} \delta_{ij} = \sum_{i < j} \sum q_{ij} d_{ij}(X) \quad (105)$$

Celle-ci est compatible avec la contrainte modifiée dans (P) et (Pc) :

$$\text{diag}(XD_q^t X) = kI$$

où  $k$  est déterminé à partir de (105).

La proposition 10 montre que l'introduction d'une constante additive nécessite l'itération de deux phases

- (1) Estimation de la constante additive  $\hat{c}^{(v+1)}$  pour une configuration donnée  $X^{(v)}$ ,
- (2) Estimation de la configuration  $\hat{X}^{(v+1)}$  pour une valeur donnée  $c^{(v+1)}$  de la constante additive,

jusqu'à convergence. L'initialisation de l'algorithme peut être réalisée en utilisant les solutions du problème de l'Analyse en Coordonnées Principales avec constante additive analogue. L'extension du problème de la constante additive au cas d'une perturbation de la forme

$$\delta_{ij} = d_{ij}(X) + u_i + u_j + e_{ij} \quad u_i, u_j \in \mathbb{R} \quad (106)$$

a été présenté plus haut dans le cadre de l'étude du problème de la constante additive, associé à l'Analyse en Coordonnées Principales. L'optimisation du critère aux moindres carrés

$$\text{Min } (\| \underline{\delta} - \underline{d}_u(X) \|_{D_Q}^2 ; X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_q \underline{1} = 0, \text{diag}(XD_q^t X) = I)$$

ajuste le modèle (106) proposé par SAITO (1983, 1985, 1986) où

$$\underline{d}_u(X) = \text{Veca}(D_u(X))$$

$$D_u(X) = D(X) + \underline{u}^t \underline{1} + \underline{1}^t \underline{u}$$

Clairement  $\underline{u}^t \underline{1} + \underline{1}^t \underline{u}$  est la matrice représentative d'un élément de  $F \vee F$  et donc la résolution de ce problème ne peut être écrit dans  $F \wedge F$ . Cette caractéristique explique la complexité de l'itération proposée par SAITO (1986) pour estimer les paramètres de ce modèle.

Dans le cadre du modèle MINRES 2, l'introduction d'une constante additive correspond à une perturbation de type 2

$$\delta_{ij}^2 = d_{ij}^2(X) + c(1 - \delta_i^j) + e_{ij} \quad (107)$$

et s'intègre donc dans une stratégie de moindres carrés alternés, telle que celle qui est proposée par TAKANE et al. (1977). De plus

$$\frac{\partial}{\partial c} \text{SSTRESS}(X) = -2 \sum_{i < j} \sum q_{ij} (\delta_{ij}^2 - d_{ij}^2(X) - c)$$

et donc le minimum conditionnel de  $c$  est son estimation aux moindres carrés pondérés

$$\hat{c} = \frac{1}{q_+} \sum_{i < j} \sum q_{ij} (\delta_{ij}^2 - d_{ij}^2(X))$$

et est donc particulièrement simple à estimer.  $\hat{c}$  dépend là encore de  $X$  et on impose donc de recourir à une itération de deux phases, analogue à la procédure mise en œuvre pour la méthode MINRES correspondante. Le modèle général

$$\delta_{ij}^2 = d_{ij}^2(X) + u_i^2 + u_j^2 + e_{ij} \quad , \quad u_i, u_j \in \mathbb{R} \quad (108)$$

a été exposé plus haut dans le cadre de l'étude de la constante additive associée à l'Analyse en Coordonnées Principales. L'optimisation du critère aux moindres carrés

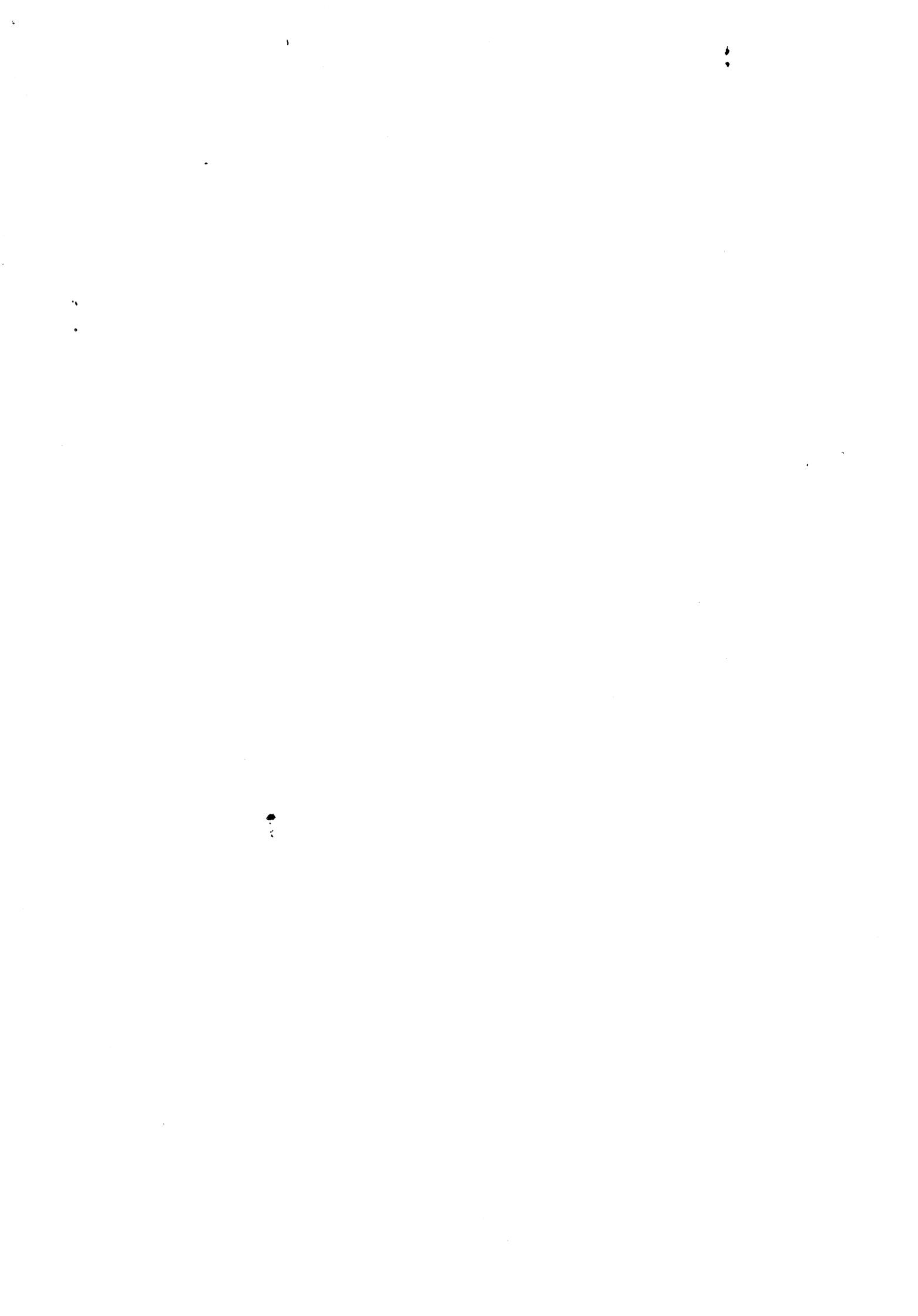
$$\text{Min } (\| \underline{\delta}^{*2} - \underline{d}_u^{*2}(X) \|_{D_Q}^2 ; X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_q \underline{1} = 0, \text{diag}(XD_q^t X) = I)$$

ajuste le modèle (108) proposé par CARROLL (1983) et WINSBERG et CARROLL (1984) où

$$\begin{aligned} \underline{d}_u^{*2}(X) &= \text{Veca}(D_u^{*2}(X)) \\ D_u^{*2}(X) &= D^{*2}(X) + \underline{u}\underline{1} + \underline{1}\underline{u} - 2 D_{\underline{u}} \end{aligned}$$

Ces auteurs proposent un algorithme spécifique de résolution, rendu nécessaire encore ici par le fait que  $\underline{u}\underline{1} + \underline{1}\underline{u} - 2 D_{\underline{u}}$  est la matrice représentative d'un élément de  $F \vee F$  et interdit donc la description du problème dans  $F \wedge F$ .

Ces deux modèles généraux peuvent être interprétés comme des problèmes de Codage Multidimensionnel sous contraintes de structures et sont abordés en ces termes dans d'AUBIGNY (1987).



## CHAPITRE VII

### ANALYSE ORDINALE DU TRIPLE $(I, \delta, Q_{|I|})$

Les méthodes de Codage Multidimensionnel du Triple  $(I, \delta, Q_{|I|})$  sont traitées dans les paragraphes précédents à l'aide d'algorithmes qui utilisent les propriétés numériques de  $\delta$ . On considère alors que  $\delta$  est mesurée sur une échelle de rapport lorsqu'on traite le problème sans constante additive ou d'intervalles lorsqu'on introduit une constante additive.

Nous étudions ici les méthodes de Codage Multidimensionnel Ordinal. On suppose alors que  $\delta$  est mesurée sur une échelle ordinale. Alors on retient l'information portée par la préordonnance  $\mathcal{R}_\delta$  et l'on suppose de plus que  $\mathcal{R}_\delta$  est une préordonnance propre. Lorsqu'est donnée une bijection de  $I$  sur  $[n]$ , on note  $(\mathcal{T}, \mathcal{R}_\delta)$  l'ensemble préordonné formé de l'ensemble  $\mathcal{T} = \{(i,j) \in [n] \times [n]; i < j\}$  et de la relation de préordre  $\mathcal{R}_\delta$  induite sur  $\mathcal{T}$  par  $\delta$ .  $\mathcal{T}$  est de cardinal  $n^* = n(n-1)/2$ , de sorte que l'indice de rang de la préordonnance  $\mathcal{R}_\delta$  s'écrit ici

$$\begin{aligned} r: \mathcal{T} &\rightarrow \mathbb{R} \\ t &\mapsto r(t) = n^* + 1 - \text{card}([t, \rightarrow]) \end{aligned}$$

**Remarque:** Cette définition de  $r$  est différente de celle qui est adoptée au chapitre II, puisqu'on se restreint ici aux éléments de  $\mathcal{T}$ , alors que là on se plaçait dans  $[n] \times [n]$ .

Les méthodes de Codage Multidimensionnel Ordinal se proposent de construire une représentation  $X$  de  $I$  dans un espace affine Euclidien  $E$  de dimension fixée a priori, de sorte que la fonction de distance  $d(X)$  induite par  $X$  soit équivalente à  $\delta$ . Cette propriété est en général exprimée sous la forme

$$(s \mathcal{R}_\delta t) \Rightarrow (d_s(X) \leq d_t(X)) \quad ((s,t) \in \mathcal{T} \times \mathcal{T})$$

Nous distinguons dans ce chapitre deux classes de méthodes, qui diffèrent par la façon dont la relation précédente est prise en compte. Tout d'abord nous associons à  $\delta$  la classe des fonctions de disparités obtenues par transformation de  $\delta$  par une fonction monotone, ou de la statistique de rang associée. Nous parlons alors de méthodes de codage fonctionnel des rangs. Dans une deuxième partie nous traitons le problème à l'aide

d'outils empruntés à l'algèbre des relations. La notion de comparaison par paires joue alors un rôle primordial et nous parlons de méthodes relationnelles.

### VII.1. LES MÉTHODES DE CODAGE FONCTIONNEL DES RANGS

Ecrivons la relation précédente sous la forme équivalente

$$(\delta_{ij} \leq \delta_{kl}) \Rightarrow (d_{ij}(X) \leq d_{kl}(X)) \quad ((i,j),(k,l) \in [n]^2 \times [n]^2 \quad i < j \text{ et } k < l)$$

Cette formulation du problème est en fait imprécise. CRITCHLEY (1985) formalise trois critères permettant de définir huit notions de relation monotone entre  $\delta$  et  $d(X)$ . Posons en effet la définition suivante (cf. CRITCHLEY (1985)):

**Définition 1 :** *On appelle fonction de disparité, toute transformation monotone  $\theta$  d'une dissimilarité propre  $\delta$  définie sur l'ensemble fini  $I$ . On la note*

$$\theta = f_{abc}(\delta) \tag{108}$$

*ou encore sous forme matricielle*

$$\Theta = f_{abc}(\Delta) = (f_{abc}(\delta_{ij}))_{i,j \in [n]}$$

Le caractère fini de  $I$  impose de rompre avec la convention habituelle en Statistique non paramétrique consistant à négliger les ex aequo dans l'élaboration de la théorie. Cette convention repose en effet sur une hypothèse d'absolue continuité de la loi de probabilité des éléments aléatoires étudiés. Les ex aequo décrivent alors un sous-ensemble de mesure nulle du support de la loi. BRADY (1985) montre qu'une telle hypothèse de continuité ne permet pas de décrire complètement les procédures expérimentales de collecte de données de dissimilarité. Nous ne posons pas de telles hypothèses de régularité et par suite les propriétés mathématiques de la fonction monotone  $f_{abc}$  sont particularisées par la valeur de trois paramètres  $a$ ,  $b$  et  $c$  à valeurs dans  $\{0,1\}$ :

- i)  $a = 1$  si les ex aequo sont préservés, c'est à dire si  $(\delta_{ij} = \delta_{kl}) \Rightarrow (\theta_{ij} = \theta_{kl})$   
 $= 0$  sinon
- ii)  $b = 1$  si les inégalités strictes sont préservées, c'est à dire si  $(\delta_{ij} < \delta_{kl}) \Rightarrow (\theta_{ij} < \theta_{kl})$   
 $= 0$  si  $(\delta_{ij} < \delta_{kl}) \Rightarrow (\theta_{ij} \leq \theta_{kl})$
- iii)  $c = 1$  si les conditions i) et ii) s'appliquent à toutes les paires  $(i,j) \ i \leq j$   
 $= 0$  si elles s'appliquent aux seuls termes hors diagonaux  $(i,j)$  avec  $i < j$

L'indice de rang de la préordonnance  $\mathfrak{R}_\delta$  est une fonction de disparité particulière qui vérifie l'équation

$$r = f_{110}(\delta)$$

Ces distinctions ont une longue histoire dans la méthodologie du Codage Multidimensionnel. Ainsi KRUSKAL (1964 a, b) parle d'une approche primaire (resp. secondaire) si  $a = 0$  (resp. 1). Son algorithme s'applique au cas  $b = 0$  alors que celui de Guttman suppose que  $b=1$ . ROSKAM (1969) parle alors respectivement de monotonie forte (resp. faible) si  $b=1$  (resp.  $b=0$ ). Il parle de plus de transformation semi-forte pour  $f_{01c}$  et semi-faible pour  $f_{10c}$ . de LEEUW (1977b) étudie le comportement des algorithmes proposés par KRUSKAL (1964 a, b) pour ajuster une fonction monotone en présence d'ex aequo. CRITCHLEY (1985) remarque que  $\theta$  est complètement définie mathématiquement dans le seul cas où  $a = 1$  et que  $f_{1bc}$  n'est inversible que si  $b = 1$ . Le rôle de la condition iii) est dégagé par CRITCHLEY (1985) qui montre son intérêt pour interpréter de façon nuancée les implications pratiques des théorèmes de HOLMAN (1972) et de LINGOES (1971) présentés au paragraphe §V.2. Nous ne poursuivons cependant pas cette étude ici et nous renvoyons le lecteur intéressé à CRITCHLEY (1985).

Les conditions d'inversibilité de  $f_{abc}$  de CRITCHLEY (1985) doivent de plus être mises en rapport avec les problèmes d'équivalence des deux modèles fonctionnels suivants de liaison ordinale

$$\text{Modèle 1 : } m(\delta_{ij}) = d_{ij}(X) + e_{ij} \quad (109)$$

$$\text{Modèle 2 : } \delta_{ij} = m(d_{ij}(X)) + r_{ij} \quad (110)$$

Le modèle 1 interprète les résidus  $e_{ij}$  comme la résultante de l'ajustement du modèle et des composantes non Euclidiennes de la fonction de disparité  $\theta = m(\delta)$ . Si  $Z$  est un plongement irréductible de  $(I, \theta)$  dans un espace affine Euclidien de dimension  $p$ , on a alors

$$e_{ij} = \tau_{ij} + \left( \sum_{s=r+1}^p (z_i^s - z_j^s)^2 \right)^{1/2} \quad (111)$$

où  $\tau_{ij}$  est la composante non Euclidienne de  $\theta_{ij}$ , et

$$X = [Z]_r, \quad d_{ij}(X) = \left( \sum_{s=1}^r (x_i^s - x_j^s)^2 \right)^{1/2}$$

L'ajustement du modèle 1 conduit alors à minimiser la fonction objectif

$$\Psi(\alpha, \beta, Q)(X) = \sum_{i < j} \sum q_{ij} | [m(\delta_{ij})]^\alpha - [d_{ij}(X)]^\alpha |^\beta \quad (112)$$

La minimisation de la fonction objectif  $\text{STRESS}(m, X) = \Psi(1, 2, Q)(X)$  fournit la formulation originelle de Kruskal lorsque la transformation  $m$  décrit l'ensemble des fonctions isotones sur  $\mathcal{T}$  (cf. infra). Cette optimisation caractérise la méthode MINRES dans le cas ordinal. De façon analogue, la méthode MINRES 2 s'étend au cadre ordinal, et l'on minimise alors sur le même ensemble de fonctions, la fonction objectif  $\text{SSTRESS}(m, X) = \Psi(2, 2, Q)(X)$ .

La falsifiabilité du modèle 2 est de nature différente. En effet, pour  $X$  et  $Z$  définies comme ci-dessus, on écrit

$$\begin{aligned} \delta_{ij} &= m(d_{ij}(X)) + r_{ij} \\ &= m(d_{ij}(Z)) + (m(d_{ij}(X)) - m(d_{ij}(Z))) + r_{ij} \end{aligned} \quad (113)$$

Le terme  $r_{ij}$  intègre alors les erreurs de mesure, ainsi qu'une composante  $\sigma_{ij}$  induite par la recherche d'une représentation résumée

$$\sigma_{ij} = m(d_{ij}(X)) - m(d_{ij}(Z))$$

La fonction objectif minimisée lors de l'ajustement du modèle 2 est alors

$$\Phi(\alpha, \beta, Q)(X) = \sum_{i < j} \sum q_{ij} | \delta_{ij}^\alpha - [m(d_{ij}(X))]^\alpha |^\beta \quad (112\text{bis})$$

La comparaison des équations (110) et (113) est possible lorsque d'une part,  $m$  est suffisamment régulière et d'autre part,  $d_{ij}(X)$  est proche de  $d_{ij}(Z)$ . Alors un développement en série de Taylor au voisinage de  $d_{ij}(Z)$  montre que

$$\sigma_{ij} = m'(d_{ij}(Z))(d_{ij}(X) - d_{ij}(Z)) + v_{ij}$$

où  $m'$  désigne la dérivée de la fonction  $m$  et  $v_{ij}$  s'interprète comme un coefficient de non euclidéanité et joue donc un rôle distinct de  $\tau_{ij}$ . La distinction des modèles 1 et 2 revêt une importance particulière dans l'établissement de tests statistiques en Codage Multidimensionnel. Ceux-ci sont étudiés par d'AUBIGNY (1979) et plus complètement par BRADY (1985). Ces deux auteurs font appel aux outils classiques d'analyse du modèle linéaire sous contraintes d'ordre (cf. BARLOW et al. (1972), ROBERTSON et al. (1988)) et leur étude peut être complétée par une formulation du problème dual Lagrangien, qui permet de montrer les liens étroits existant entre les tests de rapport de vraisemblance et les tests de Wald et ceux des multiplicateurs de Lagrange pour ce problème. On montre ainsi que ces trois familles de tests sont asymptotiquement équivalentes (cf. GOURIEROUX et al. (1982)). L'intérêt de tels tests est cependant

discutable, puisqu'ils reposent sur des hypothèses probabilistes décrivant le processus d'échantillonnage des paires et de randomisation du plan factoriel  $I \times I$  qui sont rarement satisfaites en pratique. Ces difficultés sont partagées par les tests multivariés de Bartlett, Wilks ou Pillai en Analyse en Composantes Principales. On doit donc leur attribuer le statut d'indicateurs consultables par analogie à une situation de référence (ou étalon) induite par une structure statistique bien connue, mais non nécessairement respectée par les données. Une modélisation statistique satisfaisante du problème n'existe pas à notre connaissance. Les travaux de SUPPES et ZINNES (1963), RAMSAY (1969, 1977, 1982), WALKER (1976) et MacKAY et ZINNES (1981) montrent de plus que les outils probabilistes actuellement développés en statistique multivariée sont insuffisants pour permettre une étude satisfaisante, voir aussi HUBERT et SUBKOVIK (1979), LINGOES (1981) et MacCALLUM (1981). C'est pourquoi d'AUBIGNY (1988 a) préfère mettre en œuvre des procédures de tests de permutation, adaptées de HUBERT et al. (1978), pour le traitement des problèmes d'inférence statistique liés à une méthodologie d'Analyse de la Variance, appliquée à l'observation d'une variable réponse  $\delta$  interprétable comme une fonction de dissimilarité, sur un plan d'expérience défini sur  $I \times I$  par un ensemble de facteurs fixes  $\{F_k ; k \in K\}$ .

D'autre part, les modèles 1 et 2 utilisent explicitement les propriétés ordinales de la fonction de dissimilarité observée  $\delta$ , ou de façon équivalente, de  $\mathcal{R}_\delta$  et de son indice de rang. Par suite nous posons la définition suivante, où  $K$  est un sous-ensemble de  $F \wedge F$  décrit par les fonctions monotones sur  $\mathcal{T}$ .

**Définition 2:** *On appelle méthode de Codage Multidimensionnel de Rang toute méthode de Codage Multidimensionnel Ordinal exprimée comme le problème d'optimisation suivant, où les fonctions critères  $\Psi(\alpha, \beta, Q)$  et  $\Phi(\alpha, \beta, Q)$  sont définies en (112) et (112bis)*

$$\text{Min}\{\Psi(\alpha, \beta, Q)(X) ; d(X) \in K\} \quad \text{Min}\{\Phi(\alpha, \beta, Q)(X) ; d(X) \in K\}$$

La formulation du problème en terme d'ajustement du modèle 2 a été peu étudiée dans le passé. Pourtant, dans le cas  $\alpha = 1$  et  $\beta = 2$ , FRIEDMAN et TIBSHIRANI (1984) montrent qu'elle présente un grand intérêt dans l'établissement d'algorithmes de Codage Multidimensionnel Lisse. On impose alors une condition supplémentaire de régularité sur  $m$ , que l'on estime alors à l'aide d'un estimateur à noyau. Cette approche est en particulier replacée dans une perspective de filtrage et validée par de nombreuses expérimentations numériques dans EL HADRI (1988), auquel nous renvoyons le lecteur intéressé.

L'approximation de  $m$  par des fonctions lisses a fait l'objet de nombreux travaux dans le cadre du modèle 1. WINSBERG et RAMSAY (1983) et HEISER (1985b) proposent en

particulier deux classes distinctes de fonctions splines adaptées au problème de Codage Multidimensionnel Ordinal. Ces auteurs montrent en particulier l'apport des méthodes de codage lisse dans une optique de protection de la méthode de Codage Multidimensionnel contre l'apparition de solutions dégénérées, caractérisées par la confusion de points à dissimilarité non nulle. Cependant les propriétés mathématiques des fonctions splines d'ajustement restent mal connues du point de vue théorique en présence de contraintes de monotonie, malgré les travaux récents en Analyse Numérique (cf. MEDINA (1985)) et la structure mathématique de  $K$  est alors incomplètement connue. L'utilisation des fonctions splines dans le cadre du modèle 1 et l'approche de Friedman et Tibshirani fondée sur l'emploi du modèle 2 sont comparées numériquement par EL HADRI (1988).

Nous limitons notre exposé au cadre du modèle 1 sans lissage, afin d'explicitier les structures Euclidiennes mises en œuvre par les méthodes de Codage Multidimensionnel Ordinal. L'utilisation de lisseurs apparaît en effet comme une sophistication technique de la méthode, dont l'intérêt pratique est réel, mais qui ne remet pas en cause les outils algébriques mis en œuvre dans ce mémoire.

### VII.1.1. La méthode MINRES

Considérons le problème posé par l'optimisation de la fonction critère  $\Psi(1, 2, Q)$ . La solution du problème de Codage Multidimensionnel de Rang réalise alors le minimum

$$\text{Min}\{\text{STRESS}(m, X) ; m \in K, X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_{q\perp} = 0, XD_q^t X = kI\}$$

De façon classique, la configuration  $X$  est construite par une méthodologie de moindres carrés alternés. Une telle procédure part d'une configuration initiale  $X^{(0)}$  et alterne jusqu'à convergence deux étapes :

Etape (1) : Pour  $X^{(v)}$  fixée, construire  $\hat{m}^{(v)}$  solution de

$$\text{Min}\{\text{STRESS}(m, X^{(v)}) ; m \in K\} \quad (114)$$

Etape (2) : Pour fixée  $\hat{m}^{(v)}$ , construire  $X^{(v+1)}$  solution de

$$\text{Min}\{\text{STRESS}(\hat{m}^{(v)}, X) ; X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_{q\perp} = 0, XD_q^t X = kI\}$$

où

$$\text{STRESS}(m, X) = \sum_{i < j} \sum q_{ij} (m(\delta_{ij}) - d_{ij}(X))^2$$

On décrit ainsi un patron d'algorithme, dans lequel l'étape (2) est un problème de Codage Multidimensionnel métrique particulier. On construit alors un plongement Euclidien approché du triple  $(I, \theta, Q_{|X|})$  associé à la disparité  $\theta$ .  $(I, \theta)$  est alors un espace quasi-métrique, de sorte que ce problème est traité par la méthode MINRES décrite au chapitre précédent.

Seule l'étape (1) nécessite un exposé complémentaire. Nous discutons ci-dessous deux méthodes développées dans le but d'associer à  $d(X)$  une fonction de disparité  $\theta = f(d(X))$ . La première construit  $\theta$  en résolvant le problème (114) et respecte complètement le patron d'algorithme donné ci-dessus. La seconde méthode n'optimise pas l'équation (114). Elle est cependant souvent utilisée dans les premières itérations de l'algorithme, de façon à protéger celui-ci contre les risques d'obtention de solutions sous-optimales, correspondant à des optima locaux de la fonction objectif.

#### VII.1.1.1. Utilisation d'une procédure de régression isotone

Lorsque  $\mathcal{R}_\delta$  est une préordonnance complète, l'algorithme du Pool Adjacent Violators (KRUSKAL (1964 a, b), BARLOW et al.(1972), ROBERTSON et al. (1988)), fournit la solution  $\hat{m}$ , construite dans la classe des fonctions en escalier. Nous montrons plus bas que la forme de  $\hat{m}$  ne dépend que de  $\mathcal{R}_\delta$  car la substitution de  $\delta$  par une disparité  $\theta$ , par exemple l'indice de rang de  $\mathcal{R}_\delta$ , fournit une même fonction de recodage  $\hat{m}$ .

d'AUBIGNY (1975) décrit le problème de régression isotone posé par (114) dans l'ensemble des fonctions de carré intégrable  $L^2(\Omega, \mathcal{A}, P)$  et interprète  $\delta$  et  $d(X)$  comme les réalisations  $\Delta(\omega)$  et  $D(X)(\omega)$  de deux vecteurs aléatoires de dimension  $n^*$ . Il montre alors que l'algorithme du Pool Adjacent Violators, construit un estimateur  $\hat{M}$ , de l'espérance conditionnelle  $E(D(X) / \mathcal{B}_\Delta)$ , où  $\mathcal{B}_\Delta$  est le  $\sigma$ -treillis inclus dans  $\mathcal{A}$  et induit par  $\Delta$  ou de façon équivalente, par la préordonnance aléatoire complète  $\mathcal{R}_\Delta$  induite par  $\Delta$  sur  $I$ . Nous démontrons alors le théorème (voir aussi CRITCHLEY (1978)):

**Théorème 1:** *L'ensemble des fonctions mesurables par rapport à  $\mathcal{B}_\Delta$  est un convexe fermé  $L^2(\Omega, \mathcal{B}_\Delta, P)$  dans  $L^2(\Omega, \mathcal{A}, P)$  et  $E(D(X) / \mathcal{B}_\Delta)$  est le projecteur orthogonal sur  $L^2(\Omega, \mathcal{B}_\Delta, P)$ .*

On peut donc décrire  $\hat{m}$  comme une estimation de la valeur de l'espérance conditionnelle  $E(D(X) / \Delta)$  au point  $\delta$ .

$$\hat{m} = \hat{E}(D(X) / \Delta = \delta).$$

et l'on pose  $\widehat{M} = \widehat{E}(D(X)/\Delta)$  ou de façon équivalente,  $\widehat{M} = \widehat{E}(D(X)/\mathfrak{B}_\Delta)$ . Cette présentation de la solution n'est pas dépourvue d'arbitraire, puisqu'elle suppose que  $\Delta$  et  $D(X)$  sont des fonctions de carré intégrable. Si elle est séduisante du point de vue mathématique, cette hypothèse est difficile à justifier pour  $\Delta$  en pratique. Cependant, les propriétés de consistance de  $\widehat{M}$  résultent directement des travaux de LEURGANS (1980) dans ce cadre. BRADY (1985) étudie les propriétés de consistance de la solution  $X$  du problème de codage Multidimensionnel dans le cadre des modèles 1 et 2.

La formulation de d'AUBIGNY (1975) place le problème dans le cadre des techniques d'estimation fonctionnelle. On peut alors s'intéresser à la recherche d'estimateurs de  $E(D(X) / \mathfrak{B}_\Delta)$  dans une classe de fonctions plus régulières que les fonctions en escalier. Cette analyse du problème fournit les fondements méthodologiques de la construction d'estimateurs lisses, dans la classe des estimateurs à noyaux, étudiée par FRIEDMAN et TIBSHIRANI (1984) ou bien dans une classe particulière de fonctions splines. Dans chacun des deux cas, l'introduction de contraintes de régularité se traduit par un choix particulier de géométries hilbertiennes dans un sous-espace hilbertien de  $L^2(\Omega, \mathfrak{A}, P)$ . Elle pose donc un problème de cohérence des critères optimisés dans les deux phases du patron d'algorithme décrit plus haut, qui n'a pas été étudié à ce jour.

En pratique,  $I$  et donc  $\mathcal{T}$  est de cardinal fini. On traite donc un problème discrétisé et dans le cas de l'estimateur  $\widehat{M}$ , on résout un problème décrit comme suit.

Le triple  $(I, \delta, Q_{I \times I})$  muni  $F \wedge F$  de la géométrie Euclidienne induite par la matrice  $D_Q$ , et le critère optimisé est alors défini par l'équation (114). De plus, la préordonnance  $\mathfrak{R}_g$  définit un cône convexe fermé dans  $F \wedge F$ ,

$$K = \{ \underline{x} \in F \wedge F \ ; \ 'A \underline{x} \leq 0 \}$$

où  $A$  est une matrice réelle d'ordre  $n^* \times m$ , dont le nombre de colonnes  $m$  dépend de la relation  $\mathfrak{R}_g$ . En particulier  $m = n^* - 1$  lorsque  $\mathfrak{R}_g$  est une relation d'ordre complète et lorsque la base  $\mathcal{F} \wedge \mathcal{F}$  de  $F \wedge F$  est ordonnée de façon à respecter la relation  $\mathfrak{R}_g$ , la matrice  $A$  a pour  $t$ -ième ligne

$$(0, \dots, 0, 1, -1, 0, \dots, 0)$$

avec le 1 en  $t$ -ième position. Posons maintenant la définition

**Définition 3:** Soit  $f$  une fonction réelle définie sur  $\mathcal{T}$ .  $f$  est isotone par rapport à  $\mathfrak{R}$  si et seulement si

$$(t \mathfrak{R} t') \Rightarrow (f(t) \leq f(t')) \quad \forall (t, t') \in \mathcal{T} \times \mathcal{T}$$

On note  $I_\delta(\mathcal{C})$  la classe des fonctions réelles définies sur  $\mathcal{C}$  et isotones par rapport à  $\mathcal{R}_\delta$ . En particulier

**Proposition 1:** *Toute fonction de dissimilarité est équivalente à  $\delta$  si et seulement si elle est représentée dans  $F \wedge F$  par un vecteur appartenant à l'intérieur relatif  $K^\circ$  de  $K$ . En particulier l'indice de rang de  $\mathcal{R}_\delta$  appartient à  $K^\circ$ , qui est donc non vide.*

$d(X)$  est une fonction réelle définie sur  $\mathcal{C}$ , de sorte que la régression isotone de  $d(X)$  par rapport à  $\delta$  associée à la pondération  $Q_{|X|}$ , où  $d(X)$  et  $Q_{|X|}$  sont connues, est l'élément  $\theta(X)$  de  $I_\delta(\mathcal{C})$  qui réalise l'optimum de

$$(P1) \text{ Min } \left\{ \frac{1}{2} \| \underline{d}(X) - \underline{x} \|_{D_Q}^2 = \sum_{i=1}^{n^*} q_i (d_i(X) - x_i)^2 ; m \in I(\mathcal{C}) \quad x_i = m(\delta_i) \right\}$$

Le problème (P1) s'écrit de façon équivalente

$$(P) \alpha = \text{Min } \left\{ \frac{1}{2} \langle \underline{x}, D_Q \underline{x} \rangle - \langle \underline{x}, \underline{b}(X) \rangle + c(X) ; \underline{x} \in K \right\} \quad (115)$$

où  $\underline{d}(X)$  et  $D_Q$  sont connues et donc

$$\underline{b}(X) = D_Q \underline{d}(X)$$

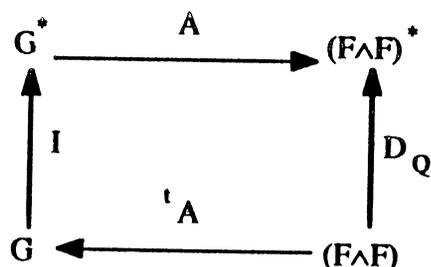
$$c(X) = \langle \underline{d}(X), D_Q \underline{d}(X) \rangle = \sum_{i < j} \sum q_{ij} d_{ij}^2(X)$$

sont des quantités connues.

**Corollaire 1:** *La régression isotone de  $d(X)$  en  $\delta$  ne dépend que de la classe d'équivalence de  $\delta$  et donc que du cône  $K$  représentant  $\mathcal{R}_\delta$  dans  $F \wedge F$ .*

Le problème (115) appartient à la classe des problèmes de programmation quadratique et peut être résolu par des algorithmes classiques en ce domaine. d'AUBIGNY et al. (1988a, b) étudient en détail les propriétés des algorithmes les plus efficaces pour traiter ce problème, dans le cas général d'une relation d'ordre partiel et / ou de contraintes de concavité. Les expérimentations numériques présentées par ces auteurs montrent en particulier que l'algorithme du Pool Adjacent Violators est optimal lorsque  $\mathcal{R}_\delta$  est une relation d'ordre complète sur  $\mathcal{C}$ .

Nous particularisons ici les relations démontrées dans ces articles aux problèmes duaux du problème (P). Considérons en effet l'isométrie de structure Euclidienne induite par  $A$ , décrite par le schéma algébrique suivant



$K$  est un cône polyédrique fermé, inclu dans  $F \wedge F$ , et par application de la dualité de Fenchel (cf. ROCKAFELLAR (1971)), on obtient l'expression du problème dual de (P) dans  $(F \wedge F)^*$  (cf. d'AUBIGNY et al. (1988a, b))

$$(D) \quad \beta = \text{Min} \left\{ \frac{1}{2} \langle \underline{y}, D_Q^{-1} \underline{y} \rangle - \langle D_Q^{-1} \underline{b}(X), \underline{y} \rangle + \frac{1}{2} \langle \underline{b}(X), D_Q^{-1} \underline{b}(X) \rangle - c(X) ; \underline{y} \in K^* \right\}$$

où  $K^*$  est le cône dual de  $K$

$$K^* = \{ \underline{y} \in F^* ; \langle \underline{x}, \underline{y} \rangle \leq 0 \quad \forall \underline{x} \in K \}$$

On montre alors (cf. d'AUBIGNY et al. (1988a, b)) en posant  $G^+$ , l'orthant positif de l'espace vectoriel  $G$  de dimension  $m$ ,

**Proposition 2:**

i) les problèmes (P) et (D) admettent une solution unique  $\underline{x}^*$  et  $\underline{y}^*$  respectivement qui sont reliées par la relation

$$\underline{y}^* = D_Q \underline{x}^* + \underline{b}(X)$$

ii) la solution est caractérisée par

$$(1) \quad {}^t A \underline{x}^* \leq 0$$

$$(2) \quad \text{il existe } \underline{\lambda}^* \in G^+ \text{ tel que } D_Q \underline{x}^* + \underline{b}(X) = -A \underline{\lambda}^* \text{ et } \langle {}^t A \underline{x}^*, \underline{\lambda}^* \rangle \leq 0$$

L'espace vectoriel  $G$  joue un rôle central dans l'établissement d'une dualité lagrangienne entre les estimations  $\underline{x}$  et les contraintes. En effet d'AUBIGNY et al. (1988a, b) notent par application directe des résultats classiques de la programmation convexe (cf. ROCKAFELLAR (1971)) que  $K^*$  s'écrit de façon équivalente

$$K^* = \{ \underline{y} = A \underline{\lambda} ; \underline{\lambda} \in G^+ \}$$

et par suite les solutions de (P) et (D) peuvent être obtenues à partir de la solution du problème dual au sens de la théorie de la dualité lagrangienne (cf. ROCKAFELLAR (1971))

$$(D_L) \quad \gamma = \text{Min} \left\{ \frac{1}{2} \langle \underline{\lambda}, {}^t A D_Q^{-1} A \underline{\lambda} \rangle - \langle {}^t A D_Q^{-1} \underline{b}(X), \underline{\lambda} \rangle ; \underline{\lambda} \in G^+ \right\}$$

d'AUBIGNY et al. (1988a, b) montrent alors le résultat suivant

**Proposition 3:**

i)  $\underline{\lambda}^*$  est solution de  $(D_L)$  si et seulement si  $\underline{y}^* = -A \underline{\lambda}^*$  est solution de  $(D)$

ii)  $\underline{\lambda}^*$  est solution de  $(D_L)$  si et seulement si  $\underline{x}^* = -D_Q^{-1}(A \underline{\lambda}^* + \underline{d}(X))$  est solution de  $(P)$

Du point de vue statistique

$$\underline{y}^* = D_Q(\underline{x}^* - \underline{d}(X)) \quad \text{et} \quad \langle \underline{x}^*, \underline{y}^* \rangle = 0$$

et donc  $\underline{\Gamma}^* = -D_Q^{-1} \underline{y}^*$  est le vecteur des résidus de la régression. Il vérifie

$$\underline{d}(X) = \underline{x}^* + \underline{\Gamma}^* \quad , \quad \langle \underline{x}^*, \underline{\Gamma}^* \rangle_{D_Q} = 0$$

et permet donc de mener une analyse des résidus en Régression Isotone, afin de juger en détail de la qualité de la solution du problème de Codage Multidimensionnel Ordinal et de détecter les observations (i,j) douteuses. D'autre part, les propriétés statistiques de la solution sont analysées plus en détail en remarquant que la relation

$$\underline{y}^* = D_Q(\underline{x}^* - \underline{d}(X)) = A \underline{\lambda}^*$$

implique que

$$\begin{aligned} 0 \leq \lambda_t^* &= \sum_{s=1}^t y_s^* \quad t = 1, 2, \dots, n^*-1 \\ &= \sum_{s=1}^t q_s (x_s^* - d_s(X)) = X_t^* - D_t(X) \quad t = 1, 2, \dots, n^*-1 \end{aligned}$$

où

$$X_t^* = \sum_{s=1}^t q_s x_s^* \quad D_t(X) = \sum_{s=1}^t q_s d_s(X) \quad Q_t = \sum_{s=1}^t q_s$$

définissent les t-ième sommes cumulées associés à  $\underline{x}^*$ ,  $\underline{d}(X)$  et à la matrice diagonale  $D_Q$ . On retrouve ainsi dans le cas d'une relation d'ordre totale, l'interprétation de l'Algorithme du Pool Adjacent Violators en termes de plus grand minorant convexe exposé par exemple par BARLOW et al. (1972). De plus la relation (115) caractérise la solution de la Régression Isotone en fonction de l'ordre de majoration entre  $\underline{x}^*$  et  $\underline{d}(X)$ , au sens de MARSHALL et OLKIN (1979). La Régression Isotone est donc liée par dualité lagrangienne à l'une des notions les plus profondes de la méthodologie statistique, l'ordre de majoration. Elle est appliquée ici avec une contrainte de positivité de la première

coordonnée dans  $G$ , liée à la contrainte de positivité imposée aux fonctions de dissimilarité.

**Remarque :** La formulation de la dualité lagrangienne donnée ici repose sur la définition d'une matrice  $A$  d'ordre  $(n^* - 1) \times n^*$  que l'on peut compléter par l'ajout d'une première ligne égale à  $(1, 0, \dots, 0)$ . On vérifie en effet de façon élémentaire que ce vecteur est linéairement indépendant des vecteurs lignes de  $A$ . On note alors  $\hat{A}$  la matrice ainsi obtenue, et  $\hat{G}$  l'espace vectoriel  ${}^t\hat{A}(F \wedge F)$ . La matrice  $\hat{A}$  est inversible et  $\hat{A}^{-1}$  est la matrice triangulaire inférieure de  $t$ -ème ligne

$${}^t\hat{a}_t = (1, 1, \dots, 1, 0, \dots, 0) = ({}^t\mathbf{1}_t, {}^t\mathbf{0}_{n^*-t})$$

On obtient ainsi

$$\lambda^* = \hat{A}^{-1} y^*$$

On peut alors identifier  $\hat{G}$  et  $F \wedge F$  et les supposer rapportés la même base. Alors  $\hat{a}_t$  décrit l'indicatrice de l'idéal principal  $[\leftarrow, t]$  pour l'ordre  $\mathcal{R}_g$ . Toute fonction de  $I_g(\mathcal{C})$  s'écrit dans le base  $\mathcal{F} \wedge \mathcal{F}$

$$\underline{m} = \hat{A}^{-1} \underline{b}, \quad b_t > 0 \quad \forall t = 2, \dots, n^*$$

On retrouve ainsi la formulation matricielle classique du codage d'une variable ordinale, exposée par exemple dans TAKANE et al. (1976) ou TENENHAUS (1983). Ainsi, tout vecteur  $\underline{b}$  de  $\hat{G}$  qui vérifie la condition de positivité des composantes d'ordre  $t$  supérieur ou égal à 2, associe à  $\delta$  une disparité  $\theta$ , représentée par le recodage ordinal  $\underline{m} = \hat{A}^{-1} \underline{b}$  de  $\underline{b}$ . Comme dit plus haut, la spécificité du problème de Codage Multidimensionnel impose de plus la contrainte  $b_1$  positif. TENENHAUS (1983) utilise cette formulation de l'algorithme du Pool Adjacent Violators dans  ${}^t\hat{A}^{-1}(F \wedge F)^*$ , pour construire dans le cas d'une préordonnance complète, un algorithme dual de Regression Isotone, plus efficace en nombre d'opérations que l'algorithme primal programmé par Kruskal. On vérifie de plus par simple identification des termes que TENENHAUS (1982) décrit une méthode de résolution du problème  $(D_L)$ .

#### VII.1.1.2. Utilisation de la méthode des images de rangs

La Régression Isotone fournit une solution de l'étape (1) du patron d'algorithme présenté ci-dessus, en termes de moindres carrés homogènes à la résolution du problème de Codage Multidimensionnel Ordinal aux moindres carrés. GUTTMAN (1968) propose une solution alternative, en termes d'images de rangs et YOUNG (1975) montre que ces deux solutions fournissent deux formes extrêmes d'un continuum de transformations possibles

des données décrites par YOUNG (1975) en termes de filtres de moyennes mobiles particulières. L'exemple de l'algorithme proposé par YATES (1972) sous le nom de "Balanced Least Squares Monotone Regression" montre cependant que les situations intermédiaires entre régression isotone et images de rangs ne conduisent pas forcément à une transformation monotone des données.

Il est souvent opposé à la méthode d'images de rangs l'absence d'un critère (par exemple de moindres carrés) d'optimalité de la solution. De plus GUTTMAN (1968) décrit la méthode en termes de permutations définies sur  $[n^*]$ . Cette présentation combinatoire des images de rangs s'oppose au cadre fonctionnel utilisé dans la section précédente pour décrire la méthode de Regression Isotone. Aussi nous paraît il utile de donner une interprétation de la méthode dans le formalisme de l'Analyse Canonique linéaire et susceptible de fournir une caractérisation de la méthode complémentaire aux résultats établis par GUTTMAN (1968) et YOUNG (1975).

Etant donnés  $\underline{\delta}$  et  $\underline{d}(X)$ , Guttman construit le vecteur des images de rangs de  $\underline{d}(X)$ , noté  $\underline{d}^*(X)$ , à l'aide d'une permutation  $\sigma$  définie sur  $[n^*]$ . Posons donc la définition suivante, exprimée sans perte de généralité dans  $F \wedge F$ .

**Définition 4:** Etant donnés deux vecteurs  $\underline{\delta}$  et  $\underline{d}(X)$  d'un espace vectoriel  $F \wedge F$ . Le vecteur  $\underline{d}^*(X)$  est l'image de rangs de  $\underline{\delta}$  construite à partir de  $\underline{d}(X)$  si et seulement si

$$d_t^*(X) = d_{\sigma(t)}(X) \quad \forall t, t \in \mathcal{T}$$

où la permutation  $\sigma$  est définie sur  $[n^*]$  par l'une des deux expressions équivalentes

$$(\tau \mathcal{R}_\delta s) \Leftrightarrow (d_t^*(X) < d_s^*(X)) \quad \forall s, t \in [n^*]$$

$$(\delta_t < \delta_s) \Leftrightarrow (d_t^*(X) < d_s^*(X)) \quad \forall s, t \in [n^*]$$

La permutation  $\sigma$  est clairement complètement déterminée par  $\mathcal{R}_\delta$ . Par suite on peut énoncer le résultat suivant:

**Proposition 4:** L'image de rangs de  $\delta$  construite à partir de  $d(X)$  ne dépend que de la classe d'équivalence de  $\delta$  et donc que du cône  $K$  représentant  $\mathcal{R}_\delta$  dans  $F \wedge F$ . C'est dire que toutes les disparités appartenant à  $K^0$  ont la même image de rangs construite à partir de  $d(X)$ .

D'autre part, la présentation de Guttman admet la possibilité d'ex aequo. Cette situation est étudiée après une description de la méthode dans le cas sans ex aequo.

Associons maintenant à  $\underline{\delta}$  le tableau  $Z_{\delta} \in \mathcal{M}_{n^+, n^+}(\mathbb{R})$  de  $r$ -ième ligne ( $r \in [n^+]$ ) le vecteur représentatif de l'indicatrice  $z^r$  définie par

$$z^r(x_t) = \begin{cases} 1 & \text{si } \text{rang}(x_t) = r \\ 0 & \text{sinon} \end{cases} \quad \forall \underline{x} \in F \wedge F$$

On associe de la même façon le tableau  $Z_{d(X)} \in \mathcal{M}_{n^+, n^+}(\mathbb{R})$  à  $\underline{d}(X)$ . L'Analyse Canonique du triplet  $(Z_{\delta}, Z_{d(X)}, I_{n^+})$  est décrite par le schéma de dualité algébrique

$$\begin{array}{ccccc} E_{\delta} & \xleftarrow{Z_{\delta}} & (F \wedge F)^* & \xrightarrow{Z_{d(X)}} & E_X \\ M_{\delta} \downarrow & & \uparrow I & & \downarrow M_X \\ E_{\delta}^* & \xrightarrow{{}^t Z_{\delta}} & (F \wedge F) & \xleftarrow{{}^t Z_{d(X)}} & E_X^* \end{array}$$

où  $E_{\delta}$  et  $E_X$  sont des espaces vectoriels isomorphes à  $F \wedge F$ . Sa résolution en termes de  $\Pi$ -stratégie (cf CAILLEZ et PAGES (1976)) est décrite par le triplet  $(B, M_{\delta}, M_X)$  où

$$M_X = V_X^{-1} = (Z_{d(X)} {}^t Z_{d(X)})^{-1}$$

$M_{\delta}$  est définie de façon analogue. Ces deux matrices sont l'identité en l'absence d'ex aequo.

$$B_{\delta} = Z_{\delta} {}^t Z_{d(X)}$$

et puisque  $M_X$  est la matrice identité,  $B$  est l'opérateur de régression linéaire et les résultats classiques de l'Analyse Canonique montrent que  $\underline{d}^*(X)$  est l'image de  $\underline{d}(X)$  par  $B$ .  $\underline{d}^*(X)$  s'interprète en termes de projection sur  $E_{\delta}$ . On vérifie alors par identification terme à terme le résultat simple suivant

**Proposition 5:**  $B$  est la matrice de la permutation  $\sigma$  dans la base  $\mathcal{F}^* \wedge \mathcal{F}^*$ .

En présence d'ex aequo,  ${}^t Z_{d(X)}$  (resp.  ${}^t Z_{\delta}$ ) est la matrice des indicatrices des rangs associées aux valeurs distinctes de  $\underline{d}(X)$  (resp. de  $\underline{\delta}$ ). Par suite, le nombre de lignes de l'application  ${}^t Z_{d(X)}$  (resp.  ${}^t Z_{\delta}$ ) est fixé par le nombre de valeurs distinctes de  $\underline{d}(X)$  (resp. de  $\underline{\delta}$ ). Par suite  $M_X$  (resp.  $M_{\delta}$ ) n'est plus égale à la matrice identité, et  $B$  n'est plus une matrice de permutation. La démarche demeure cependant fondée en termes d'Analyse Canonique et l'image d'un vecteur quelconque  $\underline{x}$  de  $(F \wedge F)^*$  par  $M_{\delta} Z_{\delta}$  code les ex aequo pour  $\delta$  par la moyenne des valeurs correspondantes de  $\underline{d}(X)$ . Enfin, le cas des données manquantes est pris en compte en munissant  $|x|$  d'une pondération appropriée, c'est à dire nulle en  $(i, j)$  si  $\delta_{ij}$  est une donnée manquante.

### VII.1.2. La méthode MINRES 2

La méthode MINRES 2 pose le problème du Codage Multidimensionnel Ordinal en termes d'optimisation de la fonction critère  $\psi(2, 2, Q)$ . La solution du problème de Codage Multidimensionnel de Rang réalise alors le minimum

$$\text{Min}\{\text{SSTRESS}(m, X) ; m \in K, X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_q \mathbf{1} = \mathbf{0}, XD_q^t X = kI\}$$

De façon classique, la configuration  $X$  est construite par une méthodologie de moindres carrés alternés. Une telle procédure part d'une configuration initiale  $X^{(0)}$  et alterne jusqu'à convergence deux étapes :

Etape (1) : Pour  $X^{(v)}$  fixée, construire  $\hat{m}^{(v)}$  solution de

$$\text{Min}\{\text{SSTRESS}(m, X^{(v)}) ; m \in K\} \quad (117)$$

Etape (2) : Pour fixée  $\hat{m}^{(v)}$ , construire  $X^{(v+1)}$  solution de

$$\text{Min}\{\text{SSTRESS}(\hat{m}^{(v)}, X) ; X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_q \mathbf{1} = \mathbf{0}, XD_q^t X = kI\}$$

où

$$\text{SSTRESS}(m, X) = \sum_{i < j} \sum q_{ij} (m(\delta_{ij}^2) - d_{ij}^2(X))^2$$

De plus l'optimisation des fonctions STRESS et SSTRESS conduit à un problème identique dans la phase (1) du patron d'algorithme, puisque  $\delta$  est positive et donc  $\delta$  et  $d(X)$  sont liées par une relation monotone si et seulement si  $\delta^{*2}$  et  $d^{*2}(X)$  le sont. La discussion des méthodes de traitement de la phase (1), présentée dans le paragraphe précédent, s'applique donc ici sans changements.

### VII.2. LES MÉTHODES RELATIONNELLES

L'étude qui précède repose sur la construction d'une transformation monotone des données

$$\theta = m(\delta) \quad m \in I(\mathcal{T})$$

qui fournit un recodage de la variable  $\delta$ , définie sur  $I$  et permet d'approcher au mieux, au sens des moindres carrés pondérés, la fonction de dissimilarité recodée, ou disparité  $\theta$ , par une dissimilarité Euclidienne  $d(X)$  induite par la représentation  $X$  de  $I$  dans un espace affine Euclidien  $E$  de faible dimension. Ce mode de description d'une relation de préordonnance  $\mathcal{R}_\delta$  s'inscrit dans la tradition fonctionnelle des méthodes de la statistique

de rang (cf. LEHMANN (1975)) et impose des contraintes d'échelle de mesure aux données. On sait en effet (cf. KENDALL (1962), GUILBAUD (1980)) que l'élaboration de telles techniques repose explicitement sur la comparaison de triplets (s,t,u) d'éléments de  $\mathcal{T}$ .

De façon classique, on oppose cette approche aux méthodes de comparaisons par paires (cf. DAVID (1988)), pour lesquelles, seule l'information portée par les paires (s,t) d'éléments de  $\mathcal{T}$ , c'est à dire les dyades (ROSKAM (1969)), est prise en compte. Ce sont alors les propriétés de la relation  $\mathcal{R}_\delta$  qui sont prises en compte par l'analyse, et nous parlons à ce propos de méthode relationnelle.

On se propose ici de traiter le problème de Codage Multidimensionnel Ordinal dans le cadre des techniques d'analyse de données de comparaisons par paires. Deux raisons motivent notre intérêt pour cette formulation du problème, malgré les difficultés informatiques imposées par la taille des corpus de données traités dans ce cas.

Tout d'abord, le formalisme mis en place permet de traiter le cas du Codage Multidimensionnel Ordinal Multiple caractérisé par la donnée d'un échantillon de mesures de dissimilarité observées sur  $I \times I$ . On dispose alors d'un ensemble de matrices de dissimilarité

$$(\Delta^k, k = 1, \dots, K) \quad \Delta^k = \Delta(\omega_k)$$

ou, de façon équivalente, d'un échantillon de préordonnances définies sur  $I$

$$(\mathcal{R}^k, k = 1, \dots, K) \quad \mathcal{R}^k = \mathcal{R}(\omega_k)$$

où  $\Delta$  (resp.  $\mathcal{R}$ ) s'interprète comme une matrice (resp. relation binaire) aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Ce problème est à la base des travaux de RAMSAY (1969, 1977, 1978, 1980) et TAKANE (1977 a, 1978). On est alors naturellement amené à utiliser les outils de la Statistique classique en analyse des comparaisons par paires lorsqu'on se propose de construire des indicateurs résumés à partir de  $(\Delta^k, k = 1, \dots, K)$  (resp.  $(\mathcal{R}^k, k = 1, \dots, K)$ ). L'estimation d'indices de concordance (DEGENNE (1972), HOLLANDER et SETHURAMAN (1973), LEE et SCHUCANY (1975), FEIGIN et COHEN (1978), GORDON (1979), ALVO et al. (1982), PETTITT (1982), FEIGIN et ALVO (1986)) fournit un exemple d'outils statistiques mis en œuvre par d'AUBIGNY (1987) pour résoudre des problèmes d'Analyse des Données, et en particulier de Codage Multidimensionnel dans ce cadre.

D'autre part cette formulation du problème permet de montrer les relations existant entre les techniques de Codage Multidimensionnel Ordinal, la Théorie des Votes et l'Analyse des Données de Préférence. Ces méthodes, en général développées indépendamment de

préoccupations de Statistique Inférentielle (cf. d'AUBIGNY (1981)), offrent alors un ensemble d'outils d'analyse, susceptible d'enrichir la méthodologie d'analyse ordinaire d'un échantillon  $\{\Delta^k, k = 1, \dots, K\}$ .

L'application des méthodes de comparaison par paires en Codage Multidimensionnel Ordinal a fait l'objet d'un certain nombre de publications par le passé. Tout d'abord, l'interprétation des travaux de van EEDEN (1956, 1957a, b, 1970) dans le cadre d'étude présent, montre que sa généralisation de l'algorithme du Pool Adjacent Violators au cas de relations d'ordre partiel, utilise la dualité lagrangienne explicitée entre  $F \wedge F$  et  $G$ . L'application de l'algorithme de régression proposé par JOHNSON (1975) fournit une alternative à la démarche suivie par Van Eeden. Ces deux propositions d'algorithmes de régression isotone conduisent à un algorithme mixte de Codage Multidimensionnel Ordinal, dans lequel la phase (1) du patron d'algorithme modèle est résolue en termes de comparaisons par paires et la phase (2) est résolue en termes fonctionnels dans  $F \wedge F$  ou mieux dans  $F_1$ .

Les méthodes de Codage Multidimensionnel développées par GUTTMAN (1969), de LEEUW (1970), ROSKAM (1970), JOHNSON (1973) et KAMENSKI (1977) résolvent les deux étapes du patron d'algorithme en termes de comparaisons par paires. La relation entre la classe d'algorithmes présentée au paragraphe précédent et les méthodes de comparaisons par paires est analysée en détail par YOUNG (1975). De plus la construction d'une approximation de la dissimilarité  $\delta$  au sens de la norme  $L_1$ , traitée par de LEEUW (1970) comme cas particulier de sa méthode dite de l'Orthant Positif est aussi étudiée par MÖBUS (1976).

### VII.2.1. Analyse Ordinale du Triple $(I, \delta, Q_{I \times I})$

Soient  $\mathcal{R}_\delta$  et  $\mathcal{R}_X$  les deux relations de préordre définies sur  $\mathcal{T}$  par

$$\begin{aligned} (s \mathcal{R}_\delta t) &\Leftrightarrow (\delta_s \leq \delta_t) \\ (s \mathcal{R}_X t) &\Leftrightarrow (d_s(X) \leq d_t(X)) \end{aligned}$$

Notons de plus  $\mathcal{R}_X \Delta \mathcal{R}_\delta$  la relation définie sur  $\mathcal{T}$  par la différence symétrique de  $\mathcal{R}_X$  et  $\mathcal{R}_\delta$ . L'indicatrice de la différence symétrique  $\mathcal{R}_X \Delta \mathcal{R}_\delta$  prend la valeur 0 si  $s$  et  $t$  sont classés dans le même ordre par  $\delta$  et  $d(X)$ , et 1 sinon. Elle définit une pondération en  $(0,1)$  sur  $\mathcal{T}$  et permet de considérer la classe d'indices d'association suivante associée à tout couple  $(\alpha, \beta)$  de nombre réels strictement positifs

$$\Phi(\alpha, \beta, Q(X)) = \sum_{s < t} q_{st}(X) |d_s^\alpha(X) - d_t^\alpha(X)|^\beta \quad (118)$$

que l'on peut normaliser en remarquant que si  $\bar{Q}(X)$  est la pondération  $(1 - Q(X))$ , alors  $\bar{Q}(X)$  est associée à l'indicatrice de la relation  $\mathcal{R}_X \cap \mathcal{R}_\delta$  lorsque  $\mathcal{R}_X$  et  $\mathcal{R}_\delta$  sont complètes. On obtient alors la décomposition additive

$$\sum_{s < t} \sum |d_s^\alpha(X) - d_t^\alpha(X)|^\beta = \Phi(\alpha, \beta, Q(X)) + \Phi(\alpha, \beta, \bar{Q}(X))$$

Choisissons alors l'indice d'association

$$F(\alpha, \beta, Q(X)) = \frac{\Phi(\alpha, \beta, Q(X))}{\Phi(\alpha, \beta, Q(X)) + \Phi(\alpha, \beta, \bar{Q}(X))}$$

c'est à dire

$$F(\alpha, \beta, Q(X)) = \frac{\sum_{s < t} \sum q_{st}(X) |d_s^\alpha(X) - d_t^\alpha(X)|^\beta}{\sum_{s < t} \sum |d_s^\alpha(X) - d_t^\alpha(X)|^\beta} \quad (119)$$

Lorsque  $\alpha = \beta = 1$ , on obtient un indice équivalent à "l'indice de valeur absolue" proposé par GUTTMAN (1969). Lorsque  $\alpha = \beta = 2$ , on retrouve la fonction objectif optimisée par JOHNSON (1973). De plus, lorsque  $Q$  est la pondération triviale, de LEEUW (1970) note que l'on obtient dans le cas  $\alpha = 1$

$$\lim_{\beta \rightarrow 0} F(\alpha, \beta, Q(X)) = \frac{m_-}{m_- + m_+}$$

c'est à dire le rapport du nombre  $m_-$  de paires classées de façon distincte par  $\delta$  et  $d(X)$  au nombre de comparaisons par paires collectées ( $m_- + m_+$ ). On reconnaît ainsi la statistique Tau-C de KENDALL (1962). de LEEUW (1970) étudie principalement le cas  $\alpha = 1$ . Cependant, les conditions d'euclidéanité d'un espace quasi-métrique  $(I, \delta^\alpha)$ ,  $\alpha < 1$  déduite des propriétés de  $(I, \delta)$  (cf. Chapitre II) conduit naturellement à étudier le cas général, tout en remarquant que les difficultés algorithmiques sont principalement liées aux valeurs de  $\beta$  (cf. de LEEUW (1970)). Cet auteur remarque de plus que

$$F(\alpha, \beta, Q(X)) = \frac{1}{2^\beta} \frac{\sum_{s < t} \sum [ |d_s^\alpha(X) - d_t^\alpha(X)| - (d_s^\alpha(X) - d_t^\alpha(X)) ]^\beta}{\sum_{s < t} \sum |d_s^\alpha(X) - d_t^\alpha(X)|^\beta}$$

de sorte que lorsque  $d(X)$  est connue, toute méthode d'optimisation de la fonction critère  $F(\alpha, \beta, Q(X))$  réalise une projection du vecteur  $y(X)$  de coordonnées

$$y_{st}(X) = d_s^\alpha(X) - d_t^\alpha(X)$$

sur l'orthant positif de l'espace vectoriel décrit par  $y(X)$ . Elle conduit donc à des calculs simples, puisque cette projection  $\hat{y}(X)$  a pour coordonnées

$$\hat{y}_{st}(X) = [y_{st}(X)]^+ \text{ où } x^+ = \text{Max}(x, 0)$$

Nous renvoyons le lecteur intéressé à de LEEUW (1970), JOHNSON (1973) et KAMENSKI (1977) pour un exposé des algorithmes résolutifs permettant de construire la solution du problème de Codage Multidimensionnel Ordinal. Notre exposé est en effet maintenant consacré à la mise en place des structures algébriques utiles à l'étude du triple  $(I, \{\delta^k, k \in [K]\}, Q_{|X|})$ .

### VII.2.2. Analyse Ordinale du Triple $(I, \{\delta^k, k \in [K]\}, Q_{|X|})$

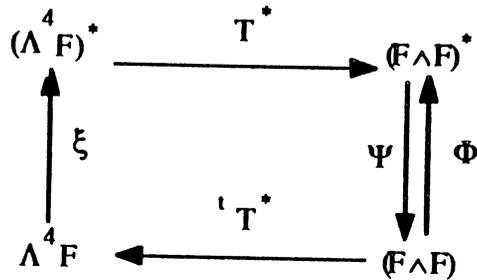
Nous montrons ici que le formalisme développé au chapitre précédent peut être étendu de façon simple afin de permettre la formalisation des méthodes d'Analyse Ordinale du Triple  $(I, \{\delta^k, k \in [K]\}, Q_{|X|})$ , où nous supposons que les fonctions de dissimilarité  $\delta^k$  (resp. les préordonnances associées  $\mathcal{R}^k$ ) sont propres. Nous regroupons les méthodes d'analyse qui ont pour but de construire une représentation approchée de  $I$ , notée  $X$ , dans un espace affine Euclidien de faible dimension, sous le vocable de Codage Multidimensionnel Ordinal Multiple. Ce problème a été très peu étudié par le passé, dans un cadre ordinal, sans doute du fait de l'importance des problèmes numériques rencontrés, mais aussi de l'absence d'un cadre théorique d'étude de ce problème. Nous définissons ici un tel cadre, et présentons brièvement le principe d'une méthode de résolution.

Notre présentation des outils algébriques nécessaires à la résolution du problème s'appuie sur les travaux de GUILBAUD (1980) que nous situons dans un cadre de dualité algébrique. Nous montrons en particulier la parenté des problèmes traités par les méthodes de Codage Multidimensionnel Ordinal Multiple avec les outils d'analyse des procédures de votes. L'originalité du Codage Multidimensionnel Ordinal provient alors de la structure de l'espace  $F \wedge F$  décrit par les observations  $\delta^k$ .

Considérons donc  $F \wedge F$  muni de la base  $\mathcal{F} \wedge \mathcal{F}$  et de façon analogue à la construction présentée au chapitres II et III, associons à  $F \wedge F$ , l'espace vectoriel  $\Lambda^4 F = (F \wedge F) \wedge (F \wedge F)$ , que nous munissons de la base  $\Lambda^4 \mathcal{F} = (\mathcal{F} \wedge \mathcal{F}) \wedge (\mathcal{F} \wedge \mathcal{F})$ . Notons de plus  $(\Lambda^4 F)^*$  le dual algébrique de  $(\Lambda^4 F)$  et munissons cet espace vectoriel de la base  $\Lambda^4 \mathcal{F}^*$ . Alors, puisque  $F \wedge F$  est de dimension  $n^*$ ,  $\Lambda^4 F$  est de dimension

$$n^{**} = \frac{n^*(n^* - 1)}{2}$$

Lorsqu'on fait choix d'une bijection de  $\mathcal{T}$  sur  $[n^*]$ ,  $\Lambda^4 F$  est l'espace de représentation des dyades, c'est à dire des paires  $(s,t)$  de  $\mathcal{T} \times \mathcal{T}$  telles que  $s < t$ . Nous construisons de plus l'application linéaire  $t^*$  à valeurs dans  $(F \wedge F)^*$  et définie sur  $(\Lambda^4 F)^*$  par la donnée de la matrice de simples contrastes  $T^*$  définie de façon analogue à la matrice  $T$  associée à l'élément  $t$  de  $\mathcal{L}((F \wedge F)^*, F^*)$ . L'application linéaire ainsi définie  $t^*$  met donc en dualité l'espace de représentation des dyades et l'espace de représentation des paires, de la même façon que  $t$  met en dualité l'espace de représentation des paires et l'espace de représentation des unités statistiques, et s'interprète donc aussi en termes de théorie des graphes. On construit alors le schéma de dualité



$t^*$  associe à tout vecteur  $y$  de  $F \wedge F$  le vecteur  $z$  de  $(s,t)$ -ième coordonnées égale à  $(y_s - y_t)$ . En particulier

$$t^*(d^{*\alpha}(X)) = (d_s^\alpha(X) - d_t^\alpha(X))_{s < t}$$

et donc la matrice des poids définie sur  $\mathcal{T} \times \mathcal{T}$  au paragraphe précédent induit une géométrie Euclidienne sur  $\Lambda^4 F$  de matrice  $D_{Q(X)}$  la matrice diagonale des poids  $q_{st}(X)$ , de sorte que (119) devient

$$F(\alpha, \beta, Q(X)) = \frac{\|t^*(d^{*\alpha}(X))\|_{D_{Q(X)}}^2}{\|t^*(d^{*\alpha}(X))\|_I^2} \tag{120}$$

Par suite, la contrainte de monotonie de la relation liant  $\delta^*$  à  $d^*(X)$  se traduit dans  $\Lambda^4 F$  par le choix d'une géométrie Euclidienne particulière dépendant de  $X$ , puisque  $q_{st}(X)$  est une fonction de  $X$  pour tous  $s$  et  $t$  dans  $[n^*]$ . La prise en compte de la pondération  $Q_{I \times I}$  fournit une généralisation immédiate de cette expression puisque nous avons montré lors de l'analyse du triple  $(I, \delta, P_I)$  que toute pondération  $P_I$  induit sur  $F \wedge F$  une matrice diagonale  $D_\varphi$  de poids  $\varphi_{ij} = p_i p_j$ . De façon analogue, nous considérons ici la matrice diagonale des poids  $D_\varphi$  définie sur  $\Lambda^4 F$  par

$$D_\varphi = \text{diag}(q_s, q_t)$$

Alors si  $D_{\varphi(X)}$  est la matrice diagonale des poids  $\varphi_{st}(X) = q_{st}(X)q_s q_t$ , on obtient la formule générale

$$F(\alpha, \beta, Q(X), \varphi) = \frac{\|T^*(\underline{d}^{*\alpha}(X))\|_{D_{\varphi(X)}}^2}{\|T^*(\underline{d}^{*\alpha}(X))\|_{D_{\varphi}}^2} \quad (121)$$

Dans un article consacré à la comparaison des indices empiriques de corrélation de Spearman et de Kendall, GUILBAUD (1980) décrit les procédures de vote en termes de permutation de l'ensemble des candidats. Par analogie, interprétons ici les paires d'unités statistiques distinctes, c'est à dire les éléments de  $\mathcal{T}$  comme des candidats et notons  $\Omega$  l'ensemble des permutations de  $[n^*]$ . On peut décrire  $\Omega$  en termes linéaires en notant  $H$  l'espace vectoriel libre de dimension  $h = (n^*)!$ , rapporté à la base des  $h$  permutations de  $\Omega$ . Alors la préordonnance  $\mathfrak{R}^k$  associée à chaque tableau observé  $\Delta^k$  est représentée par un vecteur de  $H$ . Clairement la dimension de  $H$  interdit toute analyse concrète dans cet espace vectoriel. On travaille donc de préférence dans  $F \wedge F$  ou dans  $\Lambda^4 F$ .

En théorie des votes, le décompte des voix définit une mesure de comptage sur l'ensemble des permutations des candidats. Comme il est classique en analyse linéaire des données, associons à cette mesure de comptage la métrique diagonale des poids  $D_{\underline{q}}$ .  $D_{\underline{q}}$  définit une forme quadratique sur  $H$  et nous notons

$$\underline{q} = D_{\underline{q}} \underline{1}_H$$

$\underline{q}$  est un vecteur de  $H^*$  et est interprété dans la théorie des votes comme une opinion, de sorte que l'on associe à tout vote (cf. d'AUBIGNY (1981))

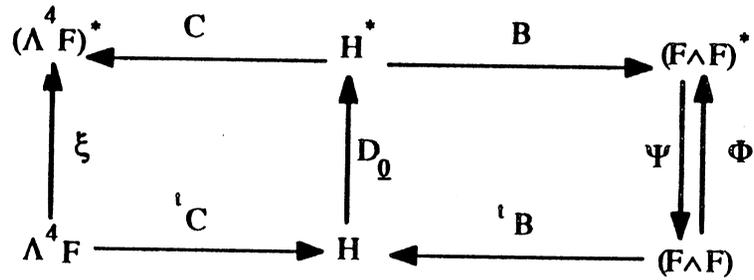
- i) le vecteur  $\underline{b}$  de  $(F \wedge F)^*$  décrivant le dépouillement du vote en termes de rangs de chaque candidat, c'est à dire ici de chaque paire  $t = (i, j)$  de  $\mathcal{T}$ .
- ii) le vecteur  $\underline{c}$  de  $(\Lambda^4 F)^*$  décrivant le dépouillement du vote en termes d'utilité affectée à chaque dyade  $(s, t)$  d'éléments de  $\mathcal{T}$  avec  $s < t$ .

L'explicitation des espaces décrits par ces vecteurs n'est certes pas classique en théorie des votes. Cependant notre formalisation est cohérente avec les résultats de GUILBAUD (1980) qui montre (dans notre formalisme) que les applications

$$\begin{aligned} c &: H^* \rightarrow (F \wedge F)^* \\ \underline{q} &\mapsto c(\underline{q}) = \underline{c} \end{aligned}$$

$$\begin{aligned} b &: H^* \rightarrow (\Lambda^4 F)^* \\ \underline{q} &\mapsto b(\underline{q}) = \underline{b} \end{aligned}$$

sont linéaires de matrices associées C et B. Nous pouvons donc résumer la situation par le double schéma de dualité



GUILBAUD (1980) illustre sur un exemple d'école la structure des matrices C et B. Nous reproduisons ici ces matrices qui correspondent au cas  $n^* = 3$ , sous une forme cohérente avec notre exposé.

$${}^t C = \begin{pmatrix} +1 & +1 & +1 \\ +1 & +1 & -1 \\ -1 & +1 & +1 \\ -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{matrix} (s,t,u) \\ (s,u,t) \\ (t,s,u) \\ (t,u,s) \\ (u,s,t) \\ (u,t,s) \end{matrix} \quad \begin{matrix} (s,t) & (s,u) & (t,u) \end{matrix}$$

$$\begin{matrix} n^{**} = 3 \\ \text{CONDORCET} \end{matrix} \quad \begin{matrix} n^* = 3 \\ \text{OPINION} \end{matrix} \quad \begin{matrix} \text{BORDA} \end{matrix}$$

Plus généralement B est l'élément de  $\mathcal{M}_{n^*,h}(\mathbb{R})$  de terme général

$$b_{sk} = (n^* + 1) - 2r(s) \quad s \in [n^*], \quad k \in [h] \tag{122}$$

où  $r(s)$  est le rang croissant de  $s$  dans la permutation  $k$ . Par suite la corrélation des vecteurs colonnes de B s'interprète comme la corrélation de Spearman des deux dépouillements associés. L'expression classique de cet indice montre que le choix de la métrique  $\Phi^{-1} = (s^*)^{-1} I_{n^*}$  sur  $(F \wedge F)^*$  permet d'identifier le produit scalaire de deux dépouillements à leur coefficient de corrélation de Spearman, puisque la variance empirique des  $n^*$  premiers entiers est

$$s^* = n^*(n^*+1)(n^*-1)/12$$

et l'on obtient la matrice des corrélations de Spearman  $R_s$  par application de l'identité:

$$R_s = {}^t B \Phi^{-1} B = \frac{12}{n^*((n^*)^2-1)} {}^t B B$$

D'autre part, le coefficient  $b_{sk}$  est égal à la différence :

$$b_{sk} = (\text{nombre d'objets classés avant } s \text{ dans } k) - (\text{nombre d'objets classés après } s \text{ dans } k) \\ = \text{card}([\leftarrow, s]^k) - \text{card}([s, \rightarrow]^k)$$

où  $[\leftarrow, s]^k$  (resp.  $[s, \rightarrow]^k$ ) est l'idéal (resp. le filtre) d'ordre défini par  $s$  pour  $\mathcal{R}^k$ . Cette expression montre que la matrice  $B$  s'écrit comme la différence ( $F - G$ ) des deux matrices  $F = (\text{card}([\leftarrow, s]^k))_{sk}$  et  $G = (\text{card}([s, \rightarrow]^k))_{sk}$ . NISHISATO (1978) montre que ces deux matrices constituent des outils de base pour l'Analyse Linéaire des Données de Préférence. d'AUBIGNY (1981) montre de plus que  $B$  vérifie la propriété de centrage en lignes et en colonnes

$$\sum_s b_{sk} = \sum_k b_{sk} = 0$$

La relation (122) est classique en Statistique non paramétrique puisqu'elle correspond au codage de  $s$  en termes de rangs centrés (HAJEK et ZIDAK (1967)). LEMAIRE (1976) et BENZECRI (1965) interprète ces codages comme les coordonnées des points dans le permutoèdre de  $H$  défini par exemple dans GUILBAUD et ROSENTHIEL (1969) (voir aussi BARBUT et al. (1970), d'AUBIGNY (1981)), pour formaliser les méthodes d'Analyse des Données de Préférences. On peut de plus utiliser cette construction algébrique pour interpréter les travaux de DIACONIS (1982) dans un cadre géométrique, ainsi que leur extension au cas des ordres partiels, due à CRITCHLOW (1985). Les problèmes numériques sont cependant alors sensiblement plus lourds, malgré la communauté des outils théoriques de description des problèmes. En particulier les expériences numériques de d'AUBIGNY et al. (1988) montrent que l'on doit alors se limiter à des tailles  $n$  raisonnables (disons  $n \leq 25$ ) pour obtenir un algorithme de Codage Multidimensionnel Ordinal qui converge suffisamment vite lorsque la dissimilarité observée définit un ordre partiel sur  $\mathcal{C}$ .

**Remarque :** L'extension de ces résultats au cadre théorique défini par  $F = L^2(\Omega, \mathcal{A}, P)$  est en fait immédiat du fait de la structure Hilbertienne de cet espace. Ce cadre de présentation des méthodes, adopté par d'AUBIGNY (1975) n'a cependant pas été retenu dans ce travail, du fait de son apport limité pour l'analyse des données. Il montre cependant de façon plus étroite l'utilité du formalisme développé dans ce mémoire pour l'étude des statistiques de rang. En effet les fonctions scores utilisées en statistique non paramétrique sont des fonctions isotones de l'indice de rang (cf. HAJEK et ZIDAK (1967)). Par suite quand on se limite aux fonctions scores de carré intégrable, on décrit en fait le cône convexe fermé  $K = L^2(\Omega, \mathcal{B}_\Delta, P)$  de  $F$ . HAJEK et ZIDAK (1967) décrivent un petit nombre de fonctions scores et van EEDEN (1970) démontre l'utilité de la régression isotone pour construire la fonction score optimale, associée à un problème de test portant sur le paramètre de position d'une loi inconnue.

D'autre part la matrice  $C$  est l'élément de  $\mathcal{M}_{n+1, n}(\mathbb{R})$  de terme général

$$c_{(s,t)k} = \begin{cases} +1 & \text{si } (s,t) \text{ est classé dans cet ordre dans } k \\ -1 & \text{si } (s,t) \text{ est classé dans l'ordre } (t,s) \text{ dans } k \end{cases}$$

**Définition 5** (GUILBAUD (1980)) : Les applications  $b$  et  $c$  de matrices  $B$  et  $C$  définies sur l'espace des opinions  $H^*$  et à valeurs dans l'espace des dépouillements de Borda et de Condorcet respectivement sont appelées respectivement application de Borda et application de Condorcet.

Guilbaud appelle  $B$  la matrice de Borda et  $C$  la matrice de Condorcet. d'AUBIGNY (1981) reprenant les travaux de NISHISATO (1978, 1980) montre que la matrice  $B$  (qu'il note alors  $T$ ) joue un rôle central dans la présentation des méthodes d'Analyse Linéaire des Données de Préférences et interprète dans ce cadre le vecteur des dépouillements de Borda  $\underline{b}$  comme un vecteur d'utilités.

D'autre part, la liaison linéaire entre  $(\Lambda^4 F)^*$  et  $(F \wedge F)^*$  est alors déduite du schéma de dualité précédent. Interprétons en effet celui-ci comme le schéma descriptif de l'Analyse Canonique du triplet  $(B, C, N)$  où  $N$  est la métrique diagonale de poids  $\frac{1}{h}$ , définie sur  $H$ . Alors (cf. CAILLEZ et PAGES (1976)), les facteurs canoniques issus de cette analyse sont obtenus de façon équivalente en terme de  $\Pi$ -stratégie, par l'Analyse en Composantes Principales du triplet  $(V_{21}, V_{11}, V_{22})$  où

$$\begin{aligned} V_{21} &= BN^t C = \frac{1}{h} B^t C \\ V_{11} &= CN^t C = \frac{1}{h} C^t C \\ V_{22} &= BN^t B = \frac{1}{h} B^t B \end{aligned}$$

L'interprétation concrète de cette analyse est éclairée par le résultat suivant, dû à Guilbaud (1980):

**Théorème 2:**

$$\begin{aligned} i) \quad \frac{n^* + 1}{3} T^* &= \frac{1}{h} B^t C \\ ii) \quad T^* C &= B \end{aligned}$$

Cet auteur appelle  $T^*$  la matrice de Kirchhoff. Il note en effet que d'après Oswald Veblen, ce type de matrices aurait été introduit par KIRCHHOFF (1848) lors de l'étude des réseaux électriques. Il rejoint ainsi l'interprétation du problème de Codage Multidimensionnel donnée plus haut en termes de réseaux électriques et montre que cette analogie doit être exploitée plus avant pour interpréter les méthodes actuelles de Codage Multidimensionnel et développer de nouveaux types d'analyse. Guilbaud observe de plus la

conséquence immédiate du théorème 2 ci-dessus, qui fournit en particulier une décomposition spectrale des opérateurs  $V_{11}$  et  $V_{22}$ .

**Corollaire 2 :** (GUILBAUD (1980):

$$i) ({}^tCC) {}^tB = \frac{(n^* + 1)!}{3} {}^tB$$

$$ii) V_{11} {}^tT^* = \frac{(n^* + 1)}{3} {}^tT^*$$

$$iii) V_{22} = \frac{(n^* + 1)}{3} T^* {}^tT^*$$

La matrice de connexion  $W_C = {}^tCC$  est un élément de  $\mathcal{Y}_H$  et a pour termes hors diagonaux les distances de Kendall entre permutations. Par suite, lorsqu'on munit  $(\Lambda^4 F)^*$  de la structure Euclidienne induite par la métrique

$$\zeta^{-1} = \frac{1}{n^{**}} I_{n^{**}}$$

on obtient (cf. GUILBAUD (1980)) le coefficient de Kendall  $\tau_C(p,q)$  entre les permutations  $p$  et  $q$  de  $[n^*]$  comme terme général de la matrice

$$K = {}^tC \zeta^{-1} C = {}^tC \frac{1}{n^{**}} I_{n^{**}} C = \frac{1}{n^{**}} {}^tC C$$

L'utilisation de ces résultats en Codage Multidimensionnel Ordinal Multiple permet en particulier d'utiliser l'une des techniques décrites dans d'AUBIGNY (1981) pour construire une fonction de dissimilarité résumé de  $\{\Delta^k; k \in [K]\}$  (ou de façon équivalente une préordonnance résumé de  $\{\mathcal{R}^k; k \in [K]\}$ ). Ce résumé fournit alors une disparité utilisable pour initialiser une procédure de Codage Multidimensionnel Ordinal Multiple. On soumet en effet ce résumé  $\Theta^{(0)}$  à un algorithme de Codage Multidimensionnel Ordinal Simple afin de construire une configuration de départ  $X^{(0)}$  de la dimension souhaitée. La construction de la solution optimale  $X^*$  est alors réalisée par itération. d'AUBIGNY (1987) associe en effet à chaque fonction de dissimilarité observée  $\delta^k$  le cône polyédrique  $C^k$  des fonctions de disparité équivalentes à  $\delta^k$ . La solution du problème d'optimisation

$$\text{Min} \left\{ \frac{1}{K} \sum_{k=1}^K \text{STRESS}^k(m, X); m \in C, X \in \mathcal{M}_{r,n}(\mathbb{R}), XD_{q1} = 0, XD_{qt}X = uI \right\}$$

où

$$\text{STRESS}^k(m, X) = \sum_{i < j} \sum q_{ij} (m(\delta_{ij}^k) - d_{ij}(X))^2$$

$$C = \bigcap_{k=1}^K C^k$$

Comme précédemment la solution est construite par un algorithme alternant deux phases:

Phase (1) : Pour une dissimilarité  $\Theta^{(v)}$  donnée, construction de  $X^{(v)}$  à l'aide d'un algorithme de Codage Multidimensionnel Ordinal, qui optimise la fonction  $\text{STRESS}(\hat{m}^{(v)}, X)$  définie dans le patron d'algorithme au paragraphe §VII-1-1.

Phase (2) : Pour une configuration  $X^{(v)}$  donnée, construction de  $\hat{m}^{(v+1)}$  par un algorithme de projection sur  $C$ . Nous montrons alors que tout algorithme de projection sur une intersection de cônes convexes fermés peut être utilisé à cette fin. Si  $C$  est non vide, nous démontrons que nous obtenons la solution souhaitée qui est la  $D_Q$ -projection de  $d(X^{(v)})$  sur  $C$ . De plus si  $C$  est vide, l'algorithme présenté dans d'AUBIGNY (1987) construit le meilleur compromis au sens de l'approximation en moyenne quadratique. La solution est alors le barycentre de ses projections sur les  $C^k$ .

## REFERENCES

- AGRESTI, A. (1984) Analysis of Ordinal Categorical Data. John Wiley. New-york.
- AITCHISON, J. (1983) "Principal Component Analysis of Compositional Data." Biometrika Vol 70, 1: 57-65.
- AITCHISON, J. (1986) The Statistical Analysis of Compositional Data. Chapman and Hall. New-york.
- ALVO, M., P. CABILLO et P.D. FEIGIN. (1982) "Asymptotic Theory for Measures of Concordance with Special Reference to Average Kendall Tau." Ann. Statist. 10: 1269-1276.
- ANASTASSAKOS, I. et G. d'AUBIGNY. (1984). "L'utilisation des Tests de Sphéricité pour la Recherche de la Dimension de l'Espace Latent en Analyse Factorielle Classique et en Analyse en Composantes Principales." Revue de Statistique Appliquée, Vol XXXII, No 3 : 45-57.
- ANDREWS, D.F. (1972) "Plots of High-Dimensional Data." Biometrics 28: 125-136.
- ARABIE, P. (1973). "Concerning Monte Carlo Evaluations of Nonmetric Multidimensional Scaling." Psychometrika 38 : 607-608.
- ARABIE, P. (1973). "The Difference Between 'Several' and 'Single' : a Reply to Spence and Young." Psychometrika 43 : 119.
- ARABIE, P., et SHEPARD, R.N. (1979). "Representation of Similarities as Additive Combinations of Discrete, Overlapping Properties." Psychological Review, Vol 86, No 2 : 87-123.
- ATTNEAVE, F. (1950). "Dimensions of Similarity". American Journal of Psychology, 63 : 515-516.
- d'AUBIGNY, G. (1975). "Description Statistique des Données Ordinales: Analyse Multidimensionnelle." Thèse de 3ème cycle. Université de Grenoble I.
- d'AUBIGNY, G. (1979). "A Probabilistic Look at Ordinal Methods In Data Analysis." In P. MANDL et M. HUSKOVA, (eds), Proceedings of the Second Prague Symposium on Asymptotic Statistics. North Holland, Amsterdam: 23-40.
- d'AUBIGNY, G. (1981). "Analyse Factorielle et Visualisation des Préférences." In P. BATTEAU et al.,(eds), Analyse et Agrégation des Préférences. Economica, Paris, pp 177-212.
- d'AUBIGNY, G. (1984). "Un Modèle Algébrique pour l'Analyse des données de dissimilarité." Troisièmes Journées Franco-Soviétiques de Statistique et Analyse des Données. Paris. Tech. Rep. IMSS (1987), Grenoble.
- d'AUBIGNY, G. (1988). "The Additive Decomposition of Some Entropy Functions and (Constrained-) Ordination Methods." Papier présenté à la XIV-th International Biometric Conference. Namur.(à paraître).
- d'AUBIGNY, G. et M.O. POLIT. (1988). "Some Optimality Properties of the Generalization of the TUCKER Method to the Analysis of n-way Tables with Specified Metrics." In L. COPPI, (ed.), Proceedings of the MULTIWAY 88 Conference. Rome.
- d'AUBIGNY, G., PHAM DINH, T., et A. YASSINE. (1988). "A Conditional Conjugate Gradient Method for Solving Isotonic and Concave Regression Problems." Computational Statistics and Data Analysis. (à paraître).
- BALESTRA, P. (1975). La Dérivation Matricielle. Sirey, Paris.
- BANACH, S. (1932). Théorie des Opérateurs Linéaires. Monografje Mathematiczne, Vol. I. Warsaw.
- BANET, T.A. et L. LEBART. (1984). "Local and Partial Principal Component Analysis and Correspondence Analysis." In COMPSTAT. Vienna : Physica Verlag, pp 113-118.
- BARBUT, M. et al. (1970). Ordres Totaux Finis. Gauthier-Villars. Paris.
- BARLOW, R.E., D.J. BARTHOLOMEW, J.M. BREMER et H.D. BRUNK (1972). Statistical Inference under Order Restrictions : The Theory of Isotonic Regression. John Wiley, New-York.

- BARTHELEMY, J.-P. et A. GUENOCHÉ (1988) Les Arbres et les Représentations des Proximités. Masson. Paris.
- BASTIN, C., J.-P. BENZECRI, C. BOURGARIT, et P. CAZES (1980). " Pratique de l'Analyse des Données." Dunod. Paris.
- BATTEAU, P., E. JACQUET-LAGREZE, et B. MONJARDET. (1981). Analyse et Aggrégation des Préférences dans les Sciences Economiques et de Gestion. Economica. Paris.
- BEALS, R. et D.H. KRANTZ. (1967). "Metrics and Geodesics Induced by Order Relations." Mathematische Zeitschrift 101 : 285-298.
- BEALS, R., D.H. KRANTZ et A. TVERSKY. (1968). "Foundations of Multidimensional Scaling." Psychological review 75 : 127-142.
- BECHTEL, G.G. (1976) Multidimensional Scaling. The Hague-Mouton. Paris.
- BELSEY, D.A., E. KUH et R.E. WELSH. (1980). Regression Diagnostics : Identifying Influential Data and Sources of Collinearity. John. Wiley : New York.
- BENINEL, F. (1987). Problèmes de Représentations Sphériques des Tableaux de Dissimilarité. Thèse de Troisième Cycle, Univ. Rennes I.
- BENTLER, P.M. (1969). " Some Extensions of Image Analysis." Psychometrika 34, 1: 77-83.
- BENZECRI, J.-P. (1964). " Sur l'Analyse Factorielle des Proximités." Publication de l'ISUP I, Vol 13.
- BENZECRI, J.-P. (1965). " Sur l'Analyse Factorielle des Proximités." Publication de l'ISUP I, Vol 14.
- BENZECRI, J.-P. (1965). L'Analyse des Préférences. Cours Polycopié. ISUP, Paris.
- BENZECRI, J.-P. et al. (1973). L'Analyse des Données. Dunod. Paris.
- BERGE, C. (1959). Espace, Topologie, Fonctions Multivoques. Dunod. Paris.
- BERGE, C. (1973). Graphes et Hypergraphes. Dunod. Paris.
- BERTRANEU, A. (1979). L'Analyse des Proximités. Bases Géométriques et Méthodes. Thèse de 3-ème cycle. Université Paul Sabatier. Toulouse.
- BETH E.W. et A. TARSKI (1956). " Equilaterality as the Only Primitive Notion of Euclidean Geometry." Indagationes Math. 18: 462-467.
- BICK, W., H. BAUER, P.J. MUELLER, et O. GIESEKE. (1978). Multidimensional Scaling and Clustering Techniques : Theory and Applications in Sociology. Institute of Applied Social Research, University of Cologne.
- BOCK, R.D. et L.V. JONES (1968). The Measurement and Prediction of Judgment and Choice. Holden-Day. San Francisco.
- BLUMENTHAL, L.M. (1938). Distance Geometry. Univ. of Missouri Studies. 13,2.
- BLUMENTHAL, L.M. (1953). Theory and Applications of Distance Geometry. Clarenton Press. Oxford.
- BLUMENTHAL, L.M. et K. MENGER (1970). Studies in Geometry. Freeman and co.. San Francisco.
- BLYTH, T.S. et M.F. JANOWITZ (1972). Residuation Theory. Pergamon Press. Oxford.
- BONIFAS, L., Y. ESCOUFIER, P.L. GONZALEZ et R. SABATIER. (1984). " Choix de Variables en Analyse en Composantes Principales." R. Statist. Appli. Vol XXXII, No. 2 : 5-15.
- BONNEFOUS, S., E. CLOUTIER et E. ROY. (1978). " Analyse Multidimensionnelle des Proximités, Techniques, Programmes APL, Exemples." Rapport CEA-R-4910. CEN Saclay. Gif sur Yvette.
- BORG, I. (1981 a). Anwendungsorientierte Multidimensionale Skalierung. Springer-Verlag. Heidelberg.
- BORG, I. (ed.) (1981 b). Multidimensional data representations: When and Why? Mathesis Press.
- BORG, I. et J.C. LINGOES. (1981). " An Alternative Approach to Confirmatory Inference and Geometric Models." Quality and Quantity 15 : 541-552.
- BORTZ, J. (1974) " Kritische Bemerkungen über den Einsatz Nicht-Euklidischer Metriken im Rahmen der Multidimensionalen Skalierung." Archiv für Psychol., bd. 126, S: 196-212.
- BOUGHAZI, M.A. (1987). Contribution à l'Etude des Algorithmes d'Optimisation en Analyse des Données. Thèse de 3ème cycle. Université de Grenoble I.
- BRADU, D. (1975). " Search and Methods for Interaction Originating in a Small Set of Outliers." Edv in Medizin und Biologie 4 : 93-100.

- BRADU, D. (1984). " Response Surface Model Diagnosis in Two-way Tables." Communications in Statist., Theor. Meth. 13, 24: 3059-3106.
- BRADY, H.E. (1985). " Statistical Consistency and Hypothesis Testing for Nonmetric Multidimensional Scaling." Psychometrika 50 : 509-537.
- BRETAGNOLLE, J., D. DACUNHA-CASTELLE, et J.L. KRIVINE (1966). " Lois Stables et Espaces L<sup>P</sup>." Ann. Inst. Henry Poincaré Vol. II, 3: 231-259.
- BROSSIER, G. (1986). " Problèmes de Représentations de Données par des Arbres." Thèse d'Etat, Université de Rennes I.
- BROWNE, M.W. (1985). " The Young-Householder Algorithm and the Least Squares Multidimensionnal Scaling of Squared Distances." (Abstract) In de LEEUW et al. (eds.), Multidimensional Data Analysis. (1986) DSWO Press, Leiden, p. 239.
- BROWNE, M.W. (1987). "The Young-Householder Algorithm and the Least Squares Multidimensionnal Scaling of Squared Distances." Research Report No 87/1, University of South Africa.
- BUNEMAN, P. (1971). " The Recovery of Trees from Measures of Dissimilarity." In FR. HODSON, D.G. KENDALL and P. TAUTU (eds.) Mathematics in the Archeological and Historical Sciences. Edimbourg University Press : Edimbourg. PP 387-395.
- BUNEMAN, P. (1974). " A note on metric properties of trees." Journal of Combinatorial Theory, (B) 17 : 48-50.
- BUSEMAN, H. (1955). The Geometry of Geodesics. Academic Press.
- CAELLI, T. (1981). " On Generating Spatial Configurations with Identical Interpoint Distance Distributions." In ROBINSON, R.W., SOUTHERN, G.W., WALLIS, W.D. eds. Lecture Notes In Mathematics. Springer Verlag, New York.
- CAILLIEZ, F. (1983). " The Analytical Solution of the Additive Constant Problem." Psychometrika 48 : 305-310.
- CAILLIEZ, F. (1984). Analyse des Données. Les Presses de l'Université de Montréal. Montréal.
- CAILLIEZ, F., PAGES, J.-P. (1976). Introduction à l'Analyse des Données. SMASH. 9 rue DUBAN, Paris.
- CARROLL, J.D. (1972). " Individual Differences and Multidimensional Scaling." In R. N. Shepard, A. K. Romney et S.B. Nerlove (eds.), Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Vol. 1. Seminar Press, New-York. pp 105-157.
- CARROLL, J.D. (1983) " Common Dimension Analysis." Paper presented at the Multidimensional Data Analysis Workshop. Paris.
- CARROLL, J.D. (1985). " Review of Multidimensional Scaling." Psychometrika 50 : 133-140.
- CARROLL, J.D. et J.J. CHANG. (1970). " Analysis of Individual Differences in Multidimensional Scaling via a N-way Generalization of 'ECKART-YOUNG' Decomposition." Psychometrika 35 : 283-319.
- CARROLL, J.D. et P. ARABIE (1980). " Multidimensional Scaling." In M.R. ROSENZWEIG and L.W. PORTER, (eds), Annual Review of Psychology 31 : 607-649.
- CARROLL, J.D. et S. PRUZANSKI. (1980). " Discrete and Hybrid Scaling Models." In E.D. Lanterman and H. Feger (eds.) Similarity and Choices. Huber : Berne.
- CAYLEY, A. (1841). " On a Theorem in the Geometry of Position." Cambridge Math. J. 2: 167-271.
- CHATTERJEE, S., HADI, A.S. (1986). " Influential Observations, High Leverage Points, and Outliers in Linear Regression." Statistical Science 1 , No 3: 379-416.
- CHINO, N. (1978). " A Graphical Technique for Representating the Asymmetric Relationship Between n Objets." Behaviometrika, 5 : 23-40.
- CLARET, A. (1987). Contribution au problème de l'Approximation d'un Tableau de Données. Thèse de Troisième Cycle. Université des Sciences et Techniques du Languedoc. Montpellier.
- CLARK, A.K. (1976). " Re-Evaluation of Monte Carlo Studies in Nonmetric Multidimensional Scaling." Psychometrika 41 : 401-404.
- CLIFF, N. (1973). " Scaling." Annual Psychological Review 24 : 473-506.

- COLLEDGE, R.G. et G. RUSHTON. (1972). "Multidimensional Scaling: Review and Geographical Applications." Geographical technical papers series No. 10. Association of American Geographers. Washington.
- CONSTANTINE, A.G. et J.C. GOWER. (1971). "Graphical Representation of Asymmetric Matrices." Appl. Statistics, 7 : 297-304.
- COOK, R.D., WEISBERG, S. (1982). Residuals and Influence in Regression. Chapman and Hall, New York.
- COOPER, L.G. (1972). "A New Solution to the Additive Constant Problem in Metric Multidimensional Scaling." Psychometrika 37 : 311-322.
- COXETER, H.S.M. (1933). "On points with Arbitrarily Assigned Mutual Distances." Proc. Cambridge Phil. Soc. Vol XXX, part I: 1-3.
- COXETER, H.S.M. (1948). Regular polytopes. London : Methuen and Co.
- COXON, A.P.M. et C.L. JONES. (1974 a). "Occupational Similarities : Subjective Aspects of Social Stratification." Quality and Quantity 22 : 369-384.
- COXON, A.P.M. et C.L. JONES. (1974 b). "Problems in Selection of Occupational Titles." The Sociological Review 22 : 369-384.
- CRIPPEN, G.M. (1977). "A Novel Approach to Calculations of Conformation: Distance Geometry." J. of Computational Physics 24 : 96-107.
- CRIPPEN, G.M. (1978). "Rapid Calculation of Coordinates from Distance Measures." J. of Computational Physics 26 : 449-452.
- CRITCHLEY, F. (1978). "Multidimensional Scaling: A Short Critique and a New Method." In. L.C.A. CORSTEN et J. HERMANS (Eds.) COMPSTAT. Physika Verlag. Vienna.
- CRITCHLEY, F. (1982 a). "Multidimensional Scaling Theory: Some Basic Mappings." Techn. Rept. Univ. of Glasgow.
- CRITCHLEY, F. (1982 b). "Multidimensional Scaling Theory: Some Basic Sets." Techn. Rept. Univ. of Glasgow.
- CRITCHLEY, F. (1985). "Dimensionality Theorems in Multidimensional Scaling and Hierarchical Cluster Analysis." Proceedings of the fourth International Symposium : Data Analysis and Informatics. North Holland. pp 85-110.
- CRITCHLOW, D.E. (1985). Metric Methods for Analysing Partially Ranked Data. Springer-Verlag. Heidelberg.
- CROQUETTE, A. (1980). "Quelques Résultats Synthétiques en Analyse Des Données Multidimensionnelles: Optimalité et Métriques à Effets Relationnels." Thèse de Troisième Cycle, Université Paul Sabatier, Toulouse.
- CUNNINGHAM, J.P. et R. SHEPARD, (1974). "Monotone Mapping of Similarities into a General Metric Space." J. of Math. Psychol. Vol. 11, 4: 335-363.
- DANIELS, H.E. (1944). "The Relation Between Measures of Correlation in the Universe of Sample Permutations." Biometrika 33 : 129-135.
- DAUGAVET, V.A. (1968). "Variant of Stepped Exponential Method of Finding Some of the First Characteristic Values of a Symmetric Matrix." URSS Comp. Math. Physics 8 (1): 212-222.
- DAUXOIS, J., POUSSE, A. (1976). "Les Analyses Factorielles en Calcul des Probabilités et en Statistique: Essais d'étude synthétique." Thèse d'Etat, Université Paul Sabatier, Toulouse.
- DAVID, H.A. (1988). The Method of Paired Comparisons. (Second edition). Charles Griffin and Co. London.
- DAVIDSON, M.L. (1983). Multidimensional Scaling. John Wiley. New York.
- DAVY de VIRVILLE, M. (1972). l'Analyse des Proximités. Thèse de 3-ième cycle. Paris.
- DEFAYS, D. (1978). "A Short Note on a Method of Seriation." British J. of Math. Statist. Psychol. 31: 49-53.
- DEGENNE, A. (1972). Techniques Ordinales en Analyse des Données Statistiques. Hachette. Paris.
- DEMPSTER, A.P., M. SCHATZOFF et N. WERMUTH. (1977). "A Simulation Study of Alternatives to Ordinary Least Squares." J. Amer. Statist. Assoc., 72 : 77-91.
- DENIS, J.-B. (1983). Interaction Entre Deux Facteurs. Thèse, Inst. National Agronomique Paris-Grignon. Paris
- DENIS, J.-B. et E. SPJOTVOLL (1980). "Plots of 2x2 Contrasts for Two-Way Tables." Doc. Interne. Dpt. of Math. and Statist., Agricultural Univ. of Norway. 60p.

- DIACONIS, P. (1982). Group Theory in Statistics. Harvard Univ. Lectures Notes.
- DILLON, W.R. et M. GOLDSTEIN. (1984). Multivariate Analysis : Methods and Applications. John Wiley: New-York.
- DIRICHLET, G.L. (1850) "Uber die Reduction der Positiven Quadratischen Formen mit drie Unbestimmten Ganzen Zahlen." J. fur die Reine und Angewandte Math. **50**: 209-227.
- DROESLER, J. (1981 a). " The Empirical Validity of Multidimensional Scaling." In Multidimensional Data Representations: When and Why. I. BORG, (Ed.), Ann Arbor, Michigan : Mathesis Press, pp 627-651.
- DROESLER, J. (1981 b). " Foundations of Multidimensional Scaling in Cayley-Klein Geometry." In Multidimensional Data Representations: When and Why. I. BORG, (ed.), Ann Arbor, Michigan : Mathesis Press, pp 652-678.
- DUBIKAJTIS, L. (1977). " Relationship Between the Notions of Distance, Area and Volume in the Euclidean Space." Prace Matematyczne T. 7: 71-82.
- DYKSTRA, R.L. (1983). " An Algorithm for Restricted Least Squares Regression." J. Amer. Statist. Assoc. **78**: 837-842.
- DYKSTRA, R.L. (1983). " Dual convex cones of order restrictions with applications." In Y.L. TONG (ed.) Inequalities in Statistics and Probability, Lectures Notes. Monograph Series, Institute of Math. Statistics, Hayward, California.
- van EEDEN, C. (1956). " Maximum Likelihood Estimation of Ordered Probabilities." Indagationes. Math. **18** : 444-455.
- van EEDEN, C. (1957a). " Maximum Likelihood Estimation of Partially or Completely Ordered Parameters I." Indagationes. Math. **19** : 128-136.
- van EEDEN, C. (1957b). " Maximum Likelihood Estimation of Partially or Completely Ordered Parameters II." Indag. Math., **19** : 201-211.
- van EEDEN, C. (1970). " Efficiency-Robust Estimation of Location." Ann. Math. Statist., **41**: 172-181.
- EFIMOV, N. (1981). Géométrie Supérieure. Editions MIR, Moscou.
- EKMAN, G. (1954). " Dimensions of Color Vision." J. of Psychology, **38**: 467-474.
- EKMAN, G. (1960). " A Direct Method for Multidimensional Ratio Scaling." Psychometrika, **28**: 33-41.
- EL HADRI, K. (1988). Lissage Monotone et Optimisation non Différentiable pour la résolution d'un problème de Codage Multidimensionnel. Thèse . Université de Grenoble II.
- ESCOUFIER, Y. (1970). " Echantillonnage dans une Population de Variables Aléatoires Réelles." Publ. Ann. Inst. Univ. Paris, **19**, fasc. 4 : 1-47.
- ESCOUFIER, Y. (1973). " Le Traitement des Variables Vectorielles." Biometrics **29** : 751-760.
- ESCOUFIER, Y. (1975). " Le Positionnement Multidimensionnel." Rev. de Statist. Appli., Vol XXIII, No 4 : 5-14.
- ESCOUFIER, Y. et A. GRORUD. (1980). " Analyse Factorielle des Matrices Carrées non Symétriques." In E. Diday et al. (eds), Data Analysis and Informatics. North Holland : Amsterdam. pp. 263-276.
- EVANS, A.B. (1983). " Distance in Finite Geometries I: Translation Planes." Geometriae Dedicata **15**: 59-68.
- FASHAM, M.J.R. (1977). " A Comparison of Non Metric Multidimensional Scaling, Principal Components Analysis, and Reciprocal Averaging for the Ordination of Simulated Coenoclines and Coenoplanes." Ecology **58**: 551-561.
- FEIGIN, P.D. et A. COHEN (1978). " On a Model for Concordance Between Judges." J. Royal Statist Soc. Ser. B **40** : 203-213.
- FEIGIN, P.D. et M. ALVO (1986). " Intergroup Diversity and Concordance for Ranking Data. An Approach Via Metrics for Permutations." Ann. Statist. Vol. 14, 2 : 691-707.
- FICHET, B. (1983). " Analyse Factorielle sur Tableaux de Dissimilarité: Application aux Données sur Signe de Présence-Absence en Médecine." Thèse d'état, Univ. Marseille.
- FICHET, B. (1985). " Distances and Euclidean Distances for Presence-Absence Characters and their Application to Factor Analysis." In de LEEUW et al. (eds.), Multidimensional Data Analysis. (1986) DSWO Press, Leiden, pp 23-46.

- FICHET, B. (1987). "The role Played by  $L_1$  in Data Analysis." in Y. DODGE (ed.) Statistical Data Analysis Based on the  $L_1$ -Norm and Related Methods. North-Holland. Amsterdam: pp. 185-194.
- FICHET, B. et E. GAUD (1982). "Quelques Resultats sur la Dimension et la Détermination d'une Préordonnance." Papier présenté aux Journées de Statistique, à Bruxelles.
- FISHER, R.A. (1922) "The Systematic Location of Genes by Means of Cross-Over Ratios." The Amer. Naturalist 56: 406-411.
- FORGAS, J.P., E. KAGEN et D. FREY. (1977). "The Cognitive Representation of Political Personalities. A Cross-cultural Comparison." International J. of Psychology 12 : 19-30.
- FRASER, D.A.S. (1968). Inference and Linear Models. McGraw Hill : New-York.
- FRASER, D.A.S. (1968). The Structure of Inference. John Wiley : New-York.
- FRECHET, M. (1935). "Sur la Définition Axiomatique d'une Classe d'Espaces Distanciés Vectoriellement Applicables sur l'Espace de Hilbert." Annals of Mathematics, 36 : 705-718.
- FRIEDMAN, J. et R. TIBSHIRANI. (1984). "The Monotone Smoothing of Scatterplots." Technometrics, Vol 26, No 3 : 243-250.
- FUNIA, S., A. HOROWITZ et F.W. YOUNG. (1976). "The Perceived Structure of American Ethnic Groups : the Use of Multidimensional Scaling Stereotype Research." Sociometry 39 : 116-130.
- GABRIEL, K.R., J. PUTTER et Y. WAX (1973). "Simultaneous Confidence Intervals for Product-Type Interaction Contrasts." J. Royal Statist. Soc. Vol. 35, 2: 234-244.
- GAUCH, H.G. (1982). Multivariate Analysis in Community Ecology. Cambridge University Press, Cambridge.
- GAUCH, H.G., WHITTAKER, R.H., SINGER, S.B. (1981). "A Comparative Study of Non Metric Ordinations." J. of Ecology 69: 135-152.
- GAUCH, H.G., WHITTAKER, R.H., WENTWORTH, T.R. (1977). "A Comparative Study of Reciprocal Averaging and Other Ordination Techniques." J. of Ecology 65: 157-174.
- GAUD, E. (1983). Représentation d'une préordonnance. Thèse de 3-ième Cycle. Université de Provence, Centre Saint-Charles, Marseille.
- GAUSS, K.F. (1831). In Göttingsche Gelehrte Anzeigen 2: 1065.
- GIFI, A. (1981). Non-Linear Multivariate Analysis. Dpt. of Data Theory, Univ. of Leiden.
- GILBERT, E.N. (1974). "Distorsion in Maps." SIAM Review 16 : 47-62.
- GLAZMAN, I. et Y. LIUBITCH. (1972). Analyse Linéaire dans les Espaces de Dimension Finie. Editions MIR, Moscou.
- GOLLEDGE, R.G. et J.N. RAYNER (eds.) (1982). Proximity and Preferences. Problems in Multidimensional Analysis of Large Set of Data. University of Minnesota Press : Minneapolis.
- GOLLOB, H.F. (1968). "A Statistical Model Which Combines Features of Factor Analysis and Analysis of Variance." Psychometrika 33: 73-116.
- GOODMAN, L.A. (1979 a) "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." J. Amer. Statist. Assoc. 74: 537-552.
- GOODMAN, L.A. (1979 b) "Multiplicative Models for Square Contingency Tables with Ordered Categories." Biometrika 66: 413-418.
- GOODMAN, L.A. (1981) "Association Models and Canonical Correlation Analysis in the Analysis of Cross-Classifications Having Ordered Categories." J. Amer. Statist. Assoc. 76: 320-334.
- GOODMAN, L.A. (1983) "The Analysis of Dependence in Cross-Classifications Having Ordered Categories Using Log-Linear Models for Frequencies and Log-Linear Models for Odds." Biometrics 39: 149-160.
- GOODMAN, L.A. et W.H. KRUSKAL (1979). Measures of Association for Cross Classifications. Springer-Verlag. Heidelberg.
- GORDON, A.D. (1979). "A Measure of Agreement Between rankings." Biometrika 66: 7-15.
- GORDON, A.D. (1983). "Classification in the Presence of Constraints." In Analyse des Données et Informatique, R.TOMASSONE (ed.), INRIA, Le Chesnay, pp 161-171.

- GOURIEROUX, C. , A. HOLLY et A. MONFORT (1982). " Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters." Econometrica, Vol. 50, No 1: 63-79.
- GOWER, J.C. (1966). "Some Distance Properties of Latent Roots and Vectors Used in Multivariate Analysis." Biometrika 53 : 325-338.
- GOWER, J.C. (1971). " Statistical Methods of Comparing Different Multivariate Analyses of the same Data." In F.R. HODSON, D.G.KENDALL and P. TAUTU (eds), Mathematics in the Archaeological and the Historical Sciences. Edimburgh University Press, pp 138-149.
- GOWER, J.C. (1977). " The Analysis of Asymmetry and Orthogonality." In J.R. Barra et al. (eds.), Progress in Statistics. North Holland : Amsterdam.
- GOWER, J.C. (1982a). " Euclidean Distance Geometry." Mathematical Scientist, 7: 1-14.
- GOWER, J.C. (1982b). " Multivariate Analysis: Ordination Multidimensional Scaling and Allied Topics." Séminaire Scientifique OTAN (ASI 82/62) Montreal, Canada, publié in W. LEDERMANN (ed.) (1984), Handbook of Applicable Mathematics, Vol. VI, Part B: 727-782.
- GOWER, J.C. (1983). " Distance Matrices and Their Euclidean Approximation." Proceedings of the third International Symposium : Data Analysis and Informatics. North Holland. pp 1-19.
- GOWER, J.C. (1985). " Euclidean Distance Matrices." In de LEEUW et al. (eds.), Multidimensional Data Analysis. (1986) DSWO Press, Leiden. pp. 67-80.
- GRAYBILL, F.A. (1961). An Introduction to Linear Statistical Models. Vol I. McGraw-Hill. New-York.
- GREEN, P.E. et F.J. CARMONE (1970). Multidimensional Scaling and Related Techniques in Marketing Analysis. Allyn and Bacon : Boston.
- GREEN, P.E. et P.R. RAO (1972). Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms. Holt Rinehart and Winston : New-York.
- GREEN, P.E. et Y WIND (1973). Multiattribute Decisions in Marketing: A Measurement Approach. The Dryden Press : Hinsdale.
- GREENACRE, M.J. (1978). Some Objective Methods of Graphical Display of a Data Matrix. Special Report, University of South Africa, Pretoria.
- GREGSON, R.A.M. (1975). Psychometrics of Similarity. Academic Press, New York.
- GREUB, W.H. (1967). Multilinear Algebra. Springer Verlag, New York.
- GUILBAUD, G.T. (1980). " Relation entre deux coefficients de corrélation de rangs." Mathématiques et Sciences Humaines, 72 : 45-59.
- GUILBAUD, G.T. et P. ROSENTHIEL, (1963). " Analyse Algébrique d'un Scrutin." Mathématiques et Sciences Humaines, 4 : 48-62.
- GUTTMAN, L. (1954). " The Principal Components of Scalable Attitudes" In LAZARFELD, P.F. (ed.), Mathematical Thinking in the Social Sciences, pp 216-348.
- GUTTMAN, L. (1965). " A Definition of Dimensionality and Distance for Graphs." In LINGOES et al.(eds.), Geometric Representations of Relational Data. (1977). Mathesis Press, Ann Arbor, pp 707-712.
- GUTTMAN, L. (1968). " A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points." Psychometrika 33 : 469-504.
- GUTTMAN, L. (1969). " Smallest space Analysis by the Absolute Value Principle." Paper presented at the Symposium Theory and Practice of Measurement. XIX International Congress of Psychology, Londres.
- HABERMAN, S.J. (1978). Analysis of Qualitative Data. Vol I. Introductory Topics. Academic Press. New-York..
- HABERMAN, S.J. (1978). The Analysis of Frequency Data.. Univ. of CHICAGO Press. Chicago.
- HABERMAN, S.J. (1979). Analysis of Qualitative Data. Vol 2. New Developments. Academic Press. New-York..
- HAJEK, J. et Z. SIDAK. (1967). Theory of Rank Tests. Academic Press. New-York.
- HAKIM, M., E.O. LOCHARD, J.P. OLIVIER et E. TEROUANNE (1973 a). " Sur les Traces de SPEARMAN. " Doc IMSS.No 3 " Travaux du Petit Séminaire sur l'Analyse Factorielle Classique." : pp 1-27.,

- HAKIM, M., E.O. LOCHARD, J.P. OLIVIER et E. TEROUANNE (1973 b). " Sur les Traces de SPEARMAN (II): Le problème des Rangs " Doc IMSS.No 4 " Travaux du Petit Séminaire sur l'Analyse Factorielle Classique.": pp 1-18.,
- HANHAM, R.Q. (1976). " Factorial Ecology in Space and Time." Environment and Planning 8: 823-828.
- HARDWICK, J. (1986). " 2R Diagnostics." BMDP Communications. Vol. 19, No 2 : 12-14.
- HARMAN, H.H. (1976). Modern Factor Analysis. Third Edition Revised. The Univ. of Chicago Press. Chicago.
- HARRIS, C.W. (1962). " Some Rao Guttman relationships." Psychometrika 27 : 247-263.
- HARSHMAN, R.A. (1970). " Foundations of the PARAFAC procedure: Models and conditions for an "exploratory" multimodal factor analysis." U.C.L.A. Working Papers in Phonetics 16.
- HARSHMAN, R.A. (1972 ). " Determination and Proof of Minimum Uniqueness Conditions for PARAFAC 1." U.C.L.A. Working Papers in Phonetics 22.
- HARTMANN, W. (1979). Geometrische Modelle zur Analyse Empirischer Daten. Akademie-Verlag. Berlin..
- HAYASHI, C. (1952). " On a Prediction of Phenomena from Qualitative Data and Quantification of Qualitative Data from the Mathematico-Statistical Point of View." Annals Inst. Statist. Math., 4: 19-30.
- HAYASHI, C. (1964). " Multidimensional quantification of the data obtained by the method of paired comparison." Annals Inst. Statist. Math., 16: 231-245.
- HAYASHI, C. (1967). " Multidimensional Quantification of the Data Obtained by the Method of Paired Comparison." Annals Inst. Statist. Math., Vol. 19, 2: 363-365.
- HAYASHI, C. (1974a). "Minimum Dimension Analysis: One of the Methods of Multidimensional Quantification (MDQ)." Behaviormetrika, 1: 1-24.
- HAYASHI, C. (1974b). " Computer Program Series I of Multidimensional Quantification Methods." Institute of Statistical Mathematics, Japan. Computer Science Monographs, 4 : 1-77.
- HAYASHI, G. (1975). " Minimum Dimension Analysis, MDA-OR and MDA-UO." In Theory, Methods and Application of Multidimensional Scaling et Related Techniques. U.S. Japan Seminar, San Diego, USA.
- HAYASHI, C. et F. HAYASHI (1978). " Comparison of Two Types of Multidimensionnal Scaling Methods. Minimum Dimension Analysis MDA-OR and MDA-UO." Ann. Inst. Statist. Math., 30, part B: 199-209.
- HEISER, W.J. (1981). " Unfolding Analysis of Proximity Data." Ph. D. Thesis, Univ. of Leiden.
- HEISER, W.J. (1985). " Undesired Nonlinearities in Nonlinear Multivariate Analysis." In E. Diday (ed.), Data Analysis and Informatics. North Holland. Amsterdam .
- HEISER, W.J. (1985b). " Multidimensional Scaling by Optimizing Goodness of Fit to a Smooth Hypothesis. " Research Report RR-07-85, Dpt. of Data Theory, Univ. of Leiden.
- HEISER, W.J., de LEEUW, J. (1981). " Multidimensional Mapping of Preference Data." Math.et Sci. Humaines 19 : 39-96.
- HEISER, W.J., MEULMAN, J. (1983). " Analyzing Rectangular Tables by Joint and Constrained Multidimensional Scaling." J. of Econometrics 22 : 139-167.
- HEISER, W.J., MEULMAN, J. (1984). " Constrained Multidimensional Scaling: More Directions Than Dimensions." In proceedings. COMPSTAT 1984: 137-142.
- HEISER, W.J., STOOP, I. (1986). " Explicit Smacof Algorithms for Individual Differences Scaling." Research Report RR-86-14.
- HENDERSON, H.V. et S.R. SEARLE. (1979). " Vec and Vech Operators for Matrices with Some Uses in Jacobian and Multivariate Statistics." La R. Canadienne Statist. Vol 7, No. 1 : 65-81.
- HERMITE, C. (1850). " Sur Differents Objets de la Théorie des Nombres." J. fur die Reine und Angewandte Math. 40: 261-315.
- HILL, M.O., GAUCH, H.G. (1980). " Detrended Correspondence Analysis, An Improved Ordination Method." Vegetatio 42: 47-58.
- HIROTSU, C. (1978). " Ordered Alternatives for Interaction Effects." Biometrika Vol. 65, 3: 561-570.

- HOAGLIN, D.C., WELSCH, R.E. (1978). "The Hat Matrix in Regression and ANOVA." The American Statistician 32, No 1: 17-22.
- HOERL, A.E. et R.W KENNARD. (1970). "Ridge Regression : Biased Estimation for non orthogonal Problems." Technometrics, 12 : 55-68
- HOLLAND, J.L. (1973). Making Individual Choices: A Theory of Careers. Englewood Cliffs. New Jersey : Prentice Hall.
- HOLLANDER, M. et J. SETHURAMAN (1978). "Testing for Agreement Between Two Groups of Judges." Biometrika 65 : 403-411.
- HOLMAN, E.W. (1972). "The Relation Between Hierarchical and Euclidean Models for Psychological Distances." Psychometrika 37: 417-423.
- HUBER, P.J. (1985). "Projection Pursuit." (With Discussion). Ann. Statist. 13: 435-525.
- HUBERT, L.J. (1977). "Nominal scale agreements as a generalized correlation." Brit. J. Math. Stat. Psych., 30 : 98-103.
- HUBERT, L.J. (1982). "Analysis of Variance Procedures based on a Proximity Measure Between Subjects." Psychological Bulletin 91, No 2: 424-430.
- HUBERT, L.J. (1983). "Inference Procedures for the Evaluation and Comparison of Proximity Matrices." In J. Felsenstein (ed.), Numerical Taxonomy. Springer Verlag : New-York. pp. 209-228.
- HUBERT, L.J. (1985). "Combinatorial Data Analysis : Association and Partial Association." Psychometrika 50 : 449-467.
- HUBERT, L.J. et M.J. SUBKOVIK (1978). "Confirmatory Analysis and Geometric Models." Psychological Bulletin 86, No 2: 361-370.
- HUBERT, L.J. et P. ARABIE (1985). "Unidimensional Scaling and Combinatorial Optimization". In de LEEUW et al. (eds), Multidimensional Data Analysis. (1986) DSWO Press. Leiden. pp 181-196.
- IGLESIAS, A. (1975). Contribution à l'Analyse des Tableaux de Distance et de Similarité. Application à la Biologie. Thèse de 3-ième cycle. Univ. Claude Bernard. Lyon.
- INDOWS, T. (1975). "An Application of Multidimensional Scaling to Study of Binocular Visual Space." Paper presented to the U.S. Japan Seminar : Theory, Methods and Applications of Multidimensional Scaling and Related Techniques. San Diego USA.
- INDOWS, T. (1979). "Alleys in Visual Space." J. of Mathematical Psychology Vol 19, No. 3: 221-258.
- ISAAC, P.D. et POOR, D.D.S. (1974). "On the Determination of Appropriate Dimensionality in Data with Error." Psychometrika, 39: 91-109.
- IWATSUBO, S. (1983). "The Analytical Solutions of Eigenvalue Problem in the Case of Applying Optimal Scoring Method to Some Types of Data." In E. DIDAY, (ed.), INRIA, Rocquencourt: 24-33.
- JAMES, G.S. (1961). "Contribution to Discussion of the Paper of D.J. BARTHOLOMEW "A Test of Homogeneity of Means Under Restrictd Alternatives."." J. Royal Statist. Soc. B, 23: 279-280.
- JAMBU, M. et M.-O. LEBEAUX. (1978). Classification automatique pour l'analyse des données. Volumes 1 et 2. Dunod. Paris.
- JOHNSON, S.C. (1967). "Hierarchical Clustering Schemes." Psychometrika, 32 : 241-254.
- JOHNSON, S.C. (1973). "Pairwise Non Metric Multidimensional Scaling." Psychometrika, 38 : 11-18.
- JOHNSON, S.C. (1975). "A Simple Method for Pairwise Monotone Regression." Psychometrika, 40 : 163-168.
- JOHNSON, D.E. et F.A. GRAYBILL ( 1972 a). "Estimation of  $\sigma^2$  in a Two-Way Classification Models with Interaction and No Replication." J. Amer. Statist. Assoc. 67: 862-868.
- JOHNSON, D.E. et F.A. GRAYBILL ( 1972 b). "The Analysis of Two-Way Model with Interaction." J. Amer. Statist. Assoc. 67: 388-394.
- JONES, L.E. et F.W. YOUNG. (1972). "Structure of the Social Environment : Longitudinal Individual Differences Scaling of an Intact Group." J. of Personality and Social Psychology 24 : 108-121.

- JORESOG, K.G. (1970). "A general method for analysing covariance structures." Biometrika 57: 239-251.
- KAISER, H.F. (1963). "Image Analysis." In W.C. HARRIS (Ed.), Problems in Measuring Change. Univ. of Wisconsin Press. Madison. pp. 156-166.
- KAMENSKI, V.S. (1977). "Methods and Models of Nonmetric Multidimensional Scaling (Survey)." Avtomatika i Telemekhanika, No 8 (translated): 118-156.
- KAMENSKI, V.S. (1979a). "A Matrix Approach of Nonmetric Multidimensional Scaling." Avtomatika i Telemekhanika, No 1 (translated): 122-131.
- KAMENSKI, V.S. (1979b). "Pairwise Nonmetric Multidimensional Scaling Method in Socioeconomic Problems." Avtomatika i Telemekhanika, No 5 (translated): 127-136.
- KATO, T. (1976). Perturbation theory for linear operators. Springer Verlag : Berlin.
- KAZMIERCZAK, J.B. (1985 a) "Analyse Logarithmique. Deux Exemples d'Applications." Rev. Statist. Appli. Vol. 33, 1: 13-24.
- KAZMIERCZAK, J.B. (1985 b). "Une Application du Principe de Yule: l'Analyse Logarithmique." Proceedings of the fourth International Symposium: Data Analysis and Informatics. North Holland. pp 393-403.
- KELLER, J.B. (1962). "Factorization of Matrices by Least Squares." Biometrika 49: 239-242.
- KENDALL, D.G. (1971 a). "Construction of Maps from 'Odds' Bits of Information." Nature 231: 158-159.
- KENDALL, D.G. (1971 b). "Maps from Marriages: An Application of Nonmetric Multidimensional Scaling to Parish Register Data." Mathematics in Archeological and Historical Sciences. Edinburgh : Univ. of Edingourgh Press.
- KENDALL, M.G. (1962). Rank Correlation Methods (Third edition). Charles Griffin and Co.: London.
- KHATRI, C.G. (1966). "A Note on a MANOVA Model Applied to Problems in Growth Curve." Ann. Inst. Statist. Math. 18 : 75-86.
- KLAHR, D.A. (1969). "Monte Carlo Investigation of the Statistical Significance of Kruskal's Scaling Procedure." Psychometrika 34 : 319-330.
- KLINGBERG, F.L. (1941). "Studies in Measurement of the Relations Between Sovereign States." Psychometrika 6 : 335-352.
- KRANE, W.R. (1978). "Least Squares Estimation of Individual Differences in Multidimensional Scaling." British J. of Math. and Stat. Psychology 31, part 2: 193-208.
- KRANTZ, D.H. (1968). "A Survey of Measurement Theory.3 In G.B. DANTZIG et A.F. VEINOTT (Eds.), Mathematics of the Decision Sciences. Amer. Math. Soc.. Providence.
- KRANTZ, D.H. (1975). "Color Measurement and Color Theory : I. Representation Theorem for Grassman Structures." J. of Mathematical Psychology 12 : 283-303.
- KRANTZ, D.H., R.D. LUCE, P. SUPPES et A.TVERSKI (1971). Foundations of Measurement. Vol.I: Additive and Polynomial Representations. Academic Press. New York.
- KRISHNAIAH, P.R. et L.N. KANAL (eds.), (1982). Handbook of Statistics 2 : Classification, Pattern Recognition and Reduction of Dimensionality. North Holland.
- KRUMHANS, C.L. (1978). "Concerning the Applicability of Geometric Models to Similarity Data: the Interrelationship Between Similarity and Spatial Density." Psychological Review, 85, No. 5 : 445-463.
- KRUSKAL, J.B. (1964 a). "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." Psychometrika 29 : 1-27.
- KRUSKAL, J.B. (1964 b). "Nonmetric Multidimensional Scaling: a Numerical Method." Psychometrika 29 : 115-129.
- KRUSKAL, J.B. (1965). "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data." J. Roy. Statist. Soc. B27: 251-263.
- KRUSKAL, J.B. (1977a). "Multidimensional Scaling and Other Methods for Discovering Structures." In K. Enslein, A. Ralston et H. S. Wilf (eds.), Statistical Methods for Digital Computers. John Wiley. New York .
- KRUSKAL, J.B. (1977b). "The relationship between multidimensional scaling and clustering." In J. Van Rysin (Ed.), Classification and Clustering. Academic Press. New-York.

- KRUSKAL, J.B. et F. CARMONE. (1966). How to use M-D-SCAL (Version 5) and other useful informations. Murray Hill, NJ : Unpublished manuscript, Bell Laboratories.
- KRUSKAL, J.B. et L.D. CARROLL. (1969). " Geometric Models and Badness of fit Functions." In P.R. Krishnaiah (ed.), Multivariate Analysis, Vol 2. Academic Press. New-York.
- KRUSKAL, J.B. et M. WISH. (1978). " Multidimensional Scaling." SAGE publ. Beverly Hills.
- KRUSKAL, J.B., F.W YOUNG et J.B. SEERY. (1973). " How to Use KYST, a Very Flexible Program to do Multidimensional Scaling and Unfolding ." Tech. Rep. Bell Labs., Murray Hill, N.J. .
- KSHIRSAGAR, A.M. (1978). Multivariate Analysis. Marcel Dekker Inc. New-York.
- LAFFON, J.J. (ed.) (1979). Aggregation and Revelation of Preferences. North Holland.
- LALOUEL, J.M. (1977). " Linkage Mapping from Pairwise Recombination Data." Heredity 38: 61-67.
- LANCASTER, P. (1964). " On Eigenvalues of Matrices Dependent on a Parameter." Numerish Mathematik, 6 : 377-387.
- LANG, S. (1967). Algebra. Addison Wesley. Reading.
- LANTERMANN, E.D. et H. FEGER (eds.), (1980). Similarity and Choices. Hans Huber Pub.. Bern.
- LAWSON, C.L. et R.J. HANSON. (1974). Solving Least Square Problems. Englewood Cliffs. New Jersey. Prentice Hall.
- LE CALVE, G. (1976). Problèmes d'Analyse des Données. Thèse d'état. Université de Rennes I
- LE CALVE, G. (1985). " Distances à Centre." Statistique et Analyse des Données, Vol 10, No 2: 29-44.
- LE CALVE, G. (1987). "  $L_1$  Embeddings of a Data Structure (I,D)." in Y. DODGE (ed.) Statistical Data Analysis Based on the  $L_1$ -Norm and Related Methods. North-Holland. Amsterdam: pp. 195-202.
- LE CALVE, G. (1988). " Similarities Functions" COMPSTAT 1988 Physica Verlag, Heidelberg: pp. 341-347.
- de LEEUW, J. (1968). " Non Metric Multidimensional Scaling." Research Note RN 010-68. Dpt. of Data Theory, University of Leiden.
- de LEEUW, J. (1970a). " The Positive Orthant Method for non Metric Multidimensional Scaling." Research Note RN 001-70. Dpt. of Data Theory, University of Leiden.
- de LEEUW, J. (1970b). " The Euclidean Distance Model." Research Note RN 002-70. Dpt. of Data Theory, University of Leiden.
- de LEEUW, J. (1973a). "Bounds for SFORM1. Part I." Research Bulletin RB-001-73. Dpt. of Data Theory, University of Leiden.
- de LEEUW, J. (1973b). "Bounds for SFORM1. Part II." Research Bulletin RB-004-73. Dpt. of Data Theory, University of Leiden.
- de LEEUW, J. (1977a). " Applications of Convex Analysis to Multidimensional Scaling." Recent Developments in Statistics. J.R. BARRA et al. (eds.), North Holland. Amsterdam: 133-146.
- de LEEUW, J. (1977b). " Correctness of Kruskal's Algorithm for monotone regression with ties." Psychometrika, 42 : 141-144.
- de LEEUW, J. (1980). " Majorization Algorithms for Individual Differences in Multidimensional Scaling." Papier présenté au symposium Multidimensional Scaling and Interindividual Differences. at the 22nd International Congress of Psychology. Leipzig.
- de LEEUW, J. et I. STOOP. (1984). " Upper Bounds for Kruskal's STRESS." Psychometrika, 439: 391-402.
- de LEEUW, J. et W. HEISER. (1977). " Convergence of Correction Matrix Algorithm for Multidimensional Scaling with Restrictions on the Configuration." In J. LINGOES (ed.), Geometric Representation of Relational Data. Mathesis Press. Ann Arbor: pp 735-752.
- de LEEUW, J. et W. HEISER. (1982). " Theory of Multidimensional Scaling." In P.R. Krishnaiah et L. Kanal (Eds.), Handbook of Statistics, Vol. 2. North Holland. Amsterdam: pp. 285-316.
- LE FOLL, Y. (1972). " L'Analyse Factorielle des Evolutions." Actes du petit séminaire sur l'Analyse Factorielle, 2 : 23-27.

- LE FOLL, Y. (1979). Sur les Propriétés de l'Analyse des Correspondances pour Diverses Formes Complètes de Données. Thèse de 3-ième cycle. Université Pierre et Marie Curie. Paris.
- LE FOLL, Y. (1982). " Pondérations des Distances en Analyse Factorielle." Stat. et Anal. des Données, Vol. 7, No 1: 13-31.
- LEAMER, E.E. (1973). " Multicolinearity. A Bayesian Interpretation." Review of Economics and Statist., 55 : 371-380.
- LEAMER, E.E. (1978). Specification Searches. John Wiley. New York.
- LEBART, L. (1969). " Analyse Statistique de la Contiguïté." Publication de l'Inst. Stat. Univ. Paris Vol. XVIII : 81-112.
- LEBART, L. (1978). " Programme d'Agrégation avec Contraintes (C.A.H. contiguïté)." Les Cahiers de l'Analyse des Données, 3 : 275-287.
- LEBART, L. (1984). " Correspondence Analysis of Graph Structures." Bulletin Technique du CESIA vol 2 : 5-19.
- LECHEVALLIER, Y. (1980). " Classification Sous Contraintes." In E. DIDAY et al., (Eds.), Optimisation en Classification Automatique. INRIA, PARIS: 677-696.
- LEDERMANN; W. (ed.), (1984). Handbook of Applicable Mathematics. Vol VI, Part B, Statistics. John Wiley. New-York.
- LEE, J.C. , GEISSER, S. (1972). " Growth Curve Prediction." Sankhya A 34 : 393-412.
- LEE, J.C. , GEISSER, S. (1975). " Applications of Growth Curve Prediction." Sankhya A 37: 239-256.
- LEE, S.Y. (1984). " Multidimensional Scaling Models with Inequality and Equality Constraints." Comm. in Statistics (B) 13: 127-140.
- LEE, S.Y., BENTLER, P.M. (1980). " Functional Relations in Multidimensional Scaling." J. of Mathematical and Statist. Psychology, 33: 142-150.
- LEFKOVITCH, L.P. (1980). " Conditional Clustering." Biometrics, 36: 91-110.
- LEHMAN, E.I. (1975). Nonparametric Statistical Methods based on Ranks. Holden Day, Inc.: San Francisco.
- LEMAIRE, J. (1976). Aggrégation Typologique des Préférences. Thèse de 3ème Cycle. Nice.
- LERMAN, I.C. (1973). " Etude Distributionnelle de Statistiques de Proximité entre Structures de Même Type : Application à la Classification Automatique." Cahier du BURQ, 19. ISUP, Paris.
- LEURGANS, S. (1982). " Asymptotic Distributions of Slope-of-Greatest-Convex-Minorant Estimators." Ann. Statist. 10 : 287-296.
- LEVINE, D.M. (1978). A Monte Carlo Study of Kruskal's Variance Based Measure on Stress." Psychometrika, 43: 307-315.
- LEW, J.S. (1975). " Preorder Relations and Pseudoconvex Metrics." Amer. J. Math. 97: 344-363.
- LEW, J.S. (1978). " Some Counterexamples in Multidimensionnal Scaling." J. Math. Psychol. 17: 247-254.
- LINDMAN, H. et T. CAELLI. (1978). " Constant Curvature Riemannian Scaling." J. of Mathematical Psychology, 17: 89-109.
- LINGOES, J.C. (1971). " Some Boundary Conditions for the Monotone Analysis of Symmetric Matrices." Psychometrika, 36 : 195-203.
- LINGOES, J.C. (1973). The Guttman-Lingoes Nonmetric Program Series. Mathesis Press, Ann Arbor.
- LINGOES, J.C. (1980). " Testing Regional Hypotheses in Multidimensional Scaling." Proceedings of the Second International Symposium: Data Analysis and Informatics. North Holland. Amsterdam. pp 191-208.
- LINGOES, J.C. et E.E. ROSKAM. (1973). " A Mathematical and Empirical Analysis of two Multidimensional Scaling Algorithms." Psychometrika Monograph Supplement, 38 : 1-93.
- LINGOES, J.C. et I. BORG (1979). " A Quasi-Statistical Model for Choosing Between Alternative Configurations Derived From Ordinally Constrained Data." The Brit. J. of Math. Statist. Psycho., 36 : 36-53.
- LINGOES, J.C., ROSKAM, E.E. et I. BORG (eds.) (1979). Geometric representations of relational data. (2-ième édition). Mathesis Press, Ann Arbor.

- LITTKENS, E. (1966). "On the Fixed Point Property of WOLD's Iterative Estimation Method for Principal Components." In P.R. KRISHNAIAH (ed.), Multivariate Analysis II: pp. 335-350.
- LUENBERGER, D.G. (1969). Optimization by Vector Space Methods. John Wiley. New-York.
- LUNEBURG, R.K. (1947). Mathematical Analysis of Binocular Vision. Princeton Univ. Press, Princeton.
- LUNEBURG, R.K. (1950). "The Metric of Binocular Visual Space." J. of the Optical Society of America, 40: 627-642.
- MacCALLUM, R.C. (1974). "Relations between Factor Analysis and Multidimensional Scaling." Psychological Bulletin, 81: 505-516.
- MacCALLUM, R.C. (1979). "3-MODE-MDS: A Computer Program to Execute Tucker's Three-Mode Multidimensional Scaling." Applied Psychological Measurement, Vol. 5, 3: 377-382.
- MacKAY, D.B. et J.L. ZINNES. (1981). "Probabilistic Scaling of Spatial Judgments." Geographical Analysis, Vol 19, No. 1 : 21-37.
- MAGNUS, J.R. (1988). Linear Structures. Charles Griffin and Co. London.
- MAGNUS, J.R. et H. NEUDECKER. (1988). Matrix Differential Calculus With Applications in Statistics and Econometrics. John Wiley. New York.
- MAILLES, J.P. (1979). Analyse Factorielle des Tableaux de Dissimilarité. Thèse de 3ème cycle. Univ. Pierre et Marie Curie, Paris.
- MANDEL, J. (1961). "Non-Additivity in Two-Way Analysis of Variance." J. of the Amer. Statist. Assoc. 56: 878-888.
- MANDEL, J. (1971). "A New Analysis of Variance Model for Non-Additive Data." Technometrics 13: 1-18.
- MARCHALL, A.W. et I. OLKIN. (1979). Inequalities Theory of Majorization and its Applications. Academic Press. New York.
- MARDIA, K.V. (1978a). "Mahalanobis Distances and Angles." In P.R. Krishnaiah (ed.), Multivariate Analysis, Vol. 4. North Holland. Amsterdam.
- MARDIA, K.V. (1978b). "Some Properties of Classical Multidimensional Scaling." Commun. Statist. A. Theory and Methods, A7 : 1233 -1241.
- MARDIA, K.V., KENT, J.T. et J. M. BIBBY. (1979). Multivariate Analysis. Academic Press. New-York.
- MATHAR, R. (1985). "The Best Euclidean Fit to a given Distance Matrix in Prescribed Dimensions." Linear Algebra and its Applications, 67 : 1-6.
- McDONALD, R.P. (1968). "A Unified Treatment of the Weighting Problem." Psychometrika 33: 351-381.
- McDONALD, R.P. (1970). "Three Common Factor Models for Groups of Variables." Psychometrika 35 : 111-128.
- McDONALD, R.P. (1975). "Descriptive Axioms for Common Factor Theory. Image Theory and Component Theory." Psychometrika 40 : 137-152.
- McGEE, V.E. (1968). "Multidimensional Scaling of N Sets of Similarity Measures: A Nonmetric Individual Differences Approach." Multivariate Behavioral Research, 3 : 233-248.
- McGEE, V.E. (1969). "Speech Perception and the Development of Appropriate Nonmetric Analytical Techniques." Tech. Rep. Darmouth College. Hanover. 66p.
- MEDINA, J. (1985). Fonctions Splines avec Conditions de Forme. Thèse. Université Fourier, Grenoble.
- MENGER, K. (1928). "Untersuchungen über Allgemeine Metrik." Math. Ann. 100: 75-163.
- MENGER, K. (1931). "New Foundations of Euclidean Geometry" Amer. J. of Math. 53 721-745.
- MEULMAN, J. (1983). "Stability of Constrained Multidimensional Scaling Solutions." Paper presented at the European Meeting of the Psychometric Society. Jouy en Josas, France.
- MEULMAN, J. (1986). A Distance Approach to Nonlinear Multivariate Analysis. DSWO Press. Leiden.
- MEULMAN, J. et J. de LEEUW. (1985). "Relations Between Constrained Multidimensional Scaling and Principal Component Analysis." Paper presented at the Fourth European Meeting of the Psychometric Society. Cambridge.

- MEULMAN, J. et W. HEISER. (1984). "Constrained Multidimensional Scaling: More Directions than Dimensions." COMPSTAT. Vienna : Physica Verlag.
- MICKO, H.C. (1970). "A Halo Model for Multidimensional Scaling." Psychometrika, 28 : 199-227.
- MINKOWSKI, H. (1891). "Über die Positiven Quadratischen Formen und über Kettenbruchähnliche Algorithmen." J. für die Reine und Angewandte Math. 107: 243-260.
- MIRKIN, B.G. (1979). Group Choice. John Wiley. New-York.
- MIRKIN, B.G. (1985). "Additive Clustering and Qualitative Factor Analysis Methods." Proceedings of the fourth International Symposium: Data Analysis and Informatics. North-Holland. pp 111-122.
- MIRKIN, B.G. (1987). "Additive Clustering and Qualitative Factor Analysis for Similarity Matrices." Journal of Classification 4 : 7-81.
- MIZERE, D. (1981). Analyse d'un Cube de Données: Décomposition Tensorielle et Liens entre Procédures de Comparaison de Tableaux Rectangulaires de Données. Thèse de 3ème cycle. Université de GRENOBLE I.
- MONLEZUN, C.J., SPEED, F.M. (1980). "The Geometry of Estimation and Hypothesis Testing in the Constrained Linear Model. The Full Rank Case." Comm. in Statistics (A) 9, 2: 213-230.
- MORGAN, C.L. (1974). "Embedding Metric Spaces in Euclidean Space." J. of Geometry. Vol. 5/1: 101-107.
- MÖBUS, C. (1976). "Nonmetric Multidimensional Scaling Without Disparities and Derivatives : a Rank Correlation Oriented Approach Through  $L_1$ -Approximation." Archiv für Psychologie, 128 : 240-266.
- MULAIK, S.A. (1972). The Foundations of Factor Analysis. McGraw-Hill. New-York.
- von NEUMANN, J. et I.J. SCHOENBERG. (1941). "Fourier Integral and Metric Geometry." Trans. Amer. Math. Soc. 50: 226-251.
- NISHISATO, S. (1978). Multidimensional Scaling : A historical Sketch and Bibliography. Dept. of Measurement, Evaluation and Computer applications. The Ontario Institute for Studies and Education. Toronto.
- NISHISATO, S. (1978b). "Optimal Scaling of Paired Comparison and Rank Order Data. An Alternative to Guttman's Formulation." Psychometrika, 43 : 463-467.
- NISHISATO, S. (1980). Analysis of Categorical Data : Dual Scaling and its Application. University of Toronto Press : Toronto.
- NOMA, E. et J. JOHNSON. (1979). "Constrained Nonmetric Multidimensional Scaling." Michigan Mathematical Psychology Program. Tech. Rep. No. 79/4. Ann Arbor. Michigan.
- OBENCHAIN, R.L. (1973). "Affine Multidimensional Scaling." Manuscript non publié. Bell Labs, Murray Hill.
- OKAMOTO, M. et J. TODA (1973). "Fourth Method of Quantification: an Principal Coordinate Analysis (with discussion)." (en Japonais) J. Japan. Statist. Soc., 3 : 41-66.
- OKAMOTO, M. et T. ISOGAI (1978). "Optimality of Multidimensional Representation in Hayashi's Fourth Method of Quantification." (en Japonais) J. Japan. Statist. Soc., 8: 63-69.
- ORTH, B. (1980). "On the Foundations of Multidimensional Scaling : An Alternative to Beals, Krantz and Tversky Approach." In E.D. LANTERMAN et H. FEJER (eds.), Similarity and Choices. Hans Huber Publ.. Bern: pp 54-69.
- OTTESTAD (1975). "Component Analysis an Alternative System." Int. Stat. Rev. Vol 43 N° 1: 83-108.
- PARTHASARATHY, K.R. et K. SCHMIDT (1972) Positive Definite Continuous Tensor Products, and Central Limit Theorems of Probability Theory. Lect. Notes in Math. No 272. Springer Verlag. Heidelberg.
- PETTITT, A.N. (1982). "Parametric Test For Agreement Amongst Groups of Judges." Biometrika 40 : 365-375.
- PIERI, M. (1908). "La Geometria Elementare Istituita Sulle Nozioni di 'Punto' e 'Sfera'." Memorie di Matematica e di Fisica della Societa Italiana delle Scienze. Ser.3, Vol. 15: 345-350.
- PIERRA, D. (1976). Méthodes de décomposition et Croisements d'Algorithmes pour des Problèmes d'Optimisation. Thèse de 3ème cycle. Univ. de GRENOBLE I.

- PIESZKO, H. (1975). " Multidimensional Scaling in Riemannian Space." J. of Mathematical Psychology 12 : 449-477.
- PLACKETT, R.L. (1981) The Analysis of Categorical Data. (2-nd ed.) Griffin. London.
- POLIT, M.O. (1986). Une  $\eta$ -ACP d'un Hypercube de données. Thèse de Troisième Cycle. Université de Grenoble II.
- POTGORNY, P. et W.R. GARNER. (1979). " Reaction Time as a Measure of Inter and Intra-Objet Visual Similarity: Letters of the Alphabet." Perception and Psychophysics 26 : 37-52.
- POTTHOFF, R., ROY, S.N.. (1964). " A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems." Biometrika 51: 313-326.
- PRENTICE, I.C. (1977). " Non Metric Ordination Methods in Ecology." J. of Ecology 65: 85-94.
- PRENTICE, I.C. (1980). " Vegetation Analysis and Order Invariant Gradient Models." Vegetatio 42: 27-34.
- QUINQUETON, J. (1981). " Le Concept de Dimension Intrinsèque en Reconnaissance des Formes." Thèse d'Etat, Univ.Pierre et Marie Curie, Paris VI.
- RAMSAY, J.O. (1969). " Some Statistical Considerations in Multidimensional Scaling." Psychometrika 34 : 167-182.
- RAMSAY, J.O. (1977). " Maximum Likelihood Estimation in Multidimensional Scaling." Psychometrika 42 : 241-266.
- RAMSAY, J.O. (1978a). " Confidence Regions for Multidimensional Scaling Analysis" Psychometrika 43 : 145-160.
- RAMSAY, J.O. (1978b). MULTISCALE : Four programs for multidimensional scaling by the method of maximum likelihood. International Education Service, Chicago.
- RAMSAY, J.O. (1980). " Some Small Sample Results for Maximum Likelihood Estimation in Multidimensional Scaling." Psychometrika 45 : 139-144.
- RAMSAY, J.O. (1982). " Some Statistical Approaches to Multidimensional Scaling data." J. Roy. Statist. Soc. A 145, part 3 : 285-311.
- RAO, C.R. (1955). " Estimation and Tests of Significance in Factor Analysis." Psychometrika 20 : 93-111.
- RAO, C.R. (1964). " The Use and Interpretation of Principal Component Analysis in Applied Research." Sankhya A 26 : 329-358.
- RAO, C.R. (1965). " The Theory of Least Squares When the Parameters are Stochastic and its Application to the Analysis of Growth Curves." Biometrika 52 : 447-458.
- RAO, C.R. (1966). " Covariance Adjustment and Related Problems in Multivariate Analysis." In P.R. KRISHNAIAH, (ed.). Multivariate Analysis . Academic Press, New York: pp.87-103.
- RAO, C.R. (1967). " Least Squares Theory Using an Estimated Dispersion Matrice and its Applications to Measurement of Signals. Proc. 5th Berkeley Symposium 1 : 355-372.
- RAO, C.R. (1969). " A Decomposition Theorem for Vector Variables with a Linear Structure." Ann. Math. Statist. 40 : 1845-1849.
- RAO, C.R. (1980). " Matrix Approximation and Reduction of Dimensionality in Multivariate Statistical Analysis. In P.R. Krishnaiah (Ed.), Multivariate Analysis, Vol. 5. North Holland. Amsterdam.
- RAO, C.R. et S.K. MITRA. (1971). Generalized Inverse of Matrices and its Applications. John Wiley. New-York.
- RESTLE, F. (1959). " A Metric and an Ordering on Sets." Psychometrika 41: 349-372.
- RICHARDSON, M.W. (1938) "Multidimensional Psychophysics." Psychol. Bull. 35: 659-660.
- ROBERT, F. (1967). " Calcul du Rapport Maximal de Deux Normes sur  $R^n$ ." RIRO 1 : 97-118.
- ROBERTSON, T., F.T. WRIGHT et R.L.DYKSTRA. (1988). Order Restricted Statistical Inference. John Wiley. New-York.
- ROCKAFELLAR, T. (1970). Convex Analysis. Princeton : University Press.
- ROSKAM, E.E. (1968). Metric Analysis Of Ordinal Data in Psychplogy : Models and Numerical Methods for Metric Analysis of Conjoint Ordial Data in Psychology. Van Voorshotten, Univ. of Leiden Press.
- ROSKAM, E.E. (1969a). A Comparison of Principles for Algorithm Construction in Nonmetric Scaling. MMPP-69-2, Ann Arbor.

- ROSKAM, E.E. (1969b). Data Theory and Algorithms for Nonmetric Scaling (Parts 1 and 2). Unpublished manuscript. Catholic Univ. of Nijmegen.
- ROSKAM, E.E. (1970). "MINISSA-I: A Fortran IV (G) Program for the Smallest Space Analysis of Square Symmetric Matrices." Behavioral Science, 15: 204-205.
- ROSKAM, E.E. (1970). "The Method of Triads for Nonmetric Multidimensional Scaling." Psychologie, 25: 404-417.
- ROSKAM, E.E. (1979). "A Survey of the Michigan-Israel-Netherlands-Integrated Serie." In J.C. Lingoes et al. (eds.), Geometrical Representations of Relational Data. Mathesis Press, Ann Arbor.
- SAITO, T. (1982). "Contributions to  $e_{ij}$ -type Quantification and Development of a New Method of Multidimensional Scaling." Behaviormetrika, No 12: 63-83.
- SAITO, T. (1983). "A Method of Multidimensional Scaling to Obtain a Sphere Configuration." Hokkaido Behavioral Science Report, Series M, No 4. Sapporo.
- SAITO, T. (1985). "Multidimensional Scaling to Represent Context Effect in Dissimilarity Judgment." Hokkaido Behavioral Science Report, Series M, No 7. Sapporo.
- SAITO, T. (1986). "A Method of Multidimensional Scaling to Explore Complex Aspects in Dissimilarity Judgment." Hokkaido Behavioral Science Report, Series M, No 8. Sapporo.
- SAITO, T. (1986). "Multidimensional Scaling to Explore Complex Aspects in Dissimilarity Judgment." Behaviormetrika, No 20: 35-62.
- SAITO, T., S. OGAWA et E. NOJIMA (1973). "A Summary Review Article on Multidimensional Scaling." Soken Kiyō, 3 (en japonais): 53-215.
- de SARBO, W.S., MOHAJAN, V. (1984). "Constrained Classification: The Use of a priori Informations in Cluster Analysis." Psychometrika 49: 187-215.
- SATTAH, S. et A. TVERSKY (1977). "Additive Similarity Trees". Psychometrika 42: 319-345.
- SCHAAFSMA, W. (1966) Hypothesis Testing Problems with Alternatives Restricted by a Number of Inequalities. Noodhof. Groningen.
- SCHAAFSMA, W. et L.J. SMID (1966). "Most Stringent Somewhere Most Powerful Test Against Alternatives Restricted by a Number of Linear Inequalities." Ann. Math. Statist. 37: 1161-1172.
- SCHIFFMAN, S.S., M.L. REYNOLDS et F.W. YOUNG (1981). Introduction to Multidimensional Scaling: Theory, Methods and Applications. Academic Press. New York.
- SCHOENBERG, I.J. (1935). "Remarks to Maurice FRECHET's Article " Sur la Définition Axiomatique d'une Classe d'Espaces Distanciés Vectoriellement Applicables sur l'Espace de Hilbert." Annals of Mathematics, 36: 724-732.
- SCHOENBERG, I.J. (1937). "On Certain Metric Spaces from Euclidean Spaces by a Change of Metric and their Imbedding in Hilbert Space." Ann. of Math. 38: 787-793.
- SCHOENBERG, I.J. (1938 a). "Metric Spaces and Completely Monotone Functions." Ann of Math., 39: 811-841.
- SCHOENBERG, I.J. (1938 b). "Metric Spaces and Positive Definite Functions." Transac. Amer. Math. Soc., 41: 522-536.
- SCHÖNEMANN, P.H. (1972). "An Algebraic Solution for a Class of Subjective Metrics Models." Psychometrika 37: 441-451.
- SCHÖNEMANN, P.H. et J.H. STEIGER. (1976). "Regression Component Analysis." British J. of Math. and Statist. Psychology 29: 175-189.
- SCOTT, D. (1956). "A Symmetric Primitive of Euclidean Geometry." Indagationes Math. 18: 456-461.
- SEN, A.K. (1970). Collective Choice and Social Welfare. Mathematical Economics Texts. Holden Day Inc.
- SHEPARD, R.N. (1957). "Stimulus and Response Generalization: A Stochastic Model Relating Generalization to Distance in Psychological Space." Psychometrika 22: 325-345.
- SHEPARD, R.N. (1958 a). "Stimulus and Response Generalization: A Stochastic Model Relating Generalization to Distance in Psychological Space." J. Exp. Psychol. 55: 509-523.
- SHEPARD, R.N. (1958 b). "Stimulus and Response Generalization: Deduction of the Generalization Gradient from a Trace Model." Psychol. Rev. 65: 242-256.

- SHEPARD, R.N. (1962 a). " The Analysis of Proximities. Multidimensional Scaling with Unknown Distance Function." Psychometrika 27 : 125-140.
- SHEPARD, R.N. (1962 b). " The Analysis of Proximities. Multidimensional Scaling with Unknown Distance Function." Psychometrika 27 : 219-246.
- SHEPARD, R.N. (1974). " Representation of Structure in Similarity Data: Problems and Prospects." Psychometrika 39 : 373-421.
- SHEPARD, R.N. (1980). " Multidimensional Scaling, Tree Fitting and Clustering." Science, Vol. 210: 390-398.
- SHEPARD, R.N., A.K. ROMNEY et S.B. NERLOVE (1972). Multidimensional Scaling: Theory and Applications in the Behavioral Sciences. Vol 1 et 2. Seminar Press. New York.
- SHERMAN, C.R. (1972). " Nonmetric Multidimensional Scaling: A Monte Carlo Study of the Basic Parameters." Psychometrika 37 : 323-355.
- SJOBERG, L.A. (1972). "A cognitive Theory of Similarity." Göteborg Psychological Reports No 10.
- de SOETE, G. (1983). " A Least Squares Algorithm for Fitting Additive Trees to Proximity Data." Psychometrika 48 : 621-626.
- de SOETE, G. (1984). "Additive Tree Representations for Incomplete Dissimilarity Data." Quality and Quantity, 18: 387-393.
- de SOETE, G., CARROLL, J.D. et W.S. DE SARBO, (1986). " Alternating Least Squares Optimal Variable Weighting Algorithms for Ultrametric and Additive Tree Representation." In W. Gaul and M. Schader (eds.), Classification as a Tool of Research. Elsevier North Holland. Amsterdam: pp. 97-103.
- de SOETE, G., DE SARBO, W.S., FURNAS, G.W. et J.D. CARROLL (1984). " The Estimation of the Ultrametric and Path Length Trees from Rectangular Proximity Data." Psychometrika 49 : 289-310.
- SPENCE, I. (1972). " A Monte Carlo Evaluation of Three Nonmetric Multidimensional Scaling Algorithms." Psychometrika, 37: 461-486.
- SPENCE, I. (1974). " On Random Rankings Studies in Nonmetric Scaling." Psychometrika, 39: 267-268.
- SPENCE, I. et GRAEF, J. (1973). " How to Use M-SPACE, a Program for the Determination of the Underlying Dimensionality of an Empirically Obtained Matrix of Proximities." Univ. of Western Ontario Research Bulletin, No 257.
- SPENCE, I. et GRAEF, J. (1974). " The Determination of the Underlying Dimensionality of an Empirically Obtained Matrix of Proximities." Multivariate Behavioral research, 9: 331-342.
- SPENCE, I. et J.C. OGILVIE. (1973). " A Table of Expected Stress Values for Random Rankings in Nonmetric Multidimensional Scaling." Multivariate Behavioral research, 8: 511-517.
- SPENCE, I. et F.W. YOUNG (1978). " Monte Carlo Studies in Nonmetric Scaling." Psychometrika, 43: 115-117.
- STEIN, C.M. (1956). " Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." Proc. third Berkeley Symp. on Math. Statist. and Proba. 1: pp 197-206. Univ. of California Press : Berveley.
- STENSON, H.H et R.L. KNOLL. (1969). " Goodness of Fit for Random Rankings in Kruskal's Nonmetric Scaling Procedure." Psychol. Bulletin, 72: 122-126.
- STUMPF, C. (1880). Tonpsychologie, Vol. I et II. Teubner. Leipzig.
- SZMIELEW, W. (1983). From Affine to Euclidean Geometry. D. Reidel Pub. Dordrecht.
- TAKANE, Y. (1977 a). " Statistical Procedures for Non Metric Multidimensional Scaling." Ph. D. Thesis. Univ. of North Carolina, Chapel Hill. Univ. Microfilm International 78-7169.
- TAKANE, Y. (1977 b). " On the Relations Among Four Methods of Multidimensional Scaling." Behaviormetrika 4: 29-43.
- TAKANE, Y. (1978). " A Maximum Likelihood Method of Multidimensional Scaling: I The Case in Which all Pairwise Orderings are Independent - Theory." Japanese Psych. Research Vol. 20, 1: 7-17.
- TAKANE, Y. (1980). " Analysis of Categorizing Behavior by a Quantification Method." Behaviormetrika 8: 75-88.
- TAKANE, Y. (1981). " Statistical Procedures for Non Metric Multidimensional Scaling." J. of Ecology 69: 135-152.

- TAKANE, Y. et J. SERGENT. (1983). " Multidimensional Scaling Models for Reaction Times and Some Different Models." Psychometrika 48 : 393-423.
- TAKANE, Y., F.W. YOUNG et J. de LEEUW (1977). " Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features." Psychometrika 42 : 7-68.
- TAKEUCHI, K., YANAI, H. et B.N. MUKHERJEE. (1982). The Foundations of Multivariate Analysis. John Wiley Eastern Limited: New Dehli.
- TAUSSKI, O. (1949). " A Recurring Theorem on Determinants." Amer. Math. Monthly, 56 : 672-676.
- TENENHAUS, M. (1977). " Analyse en Composantes Principales d'un Ensemble de Variables Nominales et Numériques." Rev. Statist. Appliq., 25: 39-56.
- TENENHAUS, M. (1983). " L'Analyse des Données Qualitatives par des Méthodes de Codage Optimal." Cahier de Recherche 229/1983 CESA. Jouy-en-Josas.
- TENENHAUS, M. (1987). " Canonical Analysis of Two Convex Polyhedral Cones and Applications." Soumis à publication: 40p. .
- TER BRAAK, C.J.F. (1985). " Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model." Biometrics 49 : 347-358.
- TER BRAAK, C.J.F. (1986). " Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis." Ecology 67 (5): 1167-1179.
- TEREKHINA, A.Y. (1973). " Methods of Nonmetric Multidimensional Scaling and Visualization. (Survey)" Avtomatika i Telemekhanika, No 7 (translated): 80-94.
- TISSERON, C. (1983). Géométries Affines Projectives et Euclidiennes. Hermann. Paris.
- TORGERSON, W.S. (1952). " Multidimensional Scaling: I. Theory and Methods." Psychometrika, 17: 401-419.
- TORGERSON, W.S. (1958). Theory and Methods of Scaling. John Wiley. New York.
- TORGERSON, W.S. (1965). " Multidimensional Scaling of Similarity." Psychometrika, 30: 379-393.
- TUKEY, J.W. (1949). " One Degree of Freedom for Non-Additivity." Biometrics Vol. 5, 3: 232-242.
- TVERSKY, A. (1977). " Features of Similarity." Psychological Review, 84: 327-352.
- TVERSKY, A. et D.H. KRANTZ. (1970). " The Dimensional Representation and Metric Structure of Similarity Data." J. of Math. Psychology 7 : 572-596.
- VELLEMAN, P.F., WELSCH, R.E. (1981). " Efficient Computing of Regression Diagnostics." The Amer. Statistician 35 , No 4: 234-242.
- VINOD, H.D. (1978). " Survey of Ridge Regression related Techniques for Improvements Over Ordinary Least Squares." Review of Economics and Statistics, 60 : 121-131.
- WAERN, Y. (1970). " A Model for Multidimensional Similarity." Perceptual and Motor Skills, 33: 15-25.
- WAGENAAR, W.A. et PADMOS, P. (1971). " Quantitative Interpretation of Stress in Kruskal's Multidimensional Scaling Technique." British J. of Math. and Statist. Psychology, 24: 101-110.
- WALKER, J.J. (1976). " Some Statistical Procedures Based on Distances." Institute of Statistics Mimeo Series No. 1096. Univ. of North Carolina, Chapel Hill, USA.
- WEEKS, D.G. et P.M. BENTLER. (1982). " Restricted Multidimensional Scaling Models for Asymmetric Proximities." Psychometrika 47 : 201-208.
- WELLS, J.H. et L.R. WILLIAMS (1975). Embeddings and Extensions in Analysis. Springer Verlag. Heidelberg.
- WHITTAKER, R.H. (1967). " Gradient Analysis of Vegetation." Biological Review 42: 207-264.
- WHITTLE, P. (1971). Optimization under Constraints. John Wiley. New-York.
- WILSON, W.A. (1931) " On Semimetric Spaces." Amer. J. Math. 53: 361-373.
- WILSON, W.A. (1932) " A Relation Between Metric and Euclidean Spaces." Amer. J. Math. 54: 505-517.
- WILSON, W.A. (1931) " On the Embedding of Metric Sets in Euclidean Spaces." Amer. J. Math. 57: 322-326.
- WINSBERG, S. et J.D. CARROLL. (1984). " A Nonmetric Method for a Multidimensional Scaling Model Postulating Common and Specific Dimensions." Paper presented at the Psychometric Society Meeting.

- WINSBERG, S. et J.O.RAMSAY (1983). " Monotone Spline Transformations for Dimension Reduction." Psychometrika, 48 : 575-595.
- WOLD, H. (1966). " Estimation of Principal Components and Related Models by Iterative Least Squares." In P.R. KRISHNAIAH (ed.), Multivariate Analysis II: pp. 391-420.
- WOLD, H. et E. LYTTKENS (1969). " Non Linear Partial Least Squares (NIPALS) Estimation Procedures." Bull. ISI Vol 43: 29-51.
- WOLFRUM, C. (1976 a) " Zum Auftreten Quasiäquivalenter Lösungen bei einer Verallgemeinerung des Skalierungsverfahrens von KRUSKAL auf Metrische Räume mit einer Minkowski-Metrik." Archiv für Psychol., Bd. 128, s : 96-11.
- WOLFRUM, C. (1976 b) " Zur Bestimmung eines Optimalen Metrikkoeffizienten  $r$  mit dem Skalierungsverfahren von KRUSKAL." Zeitschrift für Experimentelle und Angewandte Psychol., Bd. XXIII, Heft 2, s.: 339-350.
- YATES, A. (1972). " New Metric Individual Differences Multidimensional Scaling With Balanced Least Squares Monotonic Regression." Communication à la Conférence of the Psychometric Society, Princeton.
- YOSHIDA, K. (1965). Functional Analysis. Springer Verlag : Berlin.
- YOUNG, F.W. (1968). " TORSCA-9: A Fortran IV program for nonmetric Multidimensional Scaling." Behavioral Science, 13: 343-344.
- YOUNG, F.W. (1975 a). " An Asymmetric Euclidean Model for Multi-process Asymmetric Data." In Theory, Methods and Application of Multidimensional Scaling et Related Techniques. U.S. Japan Seminar, San Diego, USA.
- YOUNG, F.W. (1975 b). " Methods for Describing Ordinal Data with Cardinal Methods." J. of Mathematical Psychology, 12 : 416-436.
- YOUNG, F.W. (1980). " Nonmetric Multidimensional Scaling under SAS." In Proc. of the fifth annual SAS users group international conference. CARY, NC : SAS Institute INC.
- YOUNG, F.W. et R. LEWYCKYJ (1979). ALSCAL 4 User's guide. (2nd edition). Data Analysis and Theory Associates, Chapel Hill.
- YOUNG, F.W. et M.S. TORGERSON. (1967). " TORSCA, A Fortran IV Program for the Shepard-Kruskal Multidimensional Scaling Analysis." Behavioral Science, 12 : 498.
- YOUNG, F.W., J. de LEEUW et Y. TAKANE (1976). " Quantifying Qualitative Data." In LANTERMAN et FEGER (eds.), Similarity and Choice. Huber: Bern.
- YOUNG, F.W., J. de LEEUW et Y. TAKANE (1976). " The Principal Components of Mixed Measurement Level Multivariate Data: An Alternating Least Squares Method with Optimal Scaling Features." Psychometrika 43 : 279-281.
- YOUNG, F.W., R.M. HAMER et J.R. LEWYCKYJ. (1980). " Non Metric Multidimensional Scaling." In Proc. of the fifth annual SAS users group international conference, Cary, N.C.
- YOUNG, F.W., Y. TAKANE, et J. de LEEUW (1978). " Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features." Psychometrika 41 : 505-529.
- YOUNG, G. et A.S. HOUSEHOLDER. (1938). " Discussion of Set of Points in Terms of their Mutual Distances." Psychometrika 3 : 19-22.
- YULE, G.U. (1912). " On the Methods of Measuring Association Between Two Attributes." J. Roy. Statist. Soc. 75: 579-642.
- ZELLNER, A. (1971). An introduction to Bayesian Inference in Econometrics. John Wiley. New York.
- ZINNES, J.L. et R.A. GRIGGS (1974). " Probabilistic Multidimensional Unfolding Analysis." Psychometrika 39 : 30-50.
- ZINNES, J.L. et D.B. McKAY (1981). " Multidimensional Scaling Models: The Other Side." In I. BORG (ed.), Multidimensional Data Representations: When and Why. Mathesis Press. Ann Arbor .
- ZINNES, J.L. et R.D. WOLFF (1977). " Single and Multidimensional Same Different Judgments." J. of Mathematical Psychology 16 : 30-50.



**PARTIE B**

**UN MODELE ALGEBRIQUE**

**POUR**

**L'ANALYSE DES DONNEES DE DISSIMILARITE**

**SOUS CONTRAINTES**

- Chapitre I**            **INTRODUCTION**
- Chapitre II**        **L'ANALYSE DU TRIPLE ( I,  $\delta_I$ ,  $Q_{I \times I}$  )**  
**SANS CONTRAINTES**
- Chapitre III**      **L'ANALYSE DU TRIPLE ( I,  $\delta_I$ ,  $Q_{I \times I}$  ) SOUS**  
**CONTRAINTES DE CONFIGURATION**
- Chapitre IV**      **L'ANALYSE DU TRIPLE ( I,  $\delta_I$ ,  $Q_{I \times I}$  )**  
**SOUS CONTRAINTES RELATIONNELLES**



## CHAPITRE I

### INTRODUCTION

Cet exposé est consacré à l'étude des méthodes d'analyse des données de dissimilarité sous contraintes. L'information habituellement traitée en analyse des dissimilarités ne suppose pas de contraintes. Elle est fournie par un tableau carré dont le terme générique résulte de l'observation d'un indice de similarité (ou de dissimilarité), traduisant une relation de ressemblance (resp. de dissemblance) entre unités statistiques. On dispose donc d'un système relationnel empirique simple, dont on se propose de construire un plongement homéomorphe dans un système relationnel numérique. On construit pour cela un codage des données à l'aide de méthodes de représentation des unités statistiques dans un espace métrique de dimension  $r$ , le plus souvent Euclidien. On parle alors de méthode de Codage Multidimensionnel lorsque  $r > 1$ .

Nées des travaux de SHEPARD (1962), dans le domaine de la psychophysique, les méthodes de Codage Multidimensionnel, ou Multidimensional Scaling, ont rapidement été appliquées dans l'ensemble des domaines de recherche en Psychologie et en Marketing (SHEPARD et al. (1972), GREEN ET CARMONE (1970)), en Politologie (FORGAS et al. (1977)), en Sociologie (COXON et JONES (1974), FUNIA et al. (1976), JONES et YOUNG (1972)), en Géographie (HOLLAND (1973), COLLEDGE et RUSHTON (1972)), en Cartographie (GILBERT (1974)) et en Archéologie (KENDALL (1971)), etc...

On retrouve de plus des applications de ce type de méthodes dans le domaine des sciences de la nature, comme en biochimie (CRIPPEN (1977, 1978)) ou en écologie (FASHAM (1977), GAUCH (1982), GAUCH et al. (1977,1981), HANHAM (1976), PRENTICE (1977,1980), WHITTAKER (1967)) et plus généralement en Biologie, où un fort courant de recherche s'est développé à propos de ces "Ordination Methods". Ce nom quelque peu trompeur fut en fait donné par extension au cas multidimensionnel aux méthodes de représentation des unités statistiques sur un continuum. Dans le cas unidimensionnel en effet un ordre est naturellement associé à ce type de représentations.

SHEPARD (1974) précise que son objectif initial était de mettre à l'épreuve de l'expérience une théorie et de dégager par une représentation graphique une structure spatiale latente qui en résultait. De nombreux développements axiomatiques portant sur la modélisation métrique des jugements subjectifs de dissimilarité suivirent les travaux de Shepard. Développés dans le cadre de la théorie du mesurage (BEALS et al. (1967, 1968), ORTH (1980), TVERSKY et KRANTZ (1970)) ces travaux sont cependant actuellement difficilement utilisables en analyse des données parcequ'ils imposent des conditions de densité de l'espace des unités statistiques impossibles à contrôler sur un échantillon de taille finie. Cependant, l'élaboration récente de modèles probabilistes exploitables (RAMSAY (1969, 1977, 1978, 1980, 1982), TAKANE (1977, 1981), TAKANE et SERGENT (1983), ZINNES et al. (1974, 1977, 1981)) permettent désormais de mettre en œuvre des méthodes statistiques d'analyse des données de dissimilarité de type confirmatoire. Bien que prometteuse, l'approche confirmatoire a été rarement mise en application jusqu'à présent du fait d'une complexité de mise en œuvre numérique certaine et de l'importance des expérimentations nécessaires à la construction d'un échantillon de tableaux de dissimilarité. Elle doit être distinguée de l'objet de notre étude, qui s'inscrit dans une optique exploratoire et n'impose pas de disposer d'un échantillon de matrices de dissimilarité issues d'une expérience.

Nous nous intéressons ici au cas des données de simple observation. de LEEUW et HEISER (1980) distinguent alors trois types d'applications en fonction des connaissances a priori de l'analyste. La situation la plus simple à traiter est rencontrée principalement dans le domaine des sciences de la nature. CRIPPEN (1977) en fournit un exemple en biochimie, à propos du positionnement spatial des molécules. On sait alors qu'il existe une structure spatiale Euclidienne tridimensionnelle des molécules et l'on se propose de la comparer à la configuration "optimale" obtenue par un algorithme de Codage Multidimensionnel appliqué à des données observées parfois incomplètes et le plus souvent entâchées d'erreurs de mesure.

Cette situation est distincte du cadre d'application courant en géographie (GILBERT (1974), KENDALL (1971)) et dans certaines expériences de psychologie perceptive (EKMAN (1954), INDOVS (1975, 1979)). Le caractère Euclidien de la structure spatiale sous-jacente n'est alors plus assuré et l'on construit une représentation Euclidienne ou plus généralement dans un espace de Minkowski de type  $I_p$ , approchant au mieux la structure géométrique de référence. Le choix d'une représentation Euclidienne des données présente l'avantage de la simplicité, en particulier dans la phase d'interprétation des résultats. Il est cependant trop simplificateur pour un certain nombre d'applications, telles que l'étude des perceptions visuelles (LUNEBURG (1947, 1950), INDOVS (1975, 1979), DRÖSLER (1981a)) qui nécessite le recours à des méthodes de Codage Multidimensionnel Riemanien (PIESZKO (1975) LINDMAN et CAELLI (1978)) ou

selon une géométrie de Cayley-Klein (DRÖSLER (1981b)). Ces approches ont cependant un champ d'applications actuellement limité, du fait de connaissances insuffisantes sur leurs présupposés théoriques mais aussi des problèmes d'interprétation des résultats posés par l'utilisation de représentations graphiques non Euclidiennes.

Le mode le plus répandu d'utilisation des méthodes de Codage Multidimensionnel s'inscrit dans un cadre exploratoire. On traite alors l'information contenue dans le triple  $(I, \Delta, Q_{|I|})$  où  $I$  est un ensemble d'unités statistiques,  $\Delta = (\delta_{ij})$  est un tableau carré d'ordre  $n = \text{card}(I)$ , regroupant les mesures de dissimilarité observées  $\delta_{ij}$  entre les unités statistiques  $i$  et  $j$ . La mesure de  $\delta_{ij}$  résulte le plus souvent de comparaisons par paires ou est déduite de mesures effectuées sur  $I$  (cf. DAVISON (1983), GREGSON (1975), SCHIFFMAN et al. (1981)) et  $Q_{|I|}$  est une pondération définie sur  $|I|$  qui prend en particulier en compte les problèmes liés aux données manquantes. On se propose alors de représenter les unités statistiques dans un espace affine Euclidien  $E$  de dimension aussi faible que possible, c'est à dire de construire une application  $X$  définie sur  $I$  à valeurs dans  $E$  de telle sorte que les distances  $d_{ij}(X) = d(X(i), X(j))$ , lues sur le graphique traduisent au mieux (en un sens à définir) l'information portée par  $\Delta$ . Le Codage Multidimensionnel est alors utilisé dans un but de réduction de données et de communication.

La recherche d'une interprétation plausible de la représentation des unités statistiques (ou configuration) ainsi construite repose alors sur l'inspection des positions relatives des unités statistiques. Bien que très satisfaisante pour l'esprit, la lecture des graphiques résultats en termes des distances interpoints est d'autant plus subjective que l'on dispose de peu d'informations a priori sur le domaine étudié, ou de théorie sur le processus de genèse des données. Notre capacité à trouver en toutes circonstances une explication a posteriori aux positions relatives des unités statistiques sur une représentation graphique nous expose en particulier aux conséquences fâcheuses de trois difficultés techniques rencontrées lors de la mise en œuvre d'un algorithme de Codage Multidimensionnel.

Tout d'abord, à l'exception de l'Analyse en Coordonnées Principales (GOWER (1966)) discutée plus bas, le problème posé n'est pas réductible à une technique d'analyse linéaire des données et par suite, il n'admet pas de solution analytique explicite. Les algorithmes itératifs mis en œuvre fournissent toujours une solution dont la qualité et la pertinence sont parfois difficiles à évaluer. Les algorithmes de sous-gradient mis en œuvre nécessitent une initialisation et sont de plus sensibles au choix de cette configuration de départ. Ils peuvent converger vers un optimum local, vers un point stationnaire, et leur vitesse de convergence est en général lente et de plus influencée par la courbure du graphe de la fonction objectif au voisinage de l'optimum.

En deuxième lieu, l'unicité de la configuration réalisant le minimum de la fonction objectif est d'autant moins assurée que le nombre d'unités statistiques représentées est faible relativement à la dimension de l'espace de représentation (SHERMAN (1972)). De plus des configurations très distinctes peuvent correspondre à une même valeur minimale de la fonction objectif. Une telle situation traduit par exemple un déplacement d'un sous-groupe des unités statistiques par symétrie autour d'un axe. (cf. CAELLI (1981)). Le problème s'aggrave lorsque l'on recherche des représentations non Euclidiennes, par exemple dans un espace de Minkowski.

La troisième difficulté rencontrée résulte du caractère fermé de la plupart des programmes d'analyse des données de dissimilarité actuellement disponibles. Elle justifie la présente étude. Une analyse approfondie des données, permettant de comprendre la structure de la relation de ressemblance étudiée, conduit souvent l'analyste à itérer les constructions de représentations spatiales des données. Chaque analyse joue un rôle provocateur et éveille à son esprit des hypothèses de travail qu'il souhaite mettre à l'épreuve par une nouvelle analyse du tableau de données. Cette démarche exploratoire nécessite de disposer d'un programme permettant de prendre en compte des contraintes qui traduisent ces hypothèses de travail lors de la construction de nouvelles configurations. Cette démarche doit être très nettement distinguée d'une méthodologie relevant de la statistique inférentielle, puisqu'un seul jeu de données est soumis à plusieurs analyses itérées, sous différentes hypothèses descriptives traduites par des contraintes spécifiques. La validation statistique éventuelle de la structure ainsi explicitée dans une optique inférentielle nécessite une étude confirmatoire complémentaire, fondée sur un nouveau corpus de données, collecté selon une méthodologie d'échantillonnage rigoureusement contrôlée.

Les problèmes statistiques et numériques posés par la prise en compte de contraintes lors de l'analyse de données de dissimilarité ont retenu très tôt l'attention des promoteurs des méthodes de Codage Multidimensionnel. Le plus souvent l'introduction de contraintes fut justifiée par des arguments numériques plutôt que méthodologiques. Un exemple de cet état d'esprit est fourni par les programmes distribués par McGEE (1968, 1969) et KRUSKAL et al. (1973) qui offrent des outils rudimentaires de construction d'un plongement Euclidien (ou dans un espace de Minkowski pour le second) de  $I$  à partir de la donnée d'un échantillon de  $K$  tableaux de dissimilarité  $\{\Delta_k, k=1, \dots, K\}$ . Dans ce cas, la contrainte prise en compte se traduit par le modèle descriptif

$$\Delta_k = D(X) \quad \forall k, k = 1, 2, \dots, K$$

où  $X$  est la configuration cherchée. Une telle modélisation est enrichie par la faculté de décomposer le tableau de dissimilarité traité en blocs correspondant soit à des lignes, soit à des colonnes, soit encore à des sous-tableaux diagonaux ou hors diagonaux qui permet

en particulier de construire une solution du problème d'Unfolding ou Déploiement Multidimensionnel (cf. CARROLL et ARABIE (1980), d'AUBIGNY (1981), HEISER et de LEEUW (1981), HEISER (1981)).

L'analyse d'un échantillon de tableaux de dissimilarité trouve une formalisation dans un cadre de statistique inférentielle dans les travaux de RAMSAY (1969, 1977, 1978, 1980, 1982) et TAKANE (1977) qui analysent les modèles ci-dessus en termes d'estimation dans le cadre d'une structure statistique adaptée à la résolution du problème. Dans une optique descriptive, l'analyse d'une famille de  $K$  tableaux de dissimilarité, dans le cadre d'un modèle algébrique, a donné naissance aux modèles dits de Différences Individuelles (HORAN (1969), CARROLL et CHANG (1970), BLOXOM (1968, 1974), KRANE (1978), HARSHMAN (1972)) caractérisés par la contrainte structurelle supplémentaire

$$\begin{aligned}\Delta_k &= D_k(X) \\ D_k(X) &= (d_{ij}^k(X)) \\ [d_{ij}^k(X)]^2 &= \sum_{s=1}^r m_k^s (x_i^s - x_j^s)^2\end{aligned}$$

Supposons de plus les coefficients  $m_k^s$  positifs, et appelons  $D_{M_k}$  la matrice diagonale de  $s$ -ième terme diagonal  $\sqrt{m_k^s}$ . On traduit alors la contrainte de façon équivalente, par:

$$\begin{aligned}\Delta_k &= D(Y_k) \text{ avec } Y_k = D_{M_k} X \\ [d_{ij}(Y_k)]^2 &= \sum_{s=1}^r ((Y_k)_i^s - (Y_k)_j^s)^2\end{aligned}$$

Ces modèles sont des cas particuliers de la classe de problèmes que nous étudions ici dans un cadre algébrique. L'aspect algorithmique du problème intervient ensuite, dans la mise en oeuvre des méthodes présentées.

Les modèles utilisés dans l'analyse d'un tableau de dissimilarité sont définis par des contraintes, comme tout modèle descriptif. Ces contraintes doivent être distinguées de celles qui définissent un sous-modèle du modèle de référence et qui traduisent des informations additionnelles résultant de connaissances a priori sur la structure des données provenant soit de recherches empiriques antécédentes, soit de théories connues dans le domaine étudié. Cette situation est classique en Statistique. Ainsi la régression linéaire multiple impose la contrainte d'appartenance de l'espérance mathématique de la variable réponse au sous-espace engendré par les variables explicatives. Cet impératif définit le modèle et doit être distingué de ce que l'on reconnaît usuellement comme des contraintes. Ainsi l'étude d'un modèle du type

$$y = 'X\beta + \varepsilon \tag{1}$$

$${}^1A\beta \leq c$$

(2)

considère que seule la relation linéaire  ${}^1A\beta \leq c$  définit les contraintes (cf. MONLEZUN et SPEED (1980), DENT (1980), LAWSON et HANSON (1974)).

De façon analogue, la pratique de la classification automatique impose des contraintes fortes aux données, traduites par la structure métrique de la configuration reconstruite. Ainsi les méthodes de classification hiérarchique construisent une matrice de distances ultramétriques approchant la dissimilarité de départ. Cette contrainte de modèle doit être distinguée des contraintes de contiguité récemment introduites afin d'enrichir cette méthodologie (cf. de SARBO et MAHAJAN (1984), FERLIGOJ et BATAGELJ (1982,1983), GORDON (1973a, 1973b), LEBART(1978), LECHEVALIER (1980), LEFKOVITCH (1980)).

Nous retrouvons donc dans le cas de l'analyse d'un tableau de dissimilarité une ambiguïté de termes inhérente à toute démarche exploratoire soucieuse de dégager le sous-modèle le plus adéquat d'un modèle de description de données et caractérisé par un ensemble de contraintes, dans un souci légitime de description fidèle et parsimonieuse des données. Cette démarche progressive caractérise à notre point de vue une analyse exploratoire et l'oppose à la simple analyse descriptive.

Notre étude ne porte pas sur ce qu'il est convenu d'appeler des contraintes de régularité. Celles-ci consistent par exemple à autoriser une transformation des données observées, déterminée dans une classe de transformations admissibles telles que les fonctions monotones utilisées en Codage Multidimensionnel Ordinal, ou encore les fonctions polynomiales (cf. KRUSKAL et WISH (1978)) ou les fonctions splines (monotones ou non) utilisées par WINSBERG et RAMSAY (1983). EL HADRI (1988) étudie de façon générale l'utilisation de transformations lisses des données dans les méthodes de Codage Multidimensionnel. L'imposition de contraintes de régularité réduit le nombre de paramètres définissant la transformation des données et améliore ainsi la stabilité des résultats. Elle relève donc de préoccupations distinctes mais complémentaires de la recherche de modèles de description des données.

Nous distinguons deux classes de contraintes. Elles correspondent à deux formulations distinctes des informations a priori et réfèrent en pratique à deux situations expérimentales différentes. La première classe impose des contraintes à la configuration reconstruite  $X$  et nous parlons alors de contraintes de configuration alors que de façon complémentaire, la seconde classe impose des contraintes sur les distances interpoints  $D(X)$ . Nous parlons alors de contraintes relationnelles. La correspondance non linéaire liant  $X$  à  $D(X)$  rend ces deux classes de contraintes non équivalentes en général. Les contraintes de configuration font référence à la structure d'espace vectoriel normé de l'espace de représentation, alors que les contraintes sur les distances font référence à sa structure d'espace métrique.

L'élaboration de techniques d'analyse Euclidienne d'un triple  $(I, \delta_I, q_{|I|})$  mêle les outils algébriques et les outils métriques (ou géométriques). Notre intuition est souvent mise en défaut par l'identification abusive des propriétés des espaces vectoriels normés à celles des espaces métriques, plus généraux. Certains exemples d'abus sont d'ailleurs cités par LE CALVE (1985). Or les conditions axiomatiques généralement imposées aux indices de dissimilarité nous placent naturellement dans des espaces (quasi-)métriques, que les méthodes de Codage Multidimensionnel se proposent d'approcher au mieux par des structures d'espace vectoriel normé, le plus souvent hilbertien ou Euclidien, ou encore de Minkowski  $l_p^r$ . Nous décrivons cependant un certain nombre de situations dans lesquelles les deux approches sont équivalentes.

Illustrons notre propos à l'aide de quelques exemples montrant que l'introduction de contraintes permet d'affiner l'analyse classique d'un tableau de dissimilarité.

Tout d'abord, le plongement métrique (disons Euclidien pour fixer les idées) de  $I$  conduit implicitement à introduire les paramètres par groupes, puisque l'ajout d'une dimension en complément d'un modèle descriptif jugé insuffisant du point de vue du critère d'approximation minimisé, impose la prise en compte de  $n$  nouveaux paramètres ( $n = \text{cardinal}(I)$ ), à savoir les  $n$  coordonnées des points représentant les unités statistiques sur l'axe ajouté. Cette démarche impose une relation de proportionnalité liant le nombre de dimensions de l'espace de représentation et le nombre de paramètres retenus, qui disparaît lors de l'introduction de contraintes. Un exemple est fourni par le modèle proposé par BENTLER et WEEKS (1978) pour traiter les données de EKMAN (1954) sur les dissimilarités entre couleurs. Celles-ci sont représentées approximativement le long d'un cercle dans la configuration bidimensionnelle obtenue par Codage Multidimensionnel Ordinal (SHEPARD (1962)) ou Métrique (BENTLER et WEEKS (1978)) sans contraintes. Ces derniers auteurs obtiennent un modèle descriptif plus satisfaisant en imposant aux coordonnées des stimuli sur une troisième dimension une contrainte de proportionnalité à la longueur d'onde de la couleur obtenue par analyse spectrale. On écrit alors :

$$y_i^3 = ab_i \quad b_i \text{ longueurs d'onde données a priori} \quad (3)$$

La prise en compte d'une dimension supplémentaire correspond alors à l'ajout d'un seul paramètre. Ce modèle correspond à la perturbation des dissimilarités déduite de l'égalité:

$$d_{ij}^2(Y) = d_{ij}^2(X) + a^2 (b_i - b_j)^2 \quad (4)$$

où  $X$  est la configuration bidimensionnelle d'intérêt, construite par Codage Multidimensionnel sans contraintes. On est donc amené à résoudre un problème de Codage Multidimensionnel sous contraintes. Les contraintes sont exprimées dans ce

problème par la donnée d'un vecteur  $\underline{b}$  contenant les mesures des longueurs d'onde (équation (3)) ou de façon équivalente par la donnée du tableau de distances Euclidiennes de terme général  $d_{ij}^2(\underline{b}) = (b_i - b_j)^2$ .

De façon complémentaire, on peut généraliser le modèle Euclidien usuel et imposer l'ajustement d'un grand nombre de dimensions en considérant les contraintes de structure portant sur la matrice des coordonnées des unités statistiques dans la représentation cherchée de la forme :

$$Y = \begin{bmatrix} X \\ D_u \end{bmatrix} \quad Y \in \mathcal{M}_{n+r,n}(\mathbb{R})$$

où  $D_u$  est la matrice diagonale de terme générique  $u_{ii} = u_i$  ( $u_i > 0$ ) qui s'interprète comme une unicité et mesure l'effet spécifique de l'unité statistique  $i$ . Chaque dimension supplémentaire correspond alors à un paramètre à estimer  $u_i$  et au total ce modèle sous contraintes conduit à estimer le même nombre de paramètres qu'un plongement euclidien de  $I$  dans un espace affine de dimension  $(r+1)$ . On vérifie d'ailleurs en notant  $\underline{1}_r$  le vecteur de dimension  $r$  de composantes constantes égales à 1 que la matrice

$$Z = \begin{bmatrix} I_r & 0_{r,r} \\ \underline{0}_r & \underline{1}_n \end{bmatrix} Y = \begin{bmatrix} X \\ \underline{1}_n D_u \end{bmatrix} = \begin{bmatrix} X \\ \underline{u}_n \end{bmatrix}$$

définit une représentation des unités statistiques dans l'espace  $\mathbb{R}^{n+r}$ .

Notons de plus que

$$d_{ij}^2(Y) = \sum_{s=1}^{r+n} (y_i^s - y_j^s)^2 = \sum_{s=1}^r (x_i^s - x_j^s)^2 + \sum_{s=r+1}^{r+n} (u_i^s - u_j^s)^2$$

$$d_{ij}^2(Y) = \begin{cases} d_{ij}^2(X) + u_i^2 + u_j^2 & \text{si } i \neq j \\ d_{ii}^2(X) & \text{si } i = j \end{cases} \quad (5)$$

soit, sous forme matricielle

$$D^{*2}(Y) = D^{*2}(X) + \underline{s} \underline{1} + \underline{1} \underline{s} - 2D_u^2 \quad \underline{s} = (u_i^2)_{i=1, 2, \dots, n} \quad (6)$$

où  $D^{*2}(A)$  désigne la matrice carrée de terme général le carré de la distance entre deux unités statistiques repérées par leurs coordonnées dans la matrice  $A$ . L'ajout de  $n$  dimensions, sous contrainte de diagonalité de la matrice  $D_u$ , correspond donc au modèle d'Analyse Factorielle en Facteurs Communs et Spécifiques étudié par CARROLL (1983) et WINSBERG et CARROLL (1984). La justification de cette dénomination résulte de l'identité corollaire de (6) (cf. d'AUBIGNY (1984)):

$$W_Y = W_X + D_u^2 \quad (7)$$

qui montre, en termes de matrices de produits scalaires qu'il s'agit là d'un modèle d'Analyse Factorielle Classique appliqué aux unités statistiques, c'est à dire d'une Q-Analyse Factorielle. Ces auteurs présentent ce modèle en termes de contraintes portant sur les distances, alors que nous venons de montrer qu'il s'exprime de façon équivalente en termes de contraintes portant sur la configuration Y.

**Remarque 1:** lorsqu'on impose de plus dans (5) la contrainte

$$u_i = u \quad \forall i = 1, 2, \dots, n$$

on obtient l'équation classique définissant le problème de la constante additive rencontré en Analyse en Coordonnées Principales, lorsque la matrice W n'est pas gramienne (cf. TORGERSON (1958), CAILLEZ et PAGES (1976), COOPER (1972)):

$$d_{ij}^2(Y) = \begin{cases} d_{ij}^2(X) + 2u^2 & \text{si } i \neq j \\ d_{ii}^2(X) & \text{si } i = j \end{cases}$$

Soit sous forme matricielle

$$D^{*2}(Y) = D^{*2}(X) + 2u^2 (\mathbf{1}\mathbf{1}' - I) \quad (8)$$

On montre ainsi que la prise en compte de la constante additive intervenant dans l'équation (8) fournit un modèle particulier de Codage Multidimensionnel sous contraintes. Il fait intervenir n dimensions supplémentaires contrôlées par un seul paramètre, associé à la configuration

$${}^tY = ({}^tX, u \mathbf{1}_n)$$

**Remarque 2 :** L'équation (7) se particularise alors en

$$W_Y = W_X + 2u^2 I_n$$

et montre ainsi l'analogie formelle liant le problème de la constante additive et la détermination du coefficient k (=2u<sup>2</sup> ici) en Ridge Regression. Par extension on montre par le même argument l'analogie entre le modèle d'analyse des tableaux de dissimilarité en facteurs communs et facteurs spécifiques et les méthodes bayésiennes empiriques en régression linéaire. En régression, ces méthodes sont utilisées en particulier dans un but de protection de la méthode contre les colinéarités entre les variables explicatives. Celles-ci traduisent alors l'existence de valeurs propres (quasi) nulles lors de la décomposition spectrale de l'opérateur de projection WN sur le sous espace engendré par les régresseurs. Ici les variables engendrant l'espace de représentation n'étant pas observées, W n'est pas

Gramienne par construction et l'on doit étendre la protection à l'apparition possible de valeurs propres négatives.

L'interprétation du modèle (5) en termes d'analyse factorielle classique n'est pas possible dans le cas du modèle d'analyse "sphérique" proposé par SAITO (1983, 1985, 1986) et donné par

$$\delta_{ij} = \begin{cases} d_{ij}(X) - u_i - u_j & \text{si } i \neq j \\ d_{ij}(X) & \text{si } i = j \end{cases} \quad (9)$$

Soit sous forme matricielle

$$\Delta = D(X) - \mathbf{u} \mathbf{1} - \mathbf{1} \mathbf{u} + 2D_v$$

Les modèles (5) et (9) sont en effet équivalents dans le seul cas où est vérifiée la contrainte supplémentaire

$$\frac{1}{d_{ij}(X)} = \frac{1}{u_i} + \frac{1}{u_j} \quad u_i \neq 0 \quad \forall i \in I \quad (10)$$

imposée à tout couple (i,j) tel que  $d(i,j) \neq 0$ .

Dans le cas général, le modèle de SAITO ajuste des coefficients  $u_i$  réels de signe quelconque et n'impose pas de contraintes directement interprétables sur la configuration  $X$  (type 1) lorsque l'on interprète celle-ci comme la matrice d'une application linéaire. Les contraintes portent ici sur les distances et illustre le cas d'une contrainte non exprimable sur la configuration  $X$ , dans un cadre Euclidien. Cet auteur interprète les paramètres réels  $u_i$  comme des effets de contexte, qui corrigent l'évaluation des distances  $\delta_{ij}$ , en représentant l'unité statistique  $i$  sur une hypersphère de rayon  $|r_i|$  et centrée en un point  $x_i$  servant à évaluer la distance "hors contexte" entre unités statistiques. Si  $u_i > 0$ , l'effet de contexte accroît les distances, si  $u_i < 0$ , il les réduit. On s'accorde ainsi un degré de liberté supplémentaire par rapport aux contraintes de représentation exigées par l'hypothèse Euclidienne faite sur  $d_{ij}(X)$ . Écrit sous la forme

$$\delta_{ij} = d_{ij}(X) - c_{ij} \quad \text{si } i \neq j \quad (11)$$

avec ici  $c_{ij} = u_i + u_j$ , le modèle de SAITO apparaît en effet comme une tentative de modélisation du paramètre de non Euclidéanité  $c_{ij}$ . Il fournit en fait pour le cadre géométrique des plongements Euclidiens un analogue de la décomposition des distances arborées, associées aux représentations des données de dissimilarité par des arbres additifs. Ces derniers se décomposent en effet en la somme d'une distance ultramétrique

et d'une distance à centre (LECALVE (1985), BROSSIER (1986)) que l'on peut interpréter comme une mesure de l'effet de contexte.

**Remarque 3 :** On peut noter l'analogie de ces décompositions avec les pratiques de l'Analyse de la Variance d'une table à 2 entrées sans répliques de terme général  $\delta_{ij}$ : le terme  $\tau_{ij}$  fournit la partie additive d'une décomposition du tableau de dissimilarité dont la forme particulière est liée à la symétrie du tableau, le terme résiduel étant alors interprété comme la mesure d'une distance entre unités statistiques. Dans cette interprétation des données, le modèle de CARROLL est décrit par la table de terme général le carré de  $\delta_{ij}$ .

D'autre part, la prise en compte des inégalités triangulaires

$$d_{ij}(X) + d_{jk}(X) \geq d_{ik}(X)$$

montre que les paramètres  $u_i$  doivent respecter les contraintes

$$2u_i \geq \max_{j,k} (\delta_{jk} - \delta_{ij} - \delta_{ik}) \quad (12)$$

en plus de la contrainte de positivité d'une distance

$$d_{ij}(X) = u_i + u_j + \delta_{ij} \geq 0 \quad i \neq j \quad (13)$$

**Remarque 4 :** Dans le cadre de l'étude des distances arborées, la contrainte (12) s'interprète comme une majoration par  $2u_i$  de l'allongement des chemins  $(i,j)$  sur le graphe de sommets  $I$  et d'arêtes  $(i,j)$  valuées par  $\delta_{ij}$ .

SAITO construit la solution en termes d'ajustement aux moindres carrés à l'aide d'une méthode de gradient, malgré les problèmes de différentiabilité de la fonction objectif aux points  $X$  pour lesquels  $d_{ij}(X) = 0$  ( $i \neq j$ ). La construction de la solution pose en fait un problème numérique difficile, dont on rencontre un cas particulier en Analyse en Coordonnées Principales, lorsqu'on résout le problème de la constante additive associé à l'identité

$$\delta_{ij} = d_{ij}(X) - u_i - u_j \quad i \neq j$$

dont une solution est donnée par CAILLEZ (1983).

La modélisation de SAITO est voisine de la proposition de KRUMHANSL (1978) dont le modèle s'écrit pour  $i \neq j$

$$\delta_{ij} = d_{ij}(X) + \alpha \tau_i + \beta \tau_j \quad \tau_j > 0 \quad \alpha, \beta > 0 \quad (14)$$

soit sous forme matricielle

$$\Delta = D(X) - \alpha \mathbf{1}' \mathbf{1} - \beta \mathbf{1} \mathbf{1}'$$

Ce modèle a pour but de prendre en compte les mesures de densité spatiale  $\tau_i$  estimées aux constantes positives de normalisation  $\alpha$  et  $\beta$  près. Le modèle de KRUMHANSL affirme que l'évaluation des proximités entre unités statistiques présente pour le sujet interrogé un degré de difficulté dépendant de la densité spatiale du processus ponctuel sous-jacent au voisinage des deux unités statistiques comparées. Cet auteur ne fournit cependant pas d'algorithme résolutif.

La classe des modèles de Codage Multidimensionnel sous contraintes du type (11), c'est à dire formulées en termes de distances, ne se limite pas aux approches de KRUMHANSL et SAITO. En effet, ce type de modélisation permet, par exemple, de prendre en compte les asymétries observées dans la mesure des dissimilarités. Ainsi WEEKS et BENTLER (1982) proposent une modélisation unidimensionnelle de l'asymétrie décrite par l'équation

$$\delta_{ij} = \begin{cases} d_{ij}(X) + a + c_i - c_j & c_i, c_j \in \mathbb{R} \quad \text{si } i \neq j \\ d_{ij}(X) & \text{si } i = j \end{cases}$$

soit sous forme matricielle

$$\Delta = D(X) + a(\mathbf{1}' \mathbf{1} - I) + (\underline{c}' \mathbf{1} - \mathbf{1}' \underline{c})$$

que l'on peut généraliser sans peine au modèle général de décomposition d'une forme bilinéaire en

$$\Delta = D(X) + a(\mathbf{1}' \mathbf{1} - I) + \sum_{s=1}^S \gamma_s (\underline{c}_s' \mathbf{1} - \mathbf{1}' \underline{c}_s)$$

sous contrainte d'orthogonalité des composantes  $\underline{c}_s$ . Les termes  $(\underline{c}_s' \mathbf{1} - \mathbf{1}' \underline{c}_s)$  sont par construction des éléments de  $F \wedge F$  de la forme  $T_{\underline{c}_s}$  étudiée plus loin. On obtient ainsi, à une constante additive près, une décomposition particulière d'un tableau de dissimilarité  $\Delta$  en une partie symétrique, interprétée comme une distance Euclidienne et une partie antisymétrique que l'on se propose de décomposer en éléments simples dans  $F \wedge F$ .

Le chapitre II de ce mémoire est consacré à des rappels sur l'analyse sans contraintes du triple  $(I, \delta_I, Q_{I \times I})$ .  $I$  est alors un ensemble fini,  $\delta_I$  une dissimilarité définie sur  $I \times I$  et  $Q_{I \times I}$  est une pondération des éléments  $(i, j)$  de  $I \times I$ . On décrit en particulier la structure algébrique qui permet d'étudier les méthodes d'Analyse en Coordonnées Principales adaptées à la description de distances Euclidiennes dans le cas d'une pondération  $P_I$  des unités statistiques, et les méthodes de Codage Multidimensionnel qui ont pour but

d'ajuster un tableau de distances (Euclidiennes ou plus générales) à un tableau de dissimilarités observées.

Les chapitres suivants constituent la partie essentielle de notre exposé. Nous nous intéressons alors aux classes de contraintes qu'il est possible de prendre en compte afin d'enrichir l'analyse d'un tableau de dissimilarité. Nous distinguons les contraintes portant directement sur la configuration  $X$  elle-même, du cas plus général fourni par des contraintes relationnelles, exprimées à l'aide d'une ou de plusieurs matrices de dissimilarité. Dans le premier cas, une étude détaillée de la classe des contraintes linéaires nous permet de montrer les liens existant entre la méthodologie de l'Analyse en Composantes Principales sur Variables Instrumentales et les techniques d'Analyse du Triple  $(I, \delta_I, Q_{[X]})$  sous contraintes linéaires. Une écriture tensorielle nous permet d'en mieux situer le statut. Nous établissons le lien entre les modèles généraux étudiés par CARROLL et al. (1976) et les modèles d'analyse des courbes de croissance, en section III.1.1 et nous montrons dans la section III.1.2, en quoi elles généralisent les solutions du problème d'Analyse en Composantes Principales sur Variables Instrumentales, exprimé en termes d'opérateurs. Dans chacun de ces cas notre attention est portée sur l'applicabilité des méthodes à l'analyse de tableaux de dissimilarité. Nous étendons de plus cette étude à des contraintes liées aux propriétés d'invariance dans les espaces Euclidiens en III.1.3. Nous montrons que la méthode de rotations procrustes s'applique aussi au cas de données de dissimilarité permettant ainsi de traduire dans ce cadre d'étude, des hypothèses de travail sous forme de configurations de référence. Le cas des contraintes non linéaires, en particulier ordinales portant sur la configuration  $X$  est ensuite étudié (section III.2). Leur généralité en interdit le traitement par Analyse en Coordonnées Principales. Nous justifions alors l'approche heuristique développée par NOMA et JOHNSON (1979) dans le cadre des méthodes de Codage Multidimensionnel. Nous montrons en effet que les contraintes définies par des variables instrumentales forment une sous-classe de l'ensemble des contraintes qu'il est possible de traiter simplement par un algorithme de Codage Multidimensionnel tel que celui qui est proposé par de LEEUW et HEISER (1980).

Le chapitre IV est consacré à l'étude des contraintes relationnelles déterminées par la donnée d'une ou de plusieurs matrices de dissimilarité. Cette classe très générale de contraintes est tout d'abord étudiée dans le cadre de l'Analyse en Coordonnées Principales. On traite en détail le cas d'une contrainte relationnelle décrite par un graphe, afin d'en montrer les liens avec certaines pratiques plus classiques en l'analyse linéaire des données de profil. Ces contraintes de graphe sont associées à une relation binaire, non valuée et sont étudiées en section IV.1.1. La section IV.1.2 considère alors l'extension de l'approche proposée au cas d'une relation valuée. Le cas de relations floues est d'abord

traité, puis nous présentons le cas général, dans lequel la valuation du graphe des contraintes est interprétée comme un outil de description de contraintes métriques, c'est à dire portant sur les proximités entre unités statistiques. Les méthodes de Codage Multidimensionnel permettent de traiter des contraintes plus générales et font l'objet de la section IV.2. Ces méthodes intègrent par exemple les problèmes de Codage Multidimensionnel Ordinal, sous des contraintes déterminées par une ou plusieurs matrices de dissimilarité, considérées du point de vue numérique puis du point de vue ordinal. Nous résolvons le problème dans le cas où chaque matrice de contraintes définit dans l'ensemble des matrices de configuration  $X$ , un sous-ensemble de contraintes caractérisé comme un convexe fermé pour la topologie usuelle de l'espace ambiant. La méthode proposée est alors caractérisée en termes d'analyse canonique généralisée entre convexes fermés et nous proposons des algorithmes résolutifs dont les propriétés de convergence sont connues. L'extension de ces résultats au cas de contraintes définissant des ensembles proximaux est immédiate. Nous n'avons cependant pas recherché la plus grande généralité possible, dans un souci d'adéquation de notre étude au cadre usuel d'application des méthodes étudiées.

## CHAPITRE II

### L'ANALYSE DU TRIPLE $(I, \delta_I, Q_{\mathbf{I} \times \mathbf{I}})$ SANS CONTRAINTES

#### II.1. RAPPELS ET NOTATIONS

Soit  $I$  un ensemble fini d'unités statistiques de cardinal  $n$ . L'observation d'une fonction de dissimilarité  $\delta_I$  définie sur  $I \times I$  permet de construire le tableau de dissimilarité  $\Delta = \{\delta_{ij}, i, j \in I\}$ . De façon classique nous supposons ici que  $\delta_I$  définit une fonction de dissimilarité semi-propre sur  $I$ , c'est-à-dire telle que pour tout  $(i, j)$  dans  $I \times I$

- (1)  $\delta_{ij} \geq 0$
- (2)  $\delta_{ij} = \delta_{ji}$
- (3)  $\delta_{ij} = 0 \Rightarrow \delta_{ik} = \delta_{kj} \quad \forall k \neq i, j \quad k \in I$

La fonction de dissimilarité  $\delta_I$  est dite propre si la condition (3) est remplacée par la condition plus forte :

- (4)  $\delta_{ij} = 0 \Rightarrow i = j \quad \forall i, j \in I$

Nous nous intéressons ici au plongement Euclidien de  $I$ . On suppose donc l'existence d'un espace Euclidien  $E = (\mathbb{R}^p, m)$  où  $m$  est une forme quadratique définie sur  $\mathbb{R}^p$  et d'une application

$$\begin{aligned} x : I &\rightarrow E \\ i &\mapsto x(i) \end{aligned}$$

tels que les distances reconstruites  $d_{ij}(x) = m(x(i) - x(j))$  représentent au mieux, au sens d'un critère à définir, les dissimilarités observées. Supposons tout d'abord que  $\delta_I$  est une dissimilarité Euclidienne, c'est à dire qu'il existe un triplet  $(p, m, z)$  tel que  $\delta_I$  soit évalué dans  $E$  par

$$\delta_{ij} = m(z(i) - z(j)) \quad \forall i, j \in I$$

Dans la suite, on note  $z$  l'application associée à une dissimilarité Euclidienne et l'on précise le caractère Euclidien de la dissimilarité par la notation  $\delta_I(z)$ . L'analyse linéaire de la représentation  $z$  de  $I$  décrit alors classiquement  $z$  comme une application linéaire

$$z : F^* \rightarrow E$$

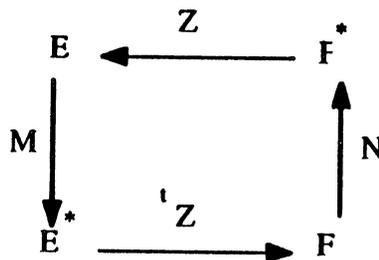
où chaque élément de  $I$  est associé à un vecteur de la base  $\{f_i^*, i \in I\}$  de l'espace vectoriel réel libre  $F^*$  de dimension  $n$ .

**Remarque:** Nous adoptons dans la suite la convention usuelle en analyse des données, consistant à identifier par abus de langage  $\mathbb{R}^n$  et  $F$  (respectivement  $\mathbb{R}^p$  et  $E$ ) lorsque l'espace vectoriel  $\mathbb{R}^n$  (resp.  $\mathbb{R}^p$ ) définit la structure Euclidienne  $F = (\mathbb{R}^n, n)$  (resp.  $E = (\mathbb{R}^p, m)$ ). L'utilisation du dual algébrique  $F^*$  de l'espace vectoriel libre  $F$ , plutôt que  $F$  lui-même, est arbitraire. Elle est retenue ici de façon à expliciter l'analogie des outils mis en œuvre, avec le formalisme de l'analyse linéaire d'un tableau  $X$  de données de profil. On peut d'autre part interpréter  $\mathbb{R}^p$  ( $p \leq n$ ) comme un sous-espace vectoriel de l'espace de référence  $\mathbb{R}^n$ .

Lorsqu'on a fait choix d'une base  $\{e^j, j \in J\}$ ,  $J = \{1, 2, \dots, p\}$  dans  $E$ , l'application linéaire  $z$  permet d'identifier  $z(i)$  au vecteur de coordonnées de sa représentation dans  $E$ :

$$z : f_i^* \rightarrow z(f_i^*) = z_i$$

Soit  $M$  la matrice associée à la forme quadratique  $m$  dans la base de  $E$  et  $N$  la matrice de la forme quadratique  $n$  dans la base de  $F$ , la situation peut être résumée par le schéma de dualité suivant



Chaque méthode d'analyse est caractérisée en fait par l'argument utilisé pour faire le lien entre la matrice des observations  $\Delta$  et les constituants du schéma de dualité donné ci-dessus.

Lorsque  $E$  et  $F$  sont munis respectivement des bases  $\{e^j, j \in J\}$  et  $\{f_i, i \in I\}$ , le produit tensoriel  $F \otimes E$  est muni naturellement de la base  $\{f_i \otimes e^j, i \in I, j \in J\}$ , que nous décrivons ici dans l'ordre lexicographique défini sur  $I \times J$ . De façon classique l'application  $z$  peut être identifiée à un tenseur de  $F \otimes E$ .  $z$  est alors décrite de façon équivalente soit par la matrice

$Z$  de  $\mathcal{M}_{p \times n}(\mathbb{R})$ , soit par le vecteur  $\underline{z} \in \mathbb{R}^{np}$  déduit de  $Z$  par l'opérateur de vectorialisation  $\text{Vec}$  (cf. HENDERSON et SEARLE (1979)):

$$\underline{z} = \text{Vec}(Z) = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_i \\ \vdots \\ z_n \end{bmatrix} \quad \text{si} \quad Z = [ z_1 \quad z_2 \quad \dots \quad z_i \quad \dots \quad z_n ]$$

On vérifie alors simplement les relations élémentaires suivantes, utilisées dans la suite de ce travail, dont on peut trouver une démonstration par exemple dans HENDERSON et SEARLE (1979), BALESTRA (1975), MAGNUS (1988) ou MAGNUS et NEUDEKER (1988) :

$$\begin{aligned} Y = AX {}^tB &\Leftrightarrow \text{Vec}(Y) = B \otimes A \text{Vec}(X) \\ &\Leftrightarrow \underline{y} = B \otimes A \underline{x} \end{aligned}$$

où l'on note  $B \otimes A$  le produit de Kronecker des matrices  $A$  et  $B$ , défini par

$$B \otimes A = [b_{ij} A]_{ij}$$

$$\begin{aligned} \text{Trace}({}^tA B) &= {}^t\text{Vec}(A) \text{Vec}(B) \\ &= {}^t\underline{a} \underline{b} \end{aligned}$$

et plus généralement, si  $X$ ,  $Y$ ,  $M$ , et  $N$  sont des matrices quelconques d'ordre tel que le produit  $MYN {}^tX$  soit défini,

$$\begin{aligned} \text{Trace}(MYN {}^tX) &= {}^t\text{Vec}(X) {}^tN \otimes M \text{Vec}(Y) = {}^t\underline{x} {}^tN \otimes M \underline{y} \\ &= {}^t\text{Vec}({}^tX) M \otimes {}^tN \text{Vec}({}^tY) \end{aligned}$$

En particulier si les matrices  $M$  et  $N$  définissent les métriques respectives sur  $E$  et  $F$ :

$$\begin{aligned} \text{Trace}(MXN {}^tX) &= {}^t\text{Vec}(X) N \otimes M \text{Vec}(X) = {}^t\underline{x} N \otimes M \underline{x} \\ &= \| \underline{x} \|_{N \otimes M}^2 \end{aligned}$$

La matrice symétrique semi-définie positive  $N \otimes M$  définit en effet une structure Euclidienne sur l'espace vectoriel  $\mathbb{R}^n \times \mathbb{R}^p$ :

$$F \otimes E = (\mathbb{R}^n \times \mathbb{R}^p, n \otimes m)$$

et l'on note, avec les abus usuels de notation:

$$N \otimes M : F \otimes E \rightarrow F^* \otimes E^*$$

Le produit scalaire

$$\langle X, Y \rangle_{N \otimes M} = \text{Trace}('XMYN)$$

permet alors de définir le coefficient de corrélation linéaire entre vecteurs centrés dans  $F \otimes E$ :

$$\begin{aligned} r(X, Y) &= \frac{\langle X, Y \rangle_{N \otimes M}}{\|X\|_{N \otimes M} \cdot \|Y\|_{N \otimes M}} = \frac{\text{Trace}('XMYN)}{\sqrt{\text{Trace}('XMXN)} \cdot \sqrt{\text{Trace}('YMYN)}} \\ &= \frac{\text{Trace}('XMYN)}{\sqrt{\text{Trace}(W_X N)} \cdot \sqrt{\text{Trace}(W_Y N)}} \end{aligned}$$

c'est à dire le coefficient de corrélation entre opérateurs  $\text{COR}(X, Y)$ , étudié entre autres par MIZERE (1981).

## II.2. L'ANALYSE EN COORDONNEES PRINCIPALES DU TRIPLE (I, $\delta_I$ , $P_I$ )

Nous considérons dans la suite le triple  $(I, \delta_I, P_I)$  formé d'un ensemble fini d'unités statistiques  $I$  ( $\text{card}(I) = n$ ), d'un tableau de données  $\Delta$  carré d'ordre  $n$ , et d'une pondération  $P_I$  des unités statistiques satisfaisant

$$p_i \geq 0 \quad \text{et} \quad \sum_{i=1}^n p_i = 1$$

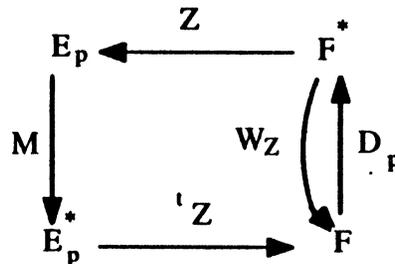
$\Delta$  regroupe les dissimilarités  $\delta_{ij}$  observées entre les unités statistiques  $i$  et  $j$  ( $i, j = 1, \dots, n$ ). Considérons de plus un espace vectoriel libre de dimension  $n$ , muni de la structure Euclidienne induite par la pondération  $P_I$ . On note  $F = (\mathbb{R}^n, n)$  et  $\{f_i, i \in I\}$  une base de  $F$ . Le dual algébrique de  $F$  est alors muni de la base duale  $\{f_i^*, i \in I\}$ , pour laquelle la matrice  $N$  représentant la forme quadratique  $n$ , est diagonale de terme générique  $n_{ii} = p_i$ , justifiant la notation  $N = D_p$ . On note alors  $F^* = (\mathbb{R}^n, n^{-1})$  le dual de  $F$  muni de sa structure Euclidienne canonique.

Les unités statistiques sont alors associées de façon canonique aux vecteurs de base  $f_i^*$  et l'on dit que  $\Delta$  est plongeable dans un espace Euclidien de dimension  $p \leq n$  si et seulement s'il existe une image Euclidienne de  $(I, \delta_I, P_I)$  notée  $(\mathcal{M}, E)$  où  $\mathcal{M}$  est l'ensemble des points  $\{M_i, i \in I\}$  de coordonnée  $z_i$  dans un espace vectoriel réel  $E$  de dimension  $p$  (On précise si nécessaire  $E = E_p$ ), muni d'une base quelconque  $\{e_j, j=1 \dots p\}$  et d'une structure métrique Euclidienne connue (ou choisie) a priori et définie par la donnée d'une matrice

symétrique semi-définie positive (on dit aussi une métrique)  $M$ . Les vecteurs  $z_i$  définissent une application linéaire  $z$  telle que :

$$\begin{aligned} z : F^* &\rightarrow E \\ \underline{f}_i^* &\rightarrow z(\underline{f}_i^*) = z_i \end{aligned}$$

GOWER (1966, 1982), CAILLEZ et PAGES (1975), MAILLES (1979) et d'AUBIGNY (1984) étudient en détail les problèmes mathématiques posés par le plongement Euclidien de  $(I, \delta_I, P_I)$ . On construit alors le schéma de dualité résumant l'étude linéaire du problème



**Définition 1:** Une dissimilarité propre  $\delta_I$  est dite Euclidienne, s'il existe un plongement Euclidien  $z$  de  $I$  dans un espace Euclidien de dimension  $p$ , noté  $E_p$ , tel que  $\delta_{ik}^2 = m(z_i - z_k)$  où la forme quadratique  $m$  définit la géométrie de  $E_p$ . On note alors  $\delta_I = \delta_I(Z)$ .  $Z$  est alors appelée une (matrice de ) configuration associée à  $I$ .

On note qu'alors la matrice  $Z$  associée à  $z$  dans ces bases définit une isométrie partielle entre les espaces Euclidiens  $F^*$  et  $E$ , que la relation (1) suivante exploite de façon classique :

$$\delta_{ij}^2 = \delta_{ij}^2(Z) = \|z_i - z_j\|_M^2 = \|f_i^* - f_j^*\|_{W_Z}^2 \quad (15)$$

où par définition,  $W_Z = {}^t Z M Z$ . Dans cette équation, la norme d'un vecteur  $\underline{x}$ , élément d'un espace Euclidien caractérisé par la métrique  $M$  s'écrit :

$$\|\underline{x}\|_M^2 = m(\underline{x}) = \langle \underline{x}, \underline{x} \rangle_M = \sum_i \sum_j m_{ij} x_i x_j$$

$$M = (m_{ij})_{i, j = 1, 2, \dots, n}$$

Le rôle central de cette construction est dévolu à la matrice

$$W = (w_{ij})_{i, j = 1, 2, \dots, n} \quad w : F^* \rightarrow F$$

déduite de  $\delta$  par la transformation due à YOUNG et HOUSEHOLDER (1938) et popularisée par TORGERSON (1958):

$$w_{ij} = w(\mathbf{f}_i^*, \mathbf{f}_j^*) = -\frac{1}{2} \left[ \delta_{ij}^2 - \delta_{i.}^2 - \delta_{.j}^2 + \delta_{..}^2 \right] \quad (16)$$

où par définition

$$\begin{aligned} \delta_{i.}^2 &= \sum_j p_j \delta_{ij}^2 \\ \delta_{..}^2 &= \sum_i p_i \delta_{i.}^2 = \sum_i \sum_j p_i p_j \delta_{ij}^2 \end{aligned}$$

On montre en effet (cf. CAILLEZ et PAGES (1976) p.252) que si  $\Delta$  est une matrice de distance Euclidienne entre unités statistiques, associée au produit scalaire  $w$ ,

$$\begin{aligned} \delta_{ij}^2 &= w(\mathbf{f}_i^* - \mathbf{f}_j^*)^2 \\ &= w(\mathbf{f}_i^*) + w(\mathbf{f}_j^*) - 2 w(\mathbf{f}_i^*, \mathbf{f}_j^*) \end{aligned}$$

et donc  $w_{ij}$  s'exprime sous une forme équivalente à (16) :

$$w(\mathbf{f}_i^*, \mathbf{f}_j^*) = \frac{1}{2} (w(\mathbf{f}_i^*) + w(\mathbf{f}_j^*) - \delta_{ij}^2) \quad (16 \text{ bis})$$

Les distances Euclidiennes  $\delta_{ij}$  sont des quantités intrinsèques qui réfèrent à la structure affine des espaces de représentation (cf. FICHET (1983)). Par suite, le traitement du problème dans le formalisme algébrique des espaces vectoriels réels nécessite le choix arbitraire d'une origine. Fixons l'origine de  $E$  au barycentre  $\mathbf{g}$  du nuage de points massiques  $\{(M_i, p_i), i=1, \dots, n\}$  et utilisons les propriétés d'isométrie partielle de l'application  $z$ . Alors d'après le théorème de HUYGHENS, l'inertie de  $\mathbf{z}_i$  dans  $E$  est donnée par

$$I_{\mathbf{z}_i} = I_{\mathbf{g}} + m(\mathbf{z}_i - \mathbf{g}) = I_{\mathbf{g}} + w(\mathbf{f}_i^*)$$

Mais, par définition de l'inertie

$$\begin{aligned} I_{\mathbf{z}_i} &= \sum_{j=1}^n p_j m(\mathbf{z}_i - \mathbf{z}_j) = \sum_{j=1}^n p_j \delta_{ij}^2 = \delta_{i.}^2 \\ I_{\mathbf{g}} &= \sum_{j=1}^n p_j m(\mathbf{z}_i - \mathbf{g}) = \sum_{j=1}^n p_j I_{\mathbf{z}_i} - I_{\mathbf{g}} \end{aligned}$$

et donc l'équation (16) exprime la relation classique liant distances et produits scalaires en fonction des dissimilarités observées et des pondérations. C'est une conséquence directe du théorème de Pythagore, dont l'expression analytique est liée au choix du barycentre  $\mathbf{g}$

des points massiques  $(M_i, p_i)$  comme origine des axes dans E. Du point de vue algébrique, le barycentre a pour coordonnées dans F le vecteur

$$g = \mathbb{1}' \mathbb{1} D_p {}^t Z$$

De sorte que la matrice  $W_Z$  s'écrit sous une forme équivalente à la relation (16)

$$W_Z = Q_1 H {}^t Q_1$$

où en notant \* le produit de Hadamard entre deux matrices

$$H = -\frac{1}{2} (\Delta(Z) * \Delta(Z)) = -\frac{1}{2} \Delta^{*2}(Z) = \left(-\frac{1}{2} \delta_{ij}^2(Z)\right) \quad (17)$$

$$Q_1 = I - \mathbb{1}' \mathbb{1} D_p$$

Malgré son caractère arbitraire (cf. GOWER (1966, 1982)), le choix d'une origine s'impose cependant puisqu'une telle approche construit la configuration X à partir de l'opérateur  $W_D$  et utilise donc indirectement la structure d'espace métrique induite par la distance Euclidienne  $\delta(z)$  à l'aide des propriétés des espace hilbertiens de dimension finie. La structure d'espace vectoriel de E suppose le choix d'une origine et impose de plus de s'affranchir des problèmes d'invariance par translation d'une distance Euclidienne.

Les distances Euclidiennes sont de plus invariantes par rotation des axes associés à la base de l'espace vectoriel sous-jacent. La coutume conduit à fixer une orientation des axes déterminée par la base des vecteurs propres de  $W_D$ . Si les valeurs propres  $\lambda_s$  de  $W_D$  sont distinctes, on obtient ainsi une solution unique qui tire sa notoriété de sa très forte analogie avec la méthodologie de l'Analyse en Composantes Principales. Aussi adoptons nous le terme d'Analyse en Coordonnées Principales, proposé par GOWER (1966) pour désigner cette démarche, de préférence aux termes ambigus de "Classical Scaling", de "Canonical Coordinate Analysis" et de "TORGERSON-GOWER Scaling" souvent employés dans les textes en langue anglaise, ou encore d'"Analyse Factorielle du Triple" utilisé en France.

La construction d'une représentation résumée retient les r plus grandes valeurs propres de  $W_D$  et construit donc une approximation par défaut des dissimilarités initiales, du fait du caractère contractant de la projection orthogonale sur une variété affine. Ecrivons en effet

$$d_{ij}^2(Z) = \sum_{s=1}^r (z_i^s - z_j^s)^2 + \sum_{s=r+1}^p (z_i^s - z_j^s)^2$$

on vérifie alors de façon immédiate les inégalités suivantes pour tout  $(i,j) \in I \times I$

$$d_{ij}^2(X_1) \leq d_{ij}^2(X_2) \leq \dots \leq d_{ij}^2(X_r) \leq \dots \leq d_{ij}^2(X_p) \leq d_{ij}^2(Z) = \delta_{ij}^2$$

où  $X_s$  est le tableau des coordonnées des points  $M_i$  en  $s$  dimensions ( $s = 1, 2, \dots, p$ ), associée aux  $s$  plus grandes valeurs propres de  $WD_p$ . D'autre part,  $d_{ij}^2(X)$  s'écrit dans  $E_r$  inclus dans  $E_p$ ,  $\dim(E_r) = r$

$$d_{ij}^2(X) = \begin{cases} \sum_{s=1}^r (x_i^s - x_j^s)^2 & \text{si on normalise } X \text{ de sorte que } X {}^tX = \Lambda_r \\ \sum_{s=1}^r \lambda_s^2 (x_i^s - x_j^s)^2 & \text{si on normalise } X \text{ de sorte que } X {}^tX = I_r \end{cases}$$

où  $\Lambda_r$  est la matrice diagonale  $\Lambda_r = \text{diag} \{(\lambda_s^2) ; s = 1, 2, \dots, r\}$ .

L'optimalité de cette méthode s'exprime en termes d'approximation des produits scalaires au sens des moindres carrés. La décomposition spectrale de  $WD_p$  assure alors des propriétés d'optimalité de la solution analogues aux résultats classiques en Analyse en Composantes Principales. En particulier, lorsque  $\Delta$  est une matrice de dissimilarité Euclidienne  $\Delta = \Delta(Z)$ , de sorte que la matrice associée  $W_Z$  est gramienne, c'est à dire symétrique et semi-définie positive, on est alors assuré de l'emboîtement des sous-espaces propres associés aux plus grandes valeurs propres de  $W_Z D_p$  et la configuration  $X_r$  de dimension  $r$ , réalise le minimum de la fonction objectif

$$\begin{aligned} \text{STRAIN} &= \text{Trace} ((W_Z - W_r) D_p (W_Z - W_r) D_p) = \| \underline{w}_Z - \underline{w}_r \|_{D_p \otimes D_p}^2 \\ &= \| \underline{w}_Z \|_{Q_r \otimes Q_r}^2 \quad Q_r = D_p (I - P_r) \end{aligned} \quad (18)$$

où  $\underline{w} = \text{Vec}(W)$  désigne le vecteur représentatif de  $W$  dans  $F \otimes F$ .

Le minimum de la fonction STRAIN est évalué parmi toutes les matrices de produit scalaire de rang  $r$  définies sur  $F^*$  et l'on a :

$$W_r = {}^t Y_r M Y_r \quad Y_r \in \mathcal{M}_{r \times r}(\mathbb{R}) \quad \underline{w}_r = {}^t Y_r \otimes {}^t Y_r \underline{m}$$

L'optimisation du critère STRAIN ne répond cependant pas à l'objectif initial des méthodes de Codage Multidimensionnel Optimal, puisque le critère (18) n'assure pas l'optimalité de la reconstitution des distances au sens des moindres carrés ordinaires. En effet, on vérifie que la fonction objectif STRAIN s'écrit aussi à l'aide du projecteur  $D_p$ -orthogonal sur le sous espace  $F_1$ , supplémentaire  $D_p$ -orthogonal de la droite des constantes dans  $F$ :

**Proposition 1:**

$$\text{STRAIN} = \frac{1}{4} \|\delta^2(Z) - d^2(X)\|_{N_P \otimes N_P}^2 \quad (18 \text{ bis})$$

où

$$N_P = D_P Q_1 = D_P \cdot p \cdot p \quad p = (p_1, p_2, \dots, p_n).$$

**Preuve:** Si  $\Delta$  est une dissimilarité Euclidienne,

$$W_z = Q_1 H(Z) Q_1 \quad \text{où } H(Z) = -\frac{1}{2} \Delta(Z) * \Delta(Z)$$

qui s'écrit aussi, avec des notations évidentes

$$w_z = Q_1 \otimes Q_1 h(Z) = -\frac{1}{2} Q_1 \otimes Q_1 \delta^2(Z)$$

de sorte que la fonction

$$\text{STRAIN} = \text{Trace}(Q_1 (\Delta^{*2}(Z) - D^{*2}(X)) Q_1 (\Delta^{*2}(Z) - D^{*2}(X)))$$

s'écrit sous la forme (18 bis). ¶

La droite des constantes a pour vecteur  $D_P$ -unitaire le vecteur  $\mathbf{1}$  de coordonnées constantes et égales à 1. Par suite le projecteur  $D_P$ -orthogonal sur  $F_1$  s'écrit

$$Q_1 = I - \mathbf{1} \mathbf{1}' D_P$$

Le rôle de l'opérateur  $Q_1$  dans la construction d'une représentation Euclidienne de  $I$  est essentiel. Ainsi le critère (18 bis) définit un problème proche de celui que traitent YOUNG et al. (1976) en termes de minimisation de la moyenne des carrés des déformations des distances interpoints :

$$\begin{aligned} \text{SSTRESS} &= \text{Trace} ((\Delta^{*2}(Z) - D^{*2}(Z))^2) \\ &= \|\delta^2(Z) - d^2(X)\|_{I \otimes I}^2 \end{aligned} \quad (19)$$

d'AUBIGNY (1984) et MEULMAN (1986) montrent que les minimisations de (18) et (19) sont deux problèmes distincts conduisant à des solutions dont les propriétés sont différentes puisque (19) fournit une approximation bilatérale des dissimilarités. Cette différence tient au rôle joué par l'opération de centrage  $Q_1$ .

Ainsi donc, le choix d'une origine, imposé par la volonté d'utiliser les outils du calcul vectoriel pour traiter un problème de géométrie affine, a pour conséquence de transformer

le problème initial d'approximation de distances en un problème d'approximation de produits scalaires non équivalent mais jugé de résolution plus aisée. GOWER (1982) discute d'autres choix d'origine, intéressants du point de vue géométrique, ainsi que de leur influence sur le plongement Euclidien de I.

Cette transformation du problème originel est génératrice de problèmes difficiles dans le cas général. Les matrices de dissimilarité observées  $\Delta$  sont alors le plus souvent non Euclidiennes. Dans ce cas, l'image de la matrice symétrique  $H = -\frac{1}{2} \Delta^2$  de terme général  $(-\frac{1}{2} \delta_{ij}^2)$  dans la transformation de Young et Householder (16) associée au projecteur  $Q_1$  est

$$W = Q_1 H Q_1 \quad \underline{w} = Q_1 \otimes Q_1 h$$

Bien que symétrique comme  $\Delta$ ,  $W$  n'est en général plus interprétable comme une matrice de produits scalaires entre vecteurs d'un espace Euclidien. Son analyse spectrale construit donc des axes de coordonnées complexes, difficiles à interpréter. Une deuxième voie de traitement consiste alors à recourir à une formalisation statistique de modèles d'erreur, décrivant la matrice de dissimilarité  $\Delta$  comme une observation bruitée d'une matrice de distances Euclidiennes. Le caractère non linéaire de la transformation de Wilkinson rend difficile le traitement statistique de tels modèles d'erreur dans le cadre de l'Analyse en Coordonnées Principales (cf. WALKER 1976). Nous traitons donc de façon directe le problème d'approximation sous-jacent.

### II.3. LE CODAGE MULTIDIMENSIONNEL AFFINE DU TRIPLE ( I , $\delta_I$ , $Q_{KI}$ )

La relation (19) est un cas particulier des techniques de Codage Multidimensionnel Affine préconisées par OBENCHAIN (1973), afin de construire des représentations graphiques de  $\Delta$ , invariantes par transformation affine de la configuration de départ :

$$Y = BZ + \underline{c} \mathbf{1}_P$$

où  $\underline{c} \in E_p$ , et  $B$  est une matrice  $p \times p$  de plein rang. Cet auteur montre alors que tout indice de la forme:

$$\psi(\alpha, \beta, \Delta) = \sum_{i=1}^n \sum_{j=1}^n q_{ij} \left| \delta_{ij}^\alpha - d_{ij}^\alpha(X) \right|^\beta \quad \alpha, \beta \in \mathbb{R}^+ \quad (20)$$

où  $q_{ij}$  est une pondération définie sur  $I \times I$ , vérifiant  $q_{ij} \geq 0$ , conduit à une méthode de Codage Multidimensionnel Affine, associé au modèle d'erreur

$$\delta_{ij}^\alpha = d_{ij}^\alpha(X) + \varepsilon_{ij} \quad \alpha \in \mathbb{R}^+$$

On suppose de plus sans perte de généralité que  $\sum q_{ij} = 1$ . Ainsi la fonction SSTRESS donnée en (19) vérifie pour une pondération constante

$$\text{SSTRESS} = \psi(2,2,\Delta(Z))$$

Cette identité reste vraie dans le cas d'une pondération  $Q_{I \times I}$  définie comme produit tensoriel  $P_1 \otimes P_1$ .

La minimisation du critère (19) conduit cependant à un problème numérique d'apparence plus compliquée que celle de la fonction STRAIN (18). En effet dans ce dernier cas, on oublie en général la complexité algorithmique de la construction des éléments propres de  $WD_p$  pour n'en retenir que l'interprétation algébrique en termes de changement de base dans un espace vectoriel Euclidien. La minimisation de la fonction SSTRESS (19) ne présente pas de solution explicite puisque l'écriture des équations normales conduit à la résolution d'un système d'équations implicites non linéaires en les paramètres, que l'on doit résoudre de façon itérative. Il en est de même pour l'ensemble des indices de la forme  $\psi(\alpha, \beta, \Delta)$ , et en particulier de la fonction STRESS =  $\psi(1,2,\Delta)$  (KAMENSKI (1978, 1979), GOWER (1982), d'AUBIGNY (1984), MEULMAN (1986)). Nous exposons plus bas dans le cadre algébrique proposé par d'AUBIGNY (1984) pour l'étude du triple  $(I, \delta_I, Q_{I \times I})$ , l'algorithme proposé par de LEEUW et HEISER (1977, 1980) pour résoudre ce problème.

### II.3.1. LES OUTILS ALGÈBRIQUES D'ANALYSE DU TRIPLE $(I, \delta_I, Q_{I \times I})$

Considérons une pondération  $Q_{I \times I}$  définie sur  $I \times I$ , à valeurs réelles, et vérifiant pour tout  $(i,j) \in I \times I$

$$(C1) \quad q_{ij} = q_{ji}$$

$$(C2) \quad q_{ij} \geq 0$$

Posons de plus :

$$q_{i+} = \sum_{j \neq i}^n q_{ij} \quad i \in I$$

On peut alors interpréter simplement du point de vue géométrique le lemme suivant dû à COXETER (1948) :

**Lemme 1:** Soit  $Q$  une matrice carrée d'ordre  $n$  vérifiant les conditions (C1) et (C2). On a pour tout élément  $x$  de  $F$

$$\sum_i \sum_j q_{ij} (x_i - x_j)^2 = 2 \left[ \sum_i q_{i+} x_i^2 - \sum_i \sum_j q_{ij} x_i x_j \right] \quad (21)$$

**Preuve:** Elle résulte d'une double application de l'égalité suivante

$$\begin{aligned} \sum_i \sum_j q_{ij} x_i x_j &= \sum_j \left[ (q_{j+} - \sum_i q_{ij}) x_j - \sum_i q_{ij} x_i \right] x_j \\ &= \sum_j q_{j+} x_j^2 - \sum_i \sum_j q_{ij} (x_j - x_i) x_j \end{aligned}$$

dans laquelle on permute le rôle de  $i$  et  $j$  dans le second terme de droite. On obtient ainsi le résultat souhaité.  $\square$

La relation (21) s'écrit encore dans  $F$

$$\|x\|_N^2 = \frac{1}{2} \sum_i \sum_j q_{ij} y_{ij}^2 = \sum_i \sum_{j>i} q_{ij} y_{ij}^2 \quad (22)$$

en notant  $y_{ij} = x_i - x_j$  le terme général de la matrice

$$Y = x \mathbf{1}' - \mathbf{1} x'$$

Remarquons de plus que  ${}^t Y = -Y$  et donc  $Y$  définit une forme bilinéaire alternée sur  $F$ . On pose alors

$$N = D_q - Q \quad Q = (q_{ij})_{i,j \in I}$$

$$n_{ij} = \begin{cases} -q_{ij} & \text{si } i \neq j \\ \sum_{k \neq i} q_{ik} & \text{si } i = j \end{cases} \quad \begin{aligned} D_q &= \text{diag}(q_{i.})_{i \in I} \\ D_Q &= \text{diag}(q_{ij})_{i,j \in I} \end{aligned}$$

de sorte que si  $y = \text{Vec}(Y)$ :

$$\|x\|_N^2 = \|y\|_{D_Q}^2$$

**Remarque :** La contrainte (C1) n'est pas fondamentale ici. On l'impose afin d'assurer le caractère symétrique de la matrice  $N$  que l'on interprète alors comme une matrice de produits scalaires sur  $F$ .

L'étude de la matrice antisymétrique  $Y$  conduit à expliciter les relations liant  $F$  à l'espace vectoriel produit extérieur de  $F$  par lui-même, noté  $F \wedge F$ .  $F \wedge F$  est un sous-espace vectoriel

de dimension  $n^* = \frac{n(n-1)}{2}$  du produit tensoriel  $F \otimes F$  que l'on peut décrire à l'aide des vecteurs de base  $f_i \wedge f_j$  associés à la base  $(f_i, i \in I)$  de  $F$  par la relation :

$$f_i \wedge f_j = f_i \otimes f_j - f_j \otimes f_i \quad i, j \in I \quad i < j$$

Pour tout  $(i, j) \in I \times I, i \neq j$ ,  $f_i \wedge f_j$  définit une application bilinéaire alternée, dont la matrice dans la base  $(f_i \otimes f_j, i, j \in I)$  de l'espace  $F \otimes F$  a pour seuls termes non nuls le terme  $(i, j)$  égal à 1 et le terme  $(j, i)$  égal à -1.

**Remarque :** D'un point de vue statistique, les éléments  $f_i \wedge f_j$  fournissent une interprétation algébrique aux codages classiquement utilisés lors du calcul des statistiques liées à l'indice TAU B proposé par Kendall pour mesurer l'association entre variables ordinales. Les codages de la forme  $Y = (X^t \mathbf{1} - \mathbf{1}^t X)$  offrent une première généralisation des codages de Kendall, à partir de laquelle on peut exprimer tous les indices d'association se déduisant de l'indice de corrélation linéaire, par transformation des données. Plus généralement, les codages  $a_{ij}$  qui s'expriment comme terme général d'une matrice antisymétrique d'ordre  $n$  et qui fondent l'étude de l'association entre variables développée par DANIELS (1944), se décomposent tous dans cette base de  $F \wedge F$ . Lorsqu'on munit cet espace d'une géométrie Euclidienne induite par une métrique  $\Phi$ , on obtient une interprétation géométrique des travaux de cet auteur ainsi qu'une généralisation immédiate de ses résultats au cas de procédures d'échantillonnage complexes ( $\Phi$  est alors associée à une mesure de dénombrement) ou de données manquantes.

La correspondance algébrique entre  $F$  et  $F \wedge F$  est étudiée par d'AUBIGNY (1984). Elle est résumée par le schéma de dualité suivant:

$$\begin{array}{ccc}
 F^* & \xleftarrow{t} & (F \wedge F)^* \\
 \uparrow N & & \uparrow \Phi \\
 F & \xrightarrow{t} & F \wedge F
 \end{array}$$

caractérisé par la relation

$$t(f_i^* \wedge f_j^*) = f_i^* - f_j^* \quad (23)$$

où de façon classique  $f_i^*$  et  $f_i^* \wedge f_j^* = (f_i \wedge f_j)^*$  désignent les vecteurs des bases duales de  $(f_i; i \in I)$  et  $(f_i \wedge f_j; (i,j) \in I \times I)$  respectivement.  $t$  est représentée dans ces bases par la matrice  $T$  définie par:

$${}^tT = \begin{bmatrix} & S_1 & & & \\ & 0 & S_2 & & \\ & & & \ddots & \\ 0 & & 0 & & S_{n-1} \end{bmatrix} \quad S_i = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ \vdots & & & & \\ \vdots & & & & \\ \vdots & & & -1 & 0 \\ 1 & & & 0 & -1 \end{bmatrix}$$

où  $S_i$  est une matrice de simples contrastes entre les  $(n-i)$  dernières coordonnées d'un vecteur  $x \in F$ . Elle est donc de taille  $(n-i) \times (n-i+1)$ . Par suite la matrice  $T$  est d'ordre  $n \times n^*$  et vérifie l'identité

$${}^tT x = y \quad (24)$$

où  $y$  donne les coordonnées du vecteur de  $F \wedge F$  associé à la matrice  $x \mathbf{1} - \mathbf{1} x$  puisque :

$$y_{ij} = x_i - x_j \quad \text{si } i < j.$$

Le plongement du problème dans le sous espace vectoriel  $F \wedge F$  de  $F \otimes F$  consiste en pratique à négliger les informations apportées par l'un des termes  $(i,j)$  et  $(j,i)$  redondants du fait de l'antisymétrie.

On note dans la suite  $F_0$  (resp.  $F_0^*$ ) le sous-ensemble de  $F$  (resp.  $F$ ) orthogonal pour la métrique identité à la droite des constantes  $\Delta_{1F}$  (resp.  $\Delta_{1F^*}$ ).

$$F = F_0 \oplus \Delta_{1F} \quad F^* = F_0^* \oplus \Delta_{1F^*}$$

Puisque les  $(n-1)$  vecteurs  $(f_i^* - f_j^*)$  sont linéairement indépendants de  $\mathbf{1}_{F^*}$  et forment une base de  $F_0^*$ , la relation (23) montre que  $t$  définit une isométrie entre  $(F \wedge F)^*$  et le sous-espace  $F_0^*$ . Lorsqu'on munit  $F$  de la géométrie Euclidienne induite par la métrique définie par la forme quadratique de rang au plus égal à  $(n-1)$

$$N = T\Phi {}^tT \quad (25)$$

$\mathbf{1}_F$  est un élément du noyau de  $N$ .  $N$  agit donc sur  $F_0$ , assurant une isométrie partielle entre  $(F_0, N)$  et  $(F \wedge F, \Phi)$ .

La pratique de l'analyse des données conduit à choisir des métriques  $\Phi$  particulières. Ainsi la minimisation de l'indice proposé par OBENCHAIN suppose le choix d'une métrique des poids, décrite dans  $F \wedge F$  par la matrice diagonale  $\Phi = D_Q$ , dont l'interprétation statistique générale est abordée plus loin. Il lui correspond par l'application  $t$ , la matrice  $N = (D_q - Q)$  décrite par le lemme de Coxeter. Deux cas particuliers sont d'utilisation courante en statistique :

$$(a) \quad q_{ij} = p_i p_j \quad p_i > 0 \quad \sum p_i = 1 \quad (\text{i.e. } Q = P \otimes P)$$

Alors  $q_{i+} = p_i$  et il correspond à  $Q$  une matrice  $N_Q = T D_Q {}^t T$  définie par

$$n_{ij} = \begin{cases} p_i (1 - p_i) & \text{si } i = j \\ - p_i p_j & \text{si } i \neq j \end{cases}$$

Interprétons  $p_i$  comme la probabilité de l'évènement  $\{i\}$ , alors la donnée d'une mesure tensorielle correspond à une hypothèse d'indépendance des unités statistiques et se traduit par une matrice  $N_Q$  associée au modèle multinomial défini par  $(p_1, \dots, p_n)$ . La matrice  $N_Q$  s'écrit alors formellement

$$N_Q = D_P (I - \mathbf{1} \mathbf{1}^t D_P) = D_P Q_1 \quad (26)$$

où  $Q_1$  est le projecteur  $D_P$ -orthogonal sur le supplémentaire de  $\Delta_{1P}$  défini plus haut.  $N_Q$  est alors égale à la matrice  $N_P$  utilisée dans l'expression (18 bis).

(b) en particulier dans le cas équipondéré ( $p_i = \frac{1}{n}$  pour tout  $i$ )

$$n_{ij} = \frac{1}{n} \left( 1 - \frac{1}{n} \right) = \frac{n-1}{n^2}$$

et donc

$$N = \frac{1}{n} \left( I - \frac{1}{n} \mathbf{1} \mathbf{1}^t \right) = \frac{1}{n} Q_0 = \frac{1}{n} I_n Q_0$$

cette égalité montre de plus que l'on peut factoriser l'opérateur  $Q_0$  en fonction de la matrice  $T$ :

$$Q_0 = T {}^t T$$

En notant  $N_Q$  la métrique  $N$  induite par les pondérations  $q_{ij}$ , l'isométrie déterminée par l'application  $t$  permet de réécrire la formule (22) sous la forme condensée:

$$\|x\|_{N_Q \otimes I}^2 = \|y\|_{D_Q}^2$$

Les conséquences de cette égalité pour la définition d'algorithmes de Codage Multidimensionnel sont traitées dans d'AUBIGNY (1984). En particulier, cette égalité se généralise par additivité au traitement d'une matrice de configuration  $X$  quelconque, c'est à dire à l'étude des quantités :

$$\begin{aligned}
\eta^2(X) &= \sum_{i < j} \sum q_{ij} d_{ij}^2(X) = \sum_{s=1}^p \left[ \sum_{i < j} q_{ij} (x_i^s - x_j^s)^2 \right] \\
&= \sum_{s=1}^p \| \underline{x}^s \|_N^2 = \text{Trace}(XN'X) \\
&= \text{Trace}(W_X N)
\end{aligned} \tag{27}$$

où l'on pose  $W_X = 'XX$ . On vérifie de plus que

$$\text{Trace}('XXN) = \| \underline{x} \|_{N \otimes I}^2$$

où  $\| \underline{x} \|_{N \otimes I}$  désigne la norme de Hilbert Schmidt de l'opérateur  $x : F^* \rightarrow E$  regardé comme un élément de  $F \otimes E$ , muni de sa structure Euclidienne canonique associée à la semi métrique  $N \otimes I$ .

**Remarque 1 :** Les outils de calcul tensoriel, repris par exemple de HENDERSON et SEARLE (1979) permettent une étude simplifiée des matrices de configuration  $X$ , construites par les méthodes d'analyse des tableaux de dissimilarité. Ces outils sont de plus complétés, lorsque  $E=F$ , par l'étude des opérateurs symétriques définis sur  $F$ . On est alors conduit le plus souvent à identifier un opérateur symétrique  $A$  avec son image dans le sous espace vectoriel  $F \vee F$  de  $F \otimes F$  de dimension  $n(n+1)/2$ , formé des tenseurs symétriques pour lesquels  $a_{ij} = a_{ji}$ . On associe alors à la matrice symétrique  $A : F^* \rightarrow F$ , le vecteur  $\underline{a} \equiv \text{Vecs}(A)$ . De façon complémentaire, à chaque matrice  $A$  à diagonale nulle et antisymétrique, est associée son image dans le sous-espace vectoriel  $F \wedge F$  de  $F \otimes F$  de dimension  $n(n-1)/2$ , formé des tenseurs antisymétriques pour lesquelles  $a_{ij} = -a_{ji}$ . On associe alors à la matrice antisymétrique  $A : F^* \rightarrow F$ , le vecteur  $\underline{a} \equiv \text{Veca}(A)$ . On rappelle de plus que les espaces  $F \vee F$  et  $F \wedge F$  sont deux sous-espaces vectoriels supplémentaires dans  $F \otimes F$ .

**Remarque 2 :** L'intérêt pratique du lemme de Coxeter repose en particulier sur l'hypothèse de symétrie de la pondération  $Q_{|x|}$ . Celle-ci n'est pas toujours vérifiée en pratique. On associe alors à  $Q_{|x|}$  une matrice symétrique définie comme la projection de  $Q$  sur l'espace  $F \vee F$  et à laquelle on applique le lemme de Coxeter. L'opération de symétrisation se traduit simplement de façon analytique, puisqu'on construit alors la matrice  $N$  de terme général:

$$n_{ij} = \begin{cases} -\frac{1}{2} (q_{ij} + q_{ji}) & \text{si } i \neq j \\ \sum_{k \neq i} \frac{1}{2} (q_{ik} + q_{ki}) & \text{si } i = j \end{cases}$$

Le cas de pondérations de signe quelconque est à notre connaissance beaucoup plus rare. On traite alors la projection de  $Q$  sur le cône des matrices semi-définies positives associées aux applications linéaires définies sur  $F$  à valeurs dans  $F^*$ , inclu dans  $F \otimes F$  afin d'associer à  $Q$  une métrique Euclidienne  $N$  sur  $F$ .

### II.3.2. PRÉSENTATION D'UN ALGORITHME DE CODAGE MULTIDIMENSIONNEL

Le problème du Codage Multidimensionnel traité de façon heuristique par SHEPARD (1962) fut formalisé par KRUSKAL (1964) en termes d'approximation au sens des moindres carrés. Cet auteur minimise en effet l'indice de STRESS qui mesure l'écart entre dissimilarités observées  $\underline{\delta}$  et distances reconstruites  $\underline{d}(X)$  au sens des moindres carrés. L'indice de STRESS est un cas particulier de la classe d'indices proposée par OBENCHAIN correspondant à:

$$\text{STRESS} = \sum_{i < j} \sum_{i < j} q_{ij} (\delta_{ij} - d_{ij}(X))^2 \quad (28)$$

$$= \sum_{i < j} \sum_{i < j} q_{ij} \delta_{ij}^2 + \sum_{i < j} \sum_{i < j} q_{ij} d_{ij}^2(X) - 2 \sum_{i < j} \sum_{i < j} q_{ij} \delta_{ij} d_{ij}(X) \quad (29)$$

Cette quantité s'écrit dans  $F \wedge F$

$$\text{STRESS} = \| \text{Veca}(\Delta) - \text{Veca}(D(X)) \|_{D_Q}^2 \quad (28 \text{ bis})$$

$$= \| \underline{\delta} - \underline{d}(X) \|_{D_Q}^2 \quad (28 \text{ ter})$$

$$= \| \underline{\delta} \|_{D_Q}^2 + \| \underline{d}(X) \|_{D_Q}^2 - 2 \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q} \quad (29 \text{ bis})$$

où  $D_Q$  est la matrice diagonale de terme générique  $q_{ij}$ . KRUSKAL (1964) construit alors une configuration caractérisée par la matrice  $X$  réelle d'ordre  $rxn$ , à l'aide d'une procédure de minimisation de l'indice de STRESS, sous des contraintes techniques de normalisation. Il traite le problème numérique sous jacent à l'aide d'une méthode de gradient particulièrement élaborée quant au choix du pas d'itération (cf. KRUSKAL (1977)). Il étend de plus cette méthode au traitement de données de dissimilarité transformées ou disparités

$$t_{ij} = f(\delta_{ij}) \quad i, j \in I$$

où  $f$  est un élément d'une classe de fonctions  $\mathcal{F}$  déterminée par les contraintes de mesure liées soit au processus de collecte des données (on suppose par exemple que  $f$  est une fonction monotone par rapport à l'ordre induit par les indices de dissimilarité

observés, cf. KRUSKAL (1964)), soit au souci d'assurer la stabilité des résultats en retenant des fonctions  $f$  dépendant d'un petit nombre de paramètres (on suppose par exemple que  $f$  est une fonction polynomiale de degré fixé, cf. KRUSKAL (1964)), soit encore un cumul des deux objectifs (on suppose par exemple que  $f$  est une fonction spline cubique monotone, cf. WINSBERG et RAMSAY (1983)). Les disparités résultent alors d'un ajustement, le plus souvent au sens des moindres carrés, des dissimilarités  $\delta$  en fonction des distances  $d(X)$  calculées pour une configuration  $X$  donnée.

$$\text{Min}(\text{STRESS} ; \mathfrak{I} \in \mathfrak{F})$$

$$\text{STRESS} = \|\mathfrak{I} - \underline{d}(X)\|_{D_Q}^2$$

De nombreuses propositions d'algorithmes ont suivi les travaux de KRUSKAL et de SHEPARD et nous présentons ici l'approche de de Leeuw et Heiser (cf. de LEEUW (1977) et de LEEUW et HEISER (1980)) du fait de ses propriétés de convergence connues et de ses relations aux travaux antérieurs de GUTTMAN (1968). Ces auteurs montrent que l'utilisation de méthodes de gradient n'est pas satisfaisante pour minimiser la fonction STRESS en fonction de  $X$  puisque cette fonction n'est pas différentiable aux points  $X$  pour lesquels il existe au moins un couple  $(i,j)$  d'éléments de  $I$ ,  $i \neq j$ , tel que  $d_{ij}^2(X) = 0$ . Ils généralisent en conséquence la démarche usuelle au cadre de l'optimisation convexe non différentiable et proposent un algorithme appartenant à la classe des méthodes de sous-gradient. En effet toute fonction de distance de la forme

$$d(X) : (i,j) \in I \times I \rightarrow d_{ij}(X) = \mu(\underline{x}_i - \underline{x}_j)$$

où  $(\underline{x}_i - \underline{x}_j)$  est le terme général de la matrice  $TX$ , est convexe et positivement homogène lorsque  $\mu(\cdot)$  est une jauge. C'est en particulier vrai pour la distance Euclidienne, utilisée ici et donc si l'on écrit le produit scalaire intervenant dans (1)

$$\rho(X) = \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(X) = \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q}$$

sous la forme équivalente:

$$\rho(X) = \sum_{i < j} \sum q_{ij}(X) d_{ij}^2(X) = \|\underline{d}(X)\|_{D_{Q(X)}}^2$$

on obtient une fonction convexe, positivement homogène, associée à la matrice diagonale  $D_{Q(X)}$ , définie à partir de la matrice symétrique et à termes positifs  $Q(X)$  de terme général

$$q_{ij}(X) = \begin{cases} q_{ij} \delta_{ij} \frac{1}{d_{ij}(X)} & \text{si } d_{ij}(X) \neq 0 \\ 0 & \text{sinon} \end{cases} \quad (30)$$

$Q(X)$  est une matrice symétrique lorsque  $D$  et  $Q$  le sont. De plus, à pondération  $q_{ij}$  constante, le terme  $q_{ij}(X)$  pondère la paire  $(i,j)$  d'autant plus fortement (resp. faiblement) que  $d_{ij}(X)$  sousestime (resp. surestime) la dissimilarité  $\delta_{ij}$ .  $Q(X)$  permet de plus de définir une forme quadratique semi définie positive sur  $F$ , à l'aide du lemme de Coxeter. On munit ainsi  $F$  d'une géométrie Euclidienne locale indicée par  $X$ , pour laquelle on peut écrire

$$\begin{aligned} \rho(X) &= \text{Trace}(X N(X) X) \\ &= \|X\|_{N(X)}^2 \end{aligned} \quad (31)$$

où  $N(X)$  est construite à partir de la pondération (30), traduite par la matrice diagonale opérant sur  $F \wedge F$

$$D_{Q(X)} = D_Q D_\Delta D_{D(X)} \quad (32)$$

#### Remarques:

(a) L'expression (30) met en lumière le rôle joué par les termes  $d_{ij}(X)$  nuls dans les problèmes de non différentiabilité de la fonction STRESS. La formule (32) utilisée par la méthode de LEEUW et HEISER recourt en effet à un choix particulier d'inverse généralisée  $D_{D(X)}^-$ . Ce choix est unique lorsque le sous-différentiel  $\partial\rho(X)$  est réduit à un point qui est alors le gradient  $\nabla\rho(X)$ , c'est à dire lorsque aucun terme  $d_{ij}(X)$  n'est nul et donc

$$D_{D(X)}^- = D_{D(X)}^{-1}$$

(b) Lorsque l'une des deux matrices  $\Delta$  ou  $Q$  n'est pas symétrique, on doit projeter la matrice  $D_Q D_\Delta$  sur le cône des matrices symétriques opérant sur  $F \wedge F$  afin d'assurer la symétrie de  $N(X)$  et par conséquent la pertinence du modèle Euclidien.

La fonction  $\rho(\cdot)$  définie en (31) est convexe et positivement homogène puisque  $d(\cdot)$  l'est. On peut donc utiliser un algorithme de programmation convexe pour maximiser cette fonction de  $X$  et ainsi minimiser la fonction STRESS, en remarquant que le premier terme de droite de l'équation (1) est fixé par les données.  $\rho(X)$  est une norme finie d'après (31) et par suite (cf. ROCKAFELLAR (1970)) son sous-différentiel  $\partial\rho(X)$  est un sous ensemble convexe non vide et compact du dual algébrique  $F_0^* \otimes E^*$  de  $F_0 \otimes E$  (ou à un isomorphisme près, de l'ensemble des applications linéaires définies sur  $F_0^*$  et à valeurs dans  $E$ ). De plus  $Y$  est par définition un élément de  $\partial\rho(X)$  si et seulement si l'on a pour toute configuration  $z$  de  $F_0 \otimes E$ :

$$\rho(Z) \geq \rho(X) + \langle Y, Z-X \rangle$$

où l'on note

$$\langle A, B \rangle = \text{Trace} ({}^tAB) = \langle \underline{a}, \underline{b} \rangle_{I \otimes I}$$

Par suite le lemme suivant, dû à de LEEUW (1977), permet de vérifier que  $Y = N(X) {}^tX$  est un sous-gradient de  $\rho$  au point  $X$ , c'est à dire un élément de  $\partial\rho(X)$ . On note en effet qu'ici  $M$  est la matrice identité, et donc  $Y$  est la matrice d'une application linéaire de  $E$  dans  $F_0^*$ .

**Lemme 2:** Pour tous  $X, Z \in F_0 \otimes E$ , on a

$$\rho(X) \geq \text{Trace}(X N(X) {}^tZ)$$

ou de façon équivalente

$$\rho(Z) \geq \langle \underline{x}, \underline{z} \rangle_{N(X) \otimes I}$$

**Preuve:** l'inégalité de Cauchy-Schwarz assure que

$$d_{ij}(Z) d_{ij}(X) \leq \sum_{s=1}^p (x_i^s - x_j^s)(z_i^s - z_j^s)$$

et donc en notant  $d_{ij}(X)$  le terme générique de  $D_{D(X)}$  :

$$\sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(Z) \leq \sum_{i < j} \sum q_{ij} \delta_{ij} d_{ij}(X) \left[ \sum_s (x_i^s - x_j^s)(z_i^s - z_j^s) \right]$$

qui conduit au résultat après arrangement des termes de droite. ¶

**Remarque 1 :** La complexité des calculs mis en œuvre lors de la minimisation de la fonction STRESS provient du fait que la métrique  $N(X)$  dépend du point  $X$ . En effet, pour une matrice symétrique semi-définie positive  $A$  fixée, le gradient de la semi-norme

$$\eta^0(Z) = \text{Trace}(Z A {}^tZ)$$

définie sur un espace vectoriel  $F_0 \otimes E$  est classiquement :

$$\nabla \eta^0(Z) = A {}^tZ \tag{33}$$

**Remarque 2 :** Le recours au formalisme tensoriel montre que l'extension des résultats de de LEEUW au cas général caractérisé par une métrique  $M$  quelconque, ne présente en fait aucune difficulté particulière. On étudie alors le produit scalaire de  $\underline{x}$  et  $\underline{z}$  dans la métrique  $M \otimes N(X)$ .

Dans  $F_0 \otimes F_0$  nous pouvons réécrire la relation (28) sous la forme (28-ter) qui explicite la nature statistique du problème traité par de LEEUW et HEISER :

$$\text{STRESS} = \| \underline{\delta} - \underline{d}(X) \|_{D(Q)}^2 \quad (28\text{-ter})$$

où  $\underline{\delta}$  est le tenseur de terme général  $\delta_{ij}$  et  $D(Q)$  est la matrice diagonale de terme général  $q_{ij}$ . De plus, du fait de l'homogénéité de la fonction distance, minimiser la fonction STRESS par rapport à  $X$  équivaut à minimiser la quantité

$$S(X, \beta) = \| \underline{\delta} - \beta \underline{d}(X) \|_{D(Q)}^2 \quad (34)$$

en fonction de  $(X, \beta)$ , avec  $\beta \in \mathbb{R}$ . Pour  $X$  fixé, on doit donc résoudre un problème classique de régression aux moindres carrés. Le minimum en  $\beta$  est atteint pour

$$b(X) = \frac{\langle \underline{\delta}, \underline{d}(X) \rangle_{D(Q)}}{\| \underline{d}(X) \|_{D(Q)}^2} = \frac{\rho(X)}{\eta^2(X)} \quad (35)$$

et le minimum de  $S(X, \beta)$  vaut alors

$$S(X, b(X)) = \| \underline{d}(X) \|_{D(Q)}^2 (1 - r^2(X)) \quad (36)$$

où nous avons noté  $r(X)$  le coefficient de corrélation linéaire entre  $\underline{\delta}$  et  $\underline{d}(X)$ :

$$r(X) = \frac{\langle \underline{\delta}, \underline{d}(X) \rangle_{D(Q)}}{\| \underline{\delta} \|_{D(Q)} \| \underline{d}(X) \|_{D(Q)}} = \frac{\rho(X)}{\| \underline{\delta} \|_{D(Q)} \eta(X)} \quad (37)$$

$\| \underline{d}(X) \|_{D(Q)}$  est alors une donnée du problème. Par suite, dans le cadre général du Codage Multidimensionnel, la minimisation du STRESS est déterminée par la propriété suivante, qui découle immédiatement de (35), (36) et (37).

**Propriété :**

- (1) si  $X^*$  minimise le STRESS  $S(X, b)$ , elle maximise  $r(X)$ .
- (2) si  $X$  maximise  $r(X)$ , le minimum de  $S(X, b)$  en  $(X, b)$  est atteint en:

$$X^* = \frac{\rho(X)}{\eta^2(X)} X \quad (38)$$

Par conséquent, la minimisation du STRESS est équivalente à la recherche de la configuration  $X$  qui réalise le maximum du rapport de deux normes:

$$q(X) = \frac{\rho(X)}{\eta(X)} = \frac{\text{Trace}(X N(X) {}^t X)}{\sqrt{\text{Trace}(X N_Q {}^t X)}}$$

L'analogie du problème usuel traité avec celui de l'Analyse en Composantes Principales est immédiate. Cependant, la complexité de ce problème résulte ici du fait que la norme  $\rho(X)$  fait intervenir une géométrie locale  $N(X)$  dépendant de  $X$ . De plus l'Analyse en Composantes Principales optimise un rapport de Rayleigh homogène au rapport

$$\frac{\rho(X)}{\eta^2(X)} \text{ et non } \frac{\rho(X)}{\eta(X)}.$$

Cependant les résultats de CROQUETTE (1980) permettent de montrer l'identité des solutions de ces deux problèmes en Analyse en Composantes Principales.

On résoud ce problème en utilisant un algorithme proposé par ROBERT (1967) qui se ramène à l'algorithme de la puissance itérée dans le cas où  $N(X)$  ne dépend plus de  $X$ . Ses propriétés de convergence sont connues (cf. ROBERT (1967), de LEEUW (1977), de LEEUW et HEISER (1980)).

**Algorithme :**

à l'itération (0) initialisation :  $X^{(0)}$  tel que  $\eta^2(X^{(0)}) = 1$

à l'itération (v)  $Y^{(v)} \in \partial\rho(X^{(v)})$

$$X^{(v+1)} \in \partial\eta^0(Y^{(v)})$$

$$\text{normalisation } \eta^2(X^{(v+1)}) = 1$$

où  $\eta^0(\cdot)$  est la jauge polaire de  $\eta(\cdot)$ . Ici  $\eta(X)$  est une norme Euclidienne, dont la norme polaire est définie de façon classique par la métrique inverse (généralisée)  $N_Q^-$ . par suite les résultats exposés plus haut montrent que l'on peut choisir

$$Y^{(v)} = N(X^{(v)}) {}^tX^{(v)}$$

$$Z^{(v+1)} = N_Q^- Y^{(v)} = N_Q^- N(X^{(v)}) {}^tX^{(v)}$$

Soit, après normalisation

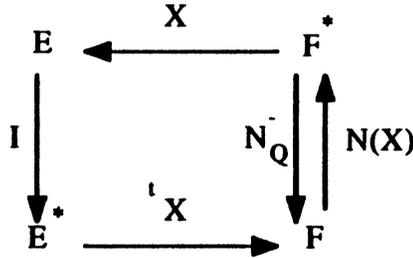
$${}^tX^{(v+1)} = \frac{1}{\|Z^{(v+1)}\|_{N_Q^-}} N_Q^- N(X^{(v)}) {}^tX^{(v)}$$

**Remarque 1 :** Au voisinage du minimum,  $\delta_{ij}$  est peu différent de  $d_{ij}(X)$  et donc  $N(X)$  est proche de  $N_Q$ .

**Remarque 2 :** La solution  $X$  est de plus une solution stationnaire de l'équation

$${}^tX = \frac{1}{\|Z\|_{N_Q^-}} N_Q^- N(X) {}^tX \quad (40)$$

où  $N_Q N(X)$  ne s'interprète pas, à notre connaissance, comme un opérateur proportionnel au projecteur orthogonal sur le sous-espace vectoriel  $F_X$  de  $F$  engendré par  $X$ . La situation correspond alors au schéma de dualité



dans lequel puisque  $X$  est fixé,  $N(X)$  l'est aussi. On résout alors un problème classique aux valeurs propres, et les axes de la représentation s'interprètent de façon identique au cas de l'Analyse en Coordonnées Principales. Les liens existant entre les deux approches sont alors clairement établis en remarquant que les vecteurs propres de l'opérateur  $N_Q N(X)$  associés à la solution optimale fournie par l'algorithme de LEEUW et HEISER, sont aussi les vecteurs propres de l'opérateur  $W_X N_Q$  où  $W_X = N(X)$ . L'analyse du triple  $(I, \delta_I, Q_{[X]})$ , exprimée dans  $F$  en termes d'optimisation de la fonction STRESS fait donc intervenir des outils algébriques analogues à ceux qui interviennent en Analyse en Coordonnées Principales, la difficulté du traitement résultant en fait de l'impossibilité de fixer a priori la géométrie de  $F$ .

Deux conséquences résultent de la remarque 2. Tout d'abord, elle montre la possibilité de traiter de façon analogue par les deux méthodes, la classe particulière de contraintes décrites dans  $F$ . Celles-ci sont étudiées dans la suite de ce travail puis étendues à une classe plus générale de contraintes traitables par des méthodes de Codage Multidimensionnel seulement. Néanmoins l'algorithme de LEEUW partage les faiblesses reconnues des méthodes de sous-gradient liées à une certaine lenteur de convergence. La deuxième conséquence est négative puisque d'AUBIGNY (1984) montre que le fait que  $N(X)$  ne puisse être fixé a priori introduit une différence importante entre les deux types de méthodes. En effet, la démonstration de la propriété d'emboîtement des sous-espaces de représentation solutions du problème d'Analyse en Coordonnées Principales en dimensions  $r$  et  $r'$  (par exemple pour  $r < r'$ ) utilise des arguments analogues à ceux qui sont classiquement employés pour l'Analyse en Composantes Principales (cf. PAGES et CAILLEZ (1976), d'AUBIGNY (1984)), dont la validité repose sur la connaissance a priori de la géométrie  $N$  définie sur  $F$ . On ne peut donc pas vérifier une telle propriété d'emboîtement des solutions dans le cas du problème d'optimisation posé en Codage Multidimensionnel puisqu'alors  $N$  est une fonction de la solution:  $N = N(X)$ .

**Remarque :** L'algorithme de de LEEUW et HEISER (1977) fournit une alternative satisfaisante aux méthodes de gradient, utilisées à tort pour optimiser le STRESS, puisque cette fonction n'est pas différentiable en certains points. Plus généralement, de tels problèmes de différentiabilité sont rencontrés lors de l'optimisation du critère général proposé par OBENCHAIN pour  $\alpha$  et  $\beta$  quelconques. Ils invalident ainsi les méthodes de gradient proposées par cet auteur. A notre connaissance, aucune étude du cas général n'a été entreprise malgré l'intérêt théorique de telles méthodes, en particulier du fait des problèmes de faible résistance aux données extrêmes qui caractérisent les méthodes de moindres carrés telles que celles qui optimisent les fonctions STRESS et SSTRESS.

### CHAPITRE III

#### L'ANALYSE DU TRIPLE $(I, \delta_I, Q_{KI})$ SOUS CONTRAINTES DE CONFIGURATION

Nous nous intéressons ici aux classes de contraintes portant directement sur la configuration  $X$  elle-même. Le cas plus général des contraintes relationnelles, exprimées à l'aide d'une ou de plusieurs matrices de dissimilarité est traité au chapitre IV.

L'introduction de contraintes portant sur la configuration est née des travaux expérimentaux menés dans le domaine de la psychophysique. En particulier NOMA et JOHNSON (1979) proposèrent de valider une théorie psychophysique de la perception, lors d'une expérience d'évaluation des similarités entre 16 ellipses construites suivant un plan d'expérience factoriel complet croisant deux facteurs:

la longueur du grand axe (GA) 1.3, 1.1, 0.9, 0.7

la longueur du petit axe (PA) 0.4, 0.3, 0.2, 0.1

Le codage multidimensionnel permet alors de décrire la surface réponse. Plus généralement, notons  $(F_l, l = 1, 2, \dots, L)$  l'ensemble des facteurs définissant un plan d'expérience,  $F_l$  est une variable qualitative à  $q_l$  modalités et soit  $q = \prod q_l$  le nombre de combinaisons de niveaux de facteurs ou patrons de réponse, parmi lesquelles on retient  $n$  patrons. Ceux-ci définissent les unités statistiques éléments de  $I$ . On évalue alors, par exemple lors d'une expérience de comparaisons par paires, la mesure de dissimilarité  $\Delta = (\delta_{ij})$ .  $\delta$  joue le rôle de variable réponse et la mise en œuvre d'une méthode de Codage Multidimensionnel permet de décrire les propriétés métriques d'une surface réponse estimée. L'originalité de cette approche est double. Tout d'abord on se propose ici de décrire la surface réponse à partir de propriétés métriques traduites par la donnée de  $\delta$ . Les facteurs de variabilité ne sont alors pas connus a priori. Cette méthode suit donc une démarche inverse de celle qui est rencontrée dans les plans d'expérience usuels pour laquelle la description approchée de la surface réponse est habituellement justifiée en termes de développements en série de Taylor d'une fonction d'autant de variables qu'il existe de facteurs contrôlés (cf. BRADU (1984)). De plus la variable réponse est évaluée ici par un procédé de comparaisons par paires conduisant en général à des observations  $\delta_{ij}$

non indépendantes. Dans le cas traité par NOMA et JOHNSON, le modèle s'écrit, sous l'hypothèse de non déformation des perceptions

$$\delta_{ij} = d_{ij}(Z) + \varepsilon_{ij} \quad (41)$$

La configuration Z respecte la structure du plan d'expérience si

$$Z = Y_1 B \quad (41\text{bis})$$

où  $Y_1$  est la matrice d'incidence du plan, et B la matrice des paramètres à estimer. Ainsi, dans le cas de l'expérience de NOMA et JOHNSON, croisant deux facteurs, la configuration bidimensionnelle obtenue par Codage Multidimensionnel doit reconstruire la grille du plan selon deux axes (qui ne sont pas forcément les axes de coordonnées principales). Les auteurs montrent qu'il en est ainsi pour trois des quatre sujets de l'expérience. Le Codage Multidimensionnel sans contrainte retrouve alors approximativement cette structure. Pour le quatrième sujet, l'introduction de contraintes s'impose (cf. NOMA et JOHNSON (1979)).

NOMA et JOHNSON introduisent des contraintes portant sur l'ordre des coordonnées des unités statistiques sur les axes de représentation. Ce n'est pas l'unique façon de traiter le problème, et BORG et LINGOES (1980) préfèrent décrire la structure du plan à l'aide de contraintes de type inégalités portant sur les distances inter unités statistiques. Leur analyse du problème relève du chapitre IV.

Lorsqu'il est interprété dans un cadre linéaire, c'est-à-dire lorsque l'on néglige l'aspect ordinal des contraintes posées par NOMA et JOHNSON, le problème traité se présente comme un cas particulier des modèles d'analyse sous contraintes d'un tableau à double entrée étudiés par CARROLL, GREEN et CARMONE (1976) et généralisés au cadre plurimodal, c'est-à-dire traitant le cas des tableaux à plus de deux entrées, par CARROLL et PRUZANSKI (1977).

### III.1. CONTRAINTES LINEAIRES PORTANT SUR LA CONFIGURATION X

#### III.1.1. L'APPROCHE DE CARROLL ET AL.: UNE FORMULATION EN TERMES DE RÉGRESSION

Nous étudions tout d'abord le modèle proposé par ces auteurs dans le cas où la donnée est un tableau à double entrée regroupant les mesures de p variables quantitatives sur n unités statistiques. Nous montrons ensuite quelles sont les adaptations possibles de leur solution au cas où la donnée est un tableau de dissimilarité.

Soit  $Z$  un tableau de données  $p \times n$ , et  $Y_1$  (resp.  $Y_2$ ) un tableau de variables externes traduisant les contraintes imposées aux  $p$  variables (resp. aux  $n$  unités statistiques).  $Y_1$  est un tableau  $p \times k_1$  contenant des mesures externes, effectuées sur les variables. On suppose que  $k_1 \leq p$  et que  $Y_1$  est de rang inférieur ou égal à  $k_1$ . Dans le cas d'un plan d'expérience,  $Y_1$  est construit de façon classique comme la matrice d'incidence du plan. Sa structure décrit les  $p$  combinaisons de facteurs qui contrôlent le choix des  $p$  variables étudiées. De même le tableau  $Y_2$ , de dimension  $n \times k_2$ , fournit soit des mesures externes caractérisant les unités statistiques, soit une matrice d'incidence du plan d'expérience qui contrôle le choix des unités statistiques. On suppose que  $k_2 \leq n$  et que  $Y_2$  est de rang inférieur ou égal à  $k_2$ . Le tableau  $Y_1$  (resp.  $Y_2$ ) définit des contraintes. Par suite la meilleure approximation  $X$  de dimension  $r$  de la configuration construite par une méthode d'analyse linéaire multidimensionnelle satisfaisant à ces contraintes, résulte formellement de deux approximations résultant respectivement de :

(1) la construction de  $\hat{Z}$  meilleure approximation de  $Z$  au sens des moindres carrés pour le modèle

$$Z = A_1 {}^t A_2 \quad (42)$$

$$A_1 = Y_1 B_1 \quad (43)$$

$$A_2 = Y_2 B_2 \quad (44)$$

(2) la construction d'un résumé  $r$  dimensionnel  $X$ , issu de l'analyse du tableau  $\hat{Z}$ .

Dans le cadre de l'analyse d'un tableau de dissimilarité Euclidienne,  $Z$  n'est pas connu car on dispose seulement du tableau de dissimilarité  $\Delta = \Delta(Z)$ . Les variables décrivant les unités statistiques ne sont pas connues à priori, on doit donc supposer que  $A_1$  est l'identité et on ne considère que la matrice de contraintes  $Y_2$  qui traduit les informations externes sur les unités statistiques. On ne retrouve cependant pas la restriction au cas linéaire du modèle de NOMA et JOHNSON qui impose des contraintes aux variables retenues pour représenter les unités statistiques sous la forme (41-bis). Nous traitons cependant ici le cas général. Notons

$$Z = A_1 {}^t A_2 = Y_1 B_1 {}^t B_2 {}^t Y_2 = Y_1 B {}^t Y_2 \quad (45)$$

où la matrice des coefficients  $B$  regroupe les paramètres non identifiables des matrices inconnues  $B_1$  et  $B_2$

$$B = B_1 {}^t B_2$$

Le choix d'un critère d'optimisation différencie les solutions proposées dans le passé. Nous montrons dans la suite que toutes celles qui traitent le problème dans le cadre de

l'Analyse en Coordonnées Principales conduisent à une même solution caractérisée comme une Analyse en Coordonnées Principales sur Variables Instrumentales par référence à l'Analyse en Composantes Principales sur Variables Instrumentales (ou ACPVI) au sens de RAO (1964), connue dans la littérature psychométrique sous le nom d'Analyse de Redondance, depuis les travaux de VAN DEN WOLLENBERG (1977).

CARROLL et al. (1976) se proposent de déterminer la configuration  $\hat{Z}$  qui réalise le minimum en T de la fonction critère suivante, mesurant l'écart entre configurations au sens des moindres carrés simples :

$$G(Z,T) = \text{Trace}((Z-T)'(Z-T))$$

Nous nous plaçons dans le cadre d'analyse un peu plus général d'une étude (Z, M, N) et nous optimisons le critère aux moindres carrés généralisés:

$$G(Z,T) = \text{Trace}((Z-T) N'(Z-T)M) = \|z - t\|_{M \otimes N}^2 \quad (46)$$

où nous cherchons  $\hat{Z}$  sous la forme (45).

$$\hat{Z} = Y_1 B' Y_2 \Leftrightarrow \hat{z} = Y_2 \otimes Y_1 b \quad \text{où } b = \text{Vec}(B)$$

nous sommes donc amenés à résoudre un problème de régression linéaire

$$z = Y_2 \otimes Y_1 b + \varepsilon \quad (47)$$

déterminé par la donnée de la variable réponse Z et par la matrice des variables explicatives (ou régresseurs)  $Y_2 \otimes Y_1$ . Celle-ci a une structure particulière. En effet elle ne vérifie les conditions d'indépendance des pseudo-observations (i,j) que dans des situations particulières. C'est par exemple le cas lorsque le tableau Z décrit le terme d'interaction lors de l'analyse d'un plan factoriel d'ordre 2 et est calculé par un double centrage du tableau initial destiné à retirer les effets principaux. Ces conditions d'indépendance ne sont pas vérifiées dans le cas qui nous intéresse et par suite l'application naïve des outils de statistique inférentielle à ce cadre n'est en général pas fondée.

**Proposition 1:** La solution Z réalisant le minimum du problème d'optimisation au sens des moindres carrés, donné par l'équation (46), sous la contrainte linéaire (45) est donnée par

$$\hat{Z} = P_1 Z' P_2 \quad (48)$$

où  $P_1$  et  $P_2$  sont les projecteurs  $M$  et  $N$  orthogonaux respectivement associés aux matrices définissant les contraintes  $Y_1$  et  $Y_2$ .

**Preuve :** La résolution des équations normales fournit la solution:

$$\hat{\mathbf{b}} = [{}^t(Y_2 \otimes Y_1) M \otimes N (Y_2 \otimes Y_1)]^{-1} {}^t(Y_2 \otimes Y_1) M \otimes N \mathbf{z}$$

où  $A^{-}$  est une inverse généralisée à gauche de  $A$ , et donc

$$\begin{aligned} \hat{\mathbf{b}} &= ({}^tY_2NY_2 \otimes {}^tY_1MY_1)^{-1} {}^tY_2N \otimes {}^tY_1M \mathbf{z} \\ \hat{\mathbf{b}} &= ({}^tY_2NY_2)^{-1} {}^tY_2N \otimes ({}^tY_1MY_1)^{-1} {}^tY_1M \mathbf{z} \end{aligned} \quad (49)$$

dont l'expression sous forme matricielle fournit une estimation des paramètres du modèle

$$\hat{\mathbf{B}} = ({}^tY_1MY_1)^{-1} {}^tY_1MZNY_2 ({}^tY_2NY_2)^{-1} \quad (49\text{-bis})$$

Le tableau approché  $\hat{\mathbf{Z}}$  se déduit immédiatement de la relation (49):

$$\begin{aligned} \hat{\mathbf{z}} &= Y_2 \otimes Y_1 \hat{\mathbf{b}} \\ &= P_2 \otimes P_1 \mathbf{z} \end{aligned} \quad (48\text{-bis})$$

soit sous forme matricielle

$$\hat{\mathbf{Z}} = P_1 \mathbf{Z} {}^tP_2$$

où  $P_1 = Y_1 ({}^tY_1MY_1)^{-1} {}^tY_1M$  est le projecteur  $M$ -orthogonal sur le sous-espace  $E_1$  de  $E$  défini par les contraintes  $Y_1$ . De la même façon,  $P_2 = Y_2 ({}^tY_2NY_2)^{-1} {}^tY_2N$  est le projecteur  $N$ -orthogonal sur le sous-espace  $F_2$  de  $F$  défini par les contraintes  $Y_2$ . On retrouve ainsi le rôle spécifique de  $Y_1$  qui agit sur l'espace  $E$ . ¶

$\hat{\mathbf{Z}}$  s'interprète donc comme la matrice associée dans les bases fixées de  $E$  et  $F$  à l'application linéaire

$$\hat{\mathbf{z}} : F_2^* \rightarrow E_1$$

dans les bases fixées de  $E$  et  $F$  où  $F_2$  est le sous-espace vectoriel de  $F$  image de  $G_2^*$  par  $Y_2$  et  $E_1$  est le sous-espace vectoriel de  $E$  image de  $G_1^*$  par  $Y_1$ .

CARROLL et al. (1976) fournissent une solution matricielle du problème dans le cas où  $M=I$  et  $N=I$ . Ils semblent ignorer l'analogie du problème traité avec l'ajustement du modèle proposé par POTTHOFF et ROY (1964) pour traiter les modèles de courbes de croissance dans le cadre de l'analyse de variance multidimensionnelle généralisée. Le tableau de données  $\mathbf{Z}$  est interprété comme l'observation sur l'échantillon  $I$  d'un vecteur aléatoire  $\mathbf{Z} = (Z^1, \dots, Z^P)$ . On suppose de plus que  $\mathbf{Z}$  appartient à une structure statistique

gaussienne de moyenne  $\mu$  et de matrice de covariance  $\Sigma$ . POTTHOFF et ROY (1964) définissent alors le modèle de courbes de croissance par l'expression des composantes de  $\mu$  comme combinaison linéaire d'un petit nombre de paramètres. La matrice aléatoire  $Z$  de  $i$ -ième colonne  $Z_i$  a pour espérance mathématique

$$E(Z) = Y_1 B {}^t Y_2$$

où les matrices  $Y_1 \in \mathcal{M}_{p,k_1}(\mathbb{R})$  et  $Y_2 \in \mathcal{M}_{n,k_2}(\mathbb{R})$  sont connues et de rang  $k_1 \leq p$  et  $k_2 \leq n$ .  $B$  est le paramètre inconnu. RAO (1966) montre alors que la matrice  $\widehat{B}$  définie par l'équation (49-bis) est l'estimateur de  $B$  au sens des moindres carrés, lorsque la matrice de covariance  $\Sigma$  définissant la structure statistique gaussienne est arbitraire, à condition de choisir

$$\begin{aligned} N &= I \\ M &= A^{-1} \\ A &= Z(I - Y_2({}^t Y_2 Y_2)^{-1} {}^t Y_2) {}^t Z \end{aligned}$$

On note pour la suite que seules les contraintes  $Y_2$  interviennent dans cette expression et que  $A$  s'interprète du point de vue statistique comme la matrice de covariance partielle empirique de  $Z$  (au coefficient  $1/n$  près), lorsque les variables  $Y_2$  sont contrôlées. Sous l'hypothèse gaussienne, KHATRI (1966) montre que  $\widehat{B}$  est aussi un estimateur de maximum de vraisemblance. RAO (1965, 1966, 1967) et KHATRI (1966) fournissent de plus les outils de base permettant de connaître la répartition d'échantillonnage de  $\widehat{B}$ , de construire les régions de confiance et de mettre en œuvre les tests de l'hypothèse

$$E(Z) = 0$$

On consultera de plus LEE et GEISSER (1972, 1975), TIMM (1979) et GEISSER (1980) pour une étude approfondie des propriétés statistiques de ce modèle. L'ensemble de ces résultats permet en effet d'enrichir la solution proposée par CARROLL et al. à l'aide des méthodes et des résultats de la statistique multivariée.

L'étude des propriétés de l'estimateur  $\widehat{B}$  fourni par l'équation (49-bis) avec des matrices  $M$  et  $N$  identités, noté ici  $\widehat{B}_1$  et retrouvé par CARROLL et al. a précédé les travaux de RAO et KHATRI (cf. GLESER et OLKIN (1970)).  $\widehat{B}_1$  est un estimateur sans biais de  $B$  qui permet aussi de construire des régions de confiance pour  $B$ . Il joue donc le même rôle que l'estimateur aux moindres carrés  $\widehat{B}_2$ , étudiés par RAO (1966). De plus GEISSER (1980) montre le lemme suivant:

**Lemme 1:**

$$\widehat{B}_2 = \widehat{B}_1 - ({}^t Y_1 Y_1)^{-1} {}^t Y_1 A U ({}^t U A U)^{-1} {}^t U Z Y_1 ({}^t Y_2 Y_2)^{-1}$$

où  $U$  est un élément quelconque de  $\mathcal{M}_{p,p-k_1}(\mathbb{R})$ , de rang  $p - k_1$  et tel que  ${}^tUY_1 = 0$ .

La comparaison de ces deux estimateurs, à l'aide de leurs matrices de covariance montre (GEISSER (1980)) que  $\widehat{B}_1$  est préférable lorsque

$$({}^tY_1Y_1)^{-1}\Sigma U = 0$$

ce qui est clairement le cas lorsque  $\Sigma = \sigma^2 I$  ou lorsque cette matrice est approximativement nulle. Plus généralement cette nullité est assurée lorsque  $\Sigma$  se décompose sous la forme classique

$$\Sigma = Y_1\Gamma {}^tY_1 + U\Theta {}^tU + \sigma^2 I$$

RAO (1967) montre que cette décomposition de  $\Sigma$  est une condition nécessaire et suffisante pour que  $\widehat{B}_1$  soit un estimateur aux moindres carrés et fournit un test d'hypothèse permettant de juger de l'adéquation de cette Structure de Covariance aux données traitées. La solution de CARROLL et al. traite en effet un cas particulier dont la pertinence n'est pas évaluée par ces auteurs.

L'introduction de métriques  $M$  et  $N$  quelconques, dans notre traitement du problème montre qu'une extension simple du modèle traité par ces auteurs permet de rendre compte de structures particulières de données, liées aux contraintes, dans un formalisme unificateur. De plus notre formulation du problème montre que la mise en œuvre de l'approche de CARROLL et al. ne nécessite pas l'écriture de logiciels particuliers.

**Remarque 1 :**  $P_1$  (resp.  $P_2$ ) est un projecteur défini sur  $E$  (resp.  $F$ ) et à valeurs dans le sous-espace vectoriel  $E_1$  (resp.  $F_2$ ) de  $E$  (resp.  $F$ ). On ne peut cependant pas interpréter  $P_2 \otimes P_1$  comme le projecteur  $N \otimes M$ -orthogonal de  $F \otimes E$  sur  $F_2 \otimes E_1$ . En effet, cet espace n'est pas un sous-espace vectoriel de  $F \otimes E$ . Si  $h$  est une forme bilinéaire définie sur  $F_2^* \times E_1^*$  (c'est-à-dire un élément de  $F_2 \otimes E_1$ ), elle n'est pas définie sur tout  $F^* \times E^*$ . Il existe cependant une infinité d'extensions  $k$  de  $h$  à  $F^* \times E^*$ .  $E$  et  $F$  étant munis d'une structure Euclidienne, nous avons adopté ici l'extension définie par

$$k(x,y) = h(P_2(x), P_1(y)) \quad \forall (x,y) \in F \times E$$

Il importe pour la suite de constater que cette "extension canonique" impose la connaissance de  $M$  et  $N$  qui interviennent dans la définition de  $P_1$  et  $P_2$ .

L'espace  $E = (\mathbb{R}^p, M)$  est une des inconnues du problème posé par l'analyse des tableaux de dissimilarité, rendant impossible la définition d'une extension unique. Par suite les propriétés d'emboîtement des sous-espaces solutions associés à des dimensions différentes, qui fondent de nombreuses méthodes d'analyse linéaire des données, telles

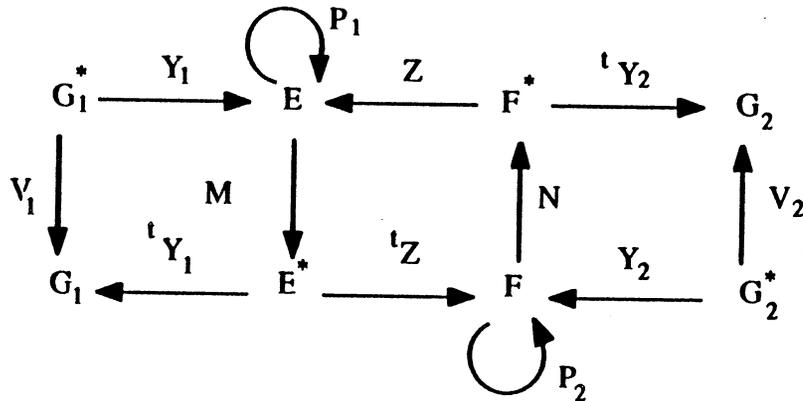
que l'Analyse en Composantes Principales ou l'Analyse Canonique, sont perdues pour cette classe de problèmes.

Lorsque M et N sont connues, l'application linéaire

$$\begin{aligned} \varphi : F_2 \otimes E_1 &\rightarrow F \otimes E \\ h &\mapsto k \end{aligned}$$

définit une bijection entre  $F_2 \otimes E_1$  et  $\varphi(F_2 \otimes E_1)$  qui est donc un sous-espace vectoriel de  $F \otimes E$  que nous notons  $\langle E_1, F_2 \rangle$ . On montre alors (d'AUBIGNY (1984), POLIT (1986), d'AUBIGNY et POLIT (1988)) que  $\varphi$  définit une isométrie de structures Euclidiennes et que  $P_2 \otimes P_1$  est le projecteur  $N \otimes M$ -orthogonal de  $F_2 \otimes E_1$  sur  $\langle E_1, F_2 \rangle$ .

L'ensemble des équations précédentes est alors décrit par le schéma algébrique suivant dans lequel on a posé  $V_1 = {}^t Y_1 M Y_1$  et  $V_2 = {}^t Y_2 N Y_2$ :



dans lequel  $\widehat{B}$  apparait comme la matrice de l'opérateur de transition  $\widehat{b} : G_2 \rightarrow G_1^*$  permettant de définir une isométrie entre  $(G_2, V_2^-)$  et  $(G_1, V_1^-)$ . Son analyse linéaire permet de décrire le potentiel d'interaction induit par la donnée de ces deux types de contraintes, par une Analyse en Composantes Principales classique du triplet  $(\widehat{B}, V_1, V_2)$ . La matrice  $Y_1$  définit la contrainte imposant une représentation des unités statistiques dans le sous-espace  $E_1 \subset E$ . On traite dans ce cadre une ACPVI menée dans l'espace des unités statistiques, c'est-à-dire une Q-ACPVI. De façon duale et plus classique la matrice  $Y_2$  définit la contrainte imposant une représentation des variables dans le sous-espace  $F_2 \subset F$ . On retrouve ainsi le problème usuel traité par une ACPVI menée dans l'espace des variables, c'est-à-dire une R-ACPVI. Par suite le cadre général du problème posé par CARROLL et al. peut être interprété comme celui d'une double ACPVI.

**Remarque 2 :** Le formalisme tensoriel montre le caractère canonique de l'imposition de telles contraintes linéaires de type égalité dans un cadre algébrique. De plus, une seule contrainte est imposée à chacun des espaces  $E_1$  de  $E$  et  $F_2$  de  $F$ . Dans le cadre de

l'Analyse en Composantes Principales sous contraintes linéaires, TEBBOUNE (1981) traite le cas d'une décomposition

$$E = \bigoplus_{l=1}^L E_l$$

où chaque sous-espace  $E_l$  correspond à une contrainte du type traité ici, ou bien

$$F = \bigoplus_{k=1}^K F_k$$

L'extension du travail de TEBBOUNE au traitement des deux décompositions additives simultanées permet une extension simple de l'approche de CARROLL et al. (1976) dans le cadre formel présenté ci-dessus. Bien que nous ne poursuivions pas cette étude ici, il convient de remarquer que les propriétés de distributivité de l'opération  $\otimes$  par rapport à l'opération  $\oplus$  conduit aux manipulations usuelles de sous-espaces vectoriels rencontrées en Analyse de la Variance. Les règles de calcul de projecteurs, décrites par exemple par MONLEZUN et SPEED (1980) appliquées aux sous-espaces de  $F \otimes E$ , s'appliquent alors sans difficultés hormis celles qui résultent des lourdeurs de notations.

Les contraintes (42 - 44) définissent un sous-modèle linéaire du modèle de description de  $I$  à partir de la configuration  $Z$ . On peut donc utiliser à titre indicatif les outils classiques de la statistique multivariée pour comparer les deux représentations obtenues. L'interprétation du problème posé par CARROLL et al. (1976) en termes de régression aux moindres carrés présente de plus un intérêt pratique, dû aux aides à l'interprétation et aux outils d'évaluation de la solution  $\hat{Z}$ , développés dans ce cadre.

Dans le cas  $N=I$ , COOK et WEISBERG (1982) montrent que le  $k$ -ième terme diagonal de la matrice du projecteur associé à la matrice  $Y$  des régresseurs, appelée usuellement Hat matrix et notée  $H$ , est compris entre  $1/n$  et  $1$  et s'écrit

$$(H)_{kk} = \frac{1}{n} + \frac{y_k^t (Y^t Y)^{-1} y_k}{n} \quad k = 1, 2, \dots, n$$

où  $y_k$  désigne la  $k$ -ème colonne de la matrice  $Y$ .  $(H)_{kk}$  s'interprète donc en terme de contribution de la  $k$ -ème observation à l'inertie globale du nuage de points définis par  $Y$ . Pour identifier les points extrêmes dans l'espace engendré par les régresseurs, ces auteurs proposent alors un indice appelé bras de levier (leverage) telle que le bras de levier de l'observation  $k$  est égale au  $k$ -ième terme diagonal de  $H$ . Dans le problème que nous traitons ici,  $H$  s'écrit

$$H = P_2 \otimes P_1$$

de  $ij$ -ème terme diagonal

$$h_{ij,ij} = (P_2)_{ii} \cdot (P_1)_{jj}$$

Par suite, le bras de levier de la case (i,j) du tableau  $Y_2 \otimes Y_1$  s'écrit comme le produit des bras de levier de la contrainte portant sur l'unité statistique i et de la contrainte portant sur la variable j.

D'autres mesures de bras de levier ont été proposées dans la littérature statistique, qui s'expriment toutes en fonction des termes diagonaux de H (cf. HARDWICK (1986), CHATTERJEE et HADI (1986)). Ainsi la distance de Mahalanobis de chaque case de  $Y_2 \otimes Y_1$  s'écrit ici

$$\text{Maha}_{ij} = (m - 1) \left( h_{ij,ij} - \frac{1}{m} \right)$$

où  $m = np$  est le nombre de pseudo-observations (i,j). On peut de plus comparer ces distances aux distances calculées à partir de  $P_1$  et de  $P_2$  respectivement.

De façon complémentaire, l'identification des points extrêmes dans l'espace engendré par le tableau Z est menée à l'aide de diagnostics portant sur les résidus (cf. CHATTERJEE et HADI (1986), HARDWICK (1986), HOAGLIN et WELSH (1978), VELLEMAN et WELSH (1981))

$$\hat{e}_{ij} = z_{ij} - \hat{z}_{ij} \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p$$

Par exemple, on calcule les résidus standardisés par l'expression (cf. HARDWICK (1986))

$$r_{ij} = \frac{\hat{e}_{ij}}{\hat{\sigma}(1 - h_{ij,ij})^{1/2}}$$

Enfin l'influence de cases influentes s'exprime de façon classique en fonction des bras de levier (CHATTERJEE et HADI (1986), HARDWICK (1986)). Par exemple la distance de Cook s'écrit ici

$$\text{Cook}_{ij} = \frac{\hat{e}_{ij} h_{ij,ij}}{\hat{\sigma}(1 - h_{ij,ij})^2}$$

et est donc aisément calculable à partir de  $P_1$ ,  $P_2$ ,  $Z$ ,  $\hat{Z}$  et de l'estimation classique  $\hat{\sigma}$  de l'erreur standard des résidus.

Ces fonctions diagnostiques prennent clairement en compte l'interaction  $y_2^j \otimes y_1^i$  entre les contraintes des deux groupes  $Y_1$  et  $Y_2$ . Leur consultation doit donc être prudente, compte tenu d'effets de structure possibles. De plus, le cadre exploratoire de notre étude ne pose pas de conditions d'indépendance des pseudo-observations (i,j).

L'évaluation globale de l'écart entre  $\hat{Z}$  et  $Z$  s'appuie sur la relation classique en régression multiple liant la variance résiduelle au carré du coefficient de corrélation linéaire.

**Propriété:**

$$\text{Trace}((Z-\hat{Z})N'(Z-\hat{Z})M) = \text{Trace}(W_Z N) \left[ 1 - \frac{\text{Trace}(ZN_2 {}^t ZM_1)}{\text{Trace}(W_Z N)} \right] \quad (50)$$

où

$$N_2 = {}^t P_2 N P_2$$

$$M_1 = {}^t P_1 M P_1$$

**Preuve :** l'équation (50) traduit dans le cadre présent la relation usuellement utilisée en régression multiple

$$\|z - \hat{z}\|_{N \otimes M}^2 = \|z\|_{N \otimes M}^2 (1 - r^2(z, \hat{z}))$$

puisque

$$\|\hat{z}\|_{N \otimes M}^2 = \text{Trace}(\hat{Z}N {}^t \hat{Z}M)$$

$$\|z\|_{N \otimes M}^2 = \text{Trace}(ZN {}^t ZM) = \text{Trace}({}^t ZMZN) = \text{Trace}(W_Z N)$$

et

$$r(z, \hat{z}) = \frac{\langle z, \hat{z} \rangle_{N \otimes M}}{\|z\|_{N \otimes M} \|\hat{z}\|_{N \otimes M}} = \text{COR}(Z, \hat{Z})$$

qui se simplifie, puisque  $P_2 \otimes P_1$  est un projecteur, en

$$r^2(z, \hat{z}) = \frac{\|\hat{z}\|_{N \otimes M}^2}{\|z\|_{N \otimes M}^2} \quad (51)$$

$$\|\hat{z}\|_{N \otimes M}^2 = \|P_2 \otimes P_1 z\|_{N \otimes M}^2 = \|z\|_A^2$$

où

$$\begin{aligned} A &= (P_2 \otimes P_1) N \otimes M (P_2 \otimes P_1) = {}^t P_2 N P_2 \otimes {}^t P_1 M P_1 \\ &= N_2 \otimes M_1 \quad \parallel \end{aligned}$$

Le carré du coefficient de corrélation  $r^2$  s'interprète alors sous la forme (51) comme le rapport de la "dispersion locale" ou dispersion expliquée par les contraintes, à la "dispersion totale" décrite par une analyse sans contraintes.

Dans le cas particulier où l'on impose pour seule contrainte  $Z = B^t Y_2$ , associée aux mesures  $Y_2$  réalisées sur les unités statistiques, on obtient

$$\hat{\underline{z}} = P_2 \otimes I \underline{z}$$

$$\hat{Z} = Z {}^t P_2$$

que l'on analyse dans les métriques M et N. On résoud donc les équations de l'Analyse en Composantes Principales sur Variables Instrumentales par application de l'Analyse en Composantes Principales à l'étude  $(\hat{Z}, N, M)$  ou de façon équivalente  $(\hat{Z}, N_2, M_1)$ . La dispersion résiduelle est déduite de la propriété précédente. Par référence à une structure gaussienne, l'application des résultats de RAO (1965, 1966, 1967) conduisent à choisir une métrique M particulière (cf. Remarque 1) de façon à construire des estimateurs aux moindres carrés satisfaisants pour  $\hat{B}$  et  $\hat{Z}$  et à disposer ainsi d'outils d'aide à l'interprétation des résultats. Les résultats de KHATRI (1966) assurent alors la consistance des estimateurs ainsi construits.

**Corollaire 1:**

$$\text{Trace}((Z-\hat{Z})N {}^t (Z-\hat{Z})M) = \text{Trace}(W_Z N) \left[ 1 - \frac{\text{Trace}(W_{\hat{Z}} N)}{\text{Trace}(W_Z N)} \right] \quad (50\text{-bis})$$

**Preuve:**

$$\hat{Z} = Z {}^t P_2$$

par suite, puisqu'ici  $M_1=M$  et que  $P_2$  est un projecteur N-symétrique, on obtient

$$\text{Tr}(Z N_2 {}^t Z M) = \text{Tr}(Z {}^t P_2 N P_2 {}^t Z M) = \text{Tr}(\hat{Z} N {}^t \hat{Z} M) = \text{Tr}(\hat{Z} M \hat{Z} N) \quad \parallel$$

Une deuxième présentation de l'Analyse en Composantes Principales sur Variables Instrumentales, utile pour l'analyse des tableaux de dissimilarité résulte d'un traitement distinct des deux types de contraintes. Posons

$$W_Z^1 = {}^t Z M_1 Z = {}^t Z P_1 M P_1 Z = \hat{Z}_1 M \hat{Z}_1$$

où la matrice  $\hat{Z}_1 = P_1 Z$  résulte de la projection des unités statistiques sur le sous-espace vectoriel  $E_1$  de E engendré par les contraintes  $Y_1$ . On vérifie alors immédiatement l'égalité

$$\text{Trace}(Z N_2 {}^t Z M_1) = \text{Trace}(W_Z^1 N_2)$$

dont on déduit le corollaire

**Corollaire 2 :** (1) Sous la contrainte  $Z = B {}^t Y_2$ , l'approximation  $\hat{Z}$  qui réalise le minimum de  $G(Z, Y)$  vérifie l'identité

$$\text{Trace}((Z-\hat{Z})N {}^t (Z-\hat{Z})M) = \text{Trace}(W_Z N) \left[ 1 - \frac{\text{Trace}(W_Z N_2)}{\text{Trace}(W_Z N)} \right] \quad (50\text{-ter})$$

(2) les vecteurs propres  $N$ -orthonormés du problème d'ACPVI ( $\hat{Z}$ ,  $M, N$ ) sont les projections sur le sous-espace  $F_2$  de  $F$  engendré par les contraintes  $Y_2$ , des vecteurs propres  $N_2$ -orthonormés de l'opérateur  $W_Z N_2$ .

**Preuve :** Le premier point découle de ce qui précède. Montrons donc le point (2).  $N_2$  est de rang au plus égal à  $k_2$  et donc  $W_Z N_2$  admet au plus  $k_2$  vecteurs propres non nuls vérifiant

$$W_Z N_2 \underline{u} = {}^t Z M Z {}^t P_2 N P_2 \underline{u} = \lambda \underline{u}$$

Par suite, le vecteur  $\underline{v} = P_2 \underline{u}$ , projection du vecteur propre  $\underline{u}$  sur  $F_2$  vérifie l'équation

$$P_2 {}^t Z M Z {}^t P_2 N \underline{v} = \lambda \underline{v}$$

c'est-à-dire

$${}^t \hat{Z} M \hat{Z} N \underline{v} = \lambda \underline{v}$$

De plus

$$1 = {}^t \underline{u} N_2 \underline{u} = {}^t \underline{u} {}^t P_2 N P_2 \underline{u} = {}^t \underline{v} N \underline{v} \quad \parallel$$

**Remarque :** La configuration  $X$  de rang  $r$  résumant les données sous les contraintes  $Y_2$  est déterminée à partir des  $r$  vecteurs propres  $\underline{v}_s$  associés au  $r$  plus grandes valeurs propres  $\lambda_s$ . Elle est donc construite à partir des  $r$  vecteurs propres  $\underline{u}_s$  correspondants. Seule la connaissance de la métrique  $N_2$  est donc nécessaire à la résolution du problème.

La décomposition spectrale de l'opérateur  $\hat{Z}$  s'écrit :

$$\hat{Z} = U D_\lambda {}^t V$$

et fournit les vecteurs propres  $\underline{u}^s$   $M$ -normés à 1 et la matrice diagonale  $D_\lambda$  regroupant les valeurs singulières  $\lambda_s$  de  $Z$  rangées par ordre décroissant. Le résumé  $X$  de dimension  $r$  est défini par les colonnes  $\underline{u}^s$  de la matrice  $U$  associées aux  $r$  plus grandes valeurs singulières de  $\hat{Z}$  (supposées distinctes ici), en posant :

$$X = [U D_\lambda]_r \quad (51)$$

Il est constitué des  $r$  premiers vecteurs propres de  $Z$  normés à la valeur singulière correspondante. On sait depuis les travaux d'Eckart et Young que parmi toutes les configurations résumées  $Y$  d'ordre  $r$  de  $Z$ , la matrice  $X$  ainsi construite minimise la quantité

$$G(Z, Y) = \text{Trace}((\hat{Z} - Y) N {}^t (\hat{Z} - Y) M) \quad (52)$$

$$= \|\hat{z} - y\|_{N \otimes M}$$

dont la valeur est à l'optimum

$$G(Z, X) = \sum_{s=r+1}^n (\lambda_s)^2$$

**Proposition 2:** La dispersion résiduelle mesurant la qualité du résumé X est donnée par:

$$G(Z, X) = \text{Trace}((Z-X)N^t(Z-X)M) = \text{Tr}(W_Z N) \left[ 1 - \frac{\sum_{s=1}^n (\lambda_s)^2}{\text{Tr}(W_Z N)} \right] + \sum_{s=r+1}^n (\lambda_s)^2 \quad (53)$$

**Preuve :** On a en effet

$$\text{Trace}((Z-X)N^t(Z-X)M) = \|\underline{z} - \underline{x}\|_{N \otimes M}^2 = \|\underline{z} - \hat{z}\|_{N \otimes M}^2 + \|\hat{z} - \underline{x}\|_{N \otimes M}^2 + 2\langle \underline{z} - \hat{z}, \hat{z} - \underline{x} \rangle_{N \otimes M}$$

par suite l'identité (53) résulte de la nullité du terme  $\langle \underline{z} - \hat{z}, \hat{z} - \underline{x} \rangle_{N \otimes M}$  qui est une conséquence de la décomposition N-orthogonale de F en

$$F = F_Z^\perp \oplus F_Z = F_Z^\perp \oplus F_{\hat{z}} \oplus F_{Z-\hat{z}}$$

puisque  $\hat{z} - X$  est à valeurs dans  $F_{\hat{z}}$  ¶

Compte tenu de la remarque 2, le modèle de CARROLL et al. (1976) consiste à définir un ensemble de contraintes décrivant le sous-espace vectoriel  $\langle E_1, F_2 \rangle$  de  $F \otimes E$ . Son application au cadre de l'analyse des tableaux de dissimilarité Euclidienne nécessite cependant de se restreindre aux cas de contraintes portant sur les seules unités statistiques puisque les variables latentes sont alors des inconnues du problème.

Dans le cas d'une analyse du triple  $(I, \delta_I(Z), Q_{|X|})$ , le coefficient  $G(Z, X)$  s'exprime en fonction de  $\Delta(Z)$ , X et Q seulement à l'aide de l'équation (27). En effet, puisque  $F_X$  est un sous-espace vectoriel de  $F_Z$  associé à la décomposition N-orthogonale

$$F_Z = F_X \oplus F_{Z-X}$$

on a, par application du théorème de Pythagore,

$$\begin{aligned} G(Z, X) &= \text{Trace}((Z-X)N^t(Z-X)M) = \text{Trace}(W_{Z-X} N) \\ &= \sum_{i < j} q_{ij} d_{ij}^2(Z-X) = \sum_{i < j} q_{ij} \delta_{ij}^2(Z) - \sum_{i < j} q_{ij} d_{ij}^2(X) \end{aligned}$$

L'indice G ne fournit donc pas un critère d'ajustement aux moindres carrés des dissimilarités Euclidiennes  $\delta_{ij}(Z)$ . On peut cependant utiliser les traductions de la proposition 2 et de son corollaire pour définir en termes de dissimilarité la perte d'information résultant respectivement de la construction de  $\hat{Z}$  et celle de X, puisque

$$\text{Trace}(W_Z N) = \sum_{i < j} \sum q_{ij} \delta_{ij}^2(Z)$$

$$\text{Trace}(W_{\hat{Z}} N) = \sum_{i < j} \sum q_{ij} d_{ij}^2(\hat{Z})$$

$$\text{Trace}(W_X N) = \sum_{i < j} \sum q_{ij} d_{ij}^2(X)$$

### III.1.2 RÉOLUTION DU PROBLEME EN TERMES D'OPÉRATEURS

Une deuxième classe de méthodes permettant de prendre en compte des contraintes résulte du choix d'un critère d'optimisation distinct. De tradition française, elle s'inspire des travaux d'ESCOUFIER (1973) présentant l'Analyse en Composantes Principales et plus généralement l'ensemble des méthodes d'analyse multidimensionnelle linéaire en termes d'opérateurs. On caractérise alors l'Analyse en Coordonnées Principales comme la recherche de la configuration X de rang donné r réalisant le minimum de la fonction critère

$$E(Z, Y) = \text{trace}((W_Z - W_Y)N (W_Z - W_Y)N) \quad (54)$$

dans laquelle la configuration Z est caractérisée par l'opérateur associé  $U_Z = W_Z N$  dont l'analyse spectrale fournit une solution de la dimension souhaitée. Cependant  $U_Z$  opère sur F et permet donc de prendre en compte des contraintes moins générales que l'approche de CARROLL et al. (1976) fondée sur l'indice  $G(Z, Y)$ . On recherche en effet le minimum de  $E(Z, Y)$  dans la classe des configurations Y vérifiant la contrainte  $Y = B_2 Y_2$ , où  $Y_2$  reçoit la même interprétation que précédemment. Par suite

$$W_Y = {}^1 Y M Y = {}^1 Y_2 {}^1 B_2 M B_2 Y_2 = {}^1 Y_2 M_2 Y_2$$

et la connaissance de Y est déduite de celle de  $B_2$  et donc de celle de la (semi-)métrique  $M_2 = {}^1 B_2 M B_2$ . La connaissance de la (semi-)métrique  $M_2$  sur E permet de construire la configuration  $\hat{Z}$  la plus proche de Z au sens où

$$E(Z, \hat{Z}) = \text{Min} \{ E(Z, Y) ; Y = B_2 Y_2 \}$$

Dans une phase de réduction des données, l'analyse classique de  $\hat{Z}$  fournit alors la configuration X de dimension r souhaitée. L'homogénéité des critères optimisés dans ces

deux phases de l'analyse justifie alors la décomposition additive du critère  $E(Z,X)$  donnée par la proposition suivante.

**Proposition 3:**

$$E(Z,X) = E(Z,\hat{Z}) + \sum_{s=r+1}^n \lambda_s^2(\hat{Z}) \quad (55)$$

où  $\lambda_s(\hat{Z})$  est la  $s$ -ième valeur propre de l'opérateur  $U_Z$ .

**Preuve:** Tensorialisons la formulation du problème. Il vient alors

$$\begin{aligned} E(Z,X) &= \text{Trace}((W_Z - W_X)N (W_Z - W_X)N) = \| \underline{w}_Z - \underline{w}_X \|_{N \otimes N}^2 \\ &= \| \underline{w}_Z - \underline{w}_{\hat{Z}} \|_{N \otimes N}^2 + \| \underline{w}_{\hat{Z}} - \underline{w}_X \|_{N \otimes N}^2 \end{aligned}$$

par simple application du théorème de Pythagore.

Réécrivons les résultats d'ESCOUFIER (1973) sous forme tensorielle : l'Analyse en Composantes Principales du tableau  $\hat{Z}$  construit la représentation résumée  $X$  de rang  $r$  réalisant le minimum du critère

$$E(\hat{Z},X) = \| \underline{w}_{\hat{Z}} - \underline{w}_X \|_{N \otimes N}^2$$

et la solution  $X$  retient les vecteurs propres associés aux  $r$  plus grandes valeurs propres de  $U_{\hat{Z}} = W_{\hat{Z}}N$ . La solution vérifie alors (cf. ESCOUFIER (1973)) :

$$E(\hat{Z},X) = \sum_{s=r+1}^n \lambda_s^2(W_{\hat{Z}}N)$$

où l'on peut noter plus simplement

$$\lambda_s^2(W_{\hat{Z}}N) = \lambda_s^2(\hat{Z}) \quad \parallel$$

Nous démontrons alors un théorème résolutif, dû initialement à MAILLES (1979) et repris par BONIFAS et al. (1984). De plus le calcul de  $\hat{M}_2$  intervient comme élément de démonstration (cf. formule (3.5), p. 444) dans SCHONEMANN (1972). Notre démonstration suit des voies distinctes, fondées sur l'utilisation du formalisme tensoriel qui nous semble donner un éclairage intéressant sur la solution et sur les aides à l'interprétation qui découlent alors directement de l'interprétation du problème en termes de régression linéaire (analyse des résidus, diagnostiques, recherche de données extrêmes, recherche de cases influentes, mesure de bras de levier,...).

**Théorème 1:** La métrique  $M$  réalisant le minimum de  $E(Z, Y)$  sous la contrainte (61)  $Z = B^t Y_2$ , est

$$\widehat{M}_2 = V_2^- Y_2 N W_Z N^t Y_2 V_2^- \quad (56)$$

où

$$V_2 = Y_2 N^t Y_2$$

**Preuve:**

$$\begin{aligned} E(Z, Y) &= \text{Trace} (N(W_Z - W_Y)N(W_Z - W_Y)) \\ &= \text{Vec}(W_Z - W_Y) N \otimes N \text{Vec}(W_Z - W_Y) \\ &= \| \underline{w}_Z - \underline{w}_Y \|_{N \otimes N}^2 \end{aligned}$$

l'inconnue du problème est  $\underline{m}_2$  et, sous forme tensorielle, on écrit

$$\begin{aligned} \underline{w}_Z &= \text{Vec}(W_Z) = {}^t Z \otimes Z \text{Vec}(M) = {}^t (Z \otimes Z) \underline{m} \\ \underline{w}_Y &= \text{Vec}(W_Y) = {}^t (Y_2 \otimes Y_2) \underline{m}_2 \end{aligned}$$

Le problème s'écrit alors en termes de régression linéaire dans  $F \otimes F$

$$\underline{w}_Z = {}^t (Y_2 \otimes Y_2) \underline{m}_2 + \underline{\epsilon}$$

La formulation du problème en termes de moindres carrés conduit à minimiser la quantité

$$\| \underline{w}_Z - {}^t (Y_2 \otimes Y_2) \underline{m}_2 \|_{N \otimes N}^2$$

dont la solution classique est

$$\begin{aligned} \widehat{\underline{m}} &= [(Y_2 \otimes Y_2)(N \otimes N) {}^t (Y_2 \otimes Y_2)]^- (Y_2 \otimes Y_2)(N \otimes N) \underline{w}_Z \\ &= (V_2 \otimes V_2)^- (Y_2 N \otimes Y_2 N) \underline{w}_Z \\ &= V_2^- Y_2 N \otimes V_2^- Y_2 N \underline{w}_Z \end{aligned}$$

soit encore, sous forme matricielle

$$\widehat{M}_2 = V_2^- Y_2 N W_Z N^t Y_2 V_2^- \quad \parallel$$

**Remarque :** Une écriture non redondante du problème devrait utiliser l'opérateur vecs du fait de la symétrie des métriques  $M$  et  $M_2$ . L'image de  $Z \otimes Z$  dans le sous-espace vectoriel de  $F \otimes F$  décrit par les opérateurs symétriques définis sur  $F$  est alors utilisée en lieu et place de  $Z \otimes Z$ . Cependant nous travaillons ici dans  $F \otimes F$  de façon à assurer une comparaison plus aisée avec les démonstrations données précédemment.

L'interprétation de la transformation de la métrique induite par la prise en compte des contraintes se déduit de l'écriture

$$\widehat{M}_2 = B_{2Z} M' B_{2Z}$$

où

$$B_{2Z} = \widehat{V}_2^{-1} V_{2Z} = V_2^{-1} Y_2 N' Z$$

est la matrice des coefficients de régression estimés dans la régression aux moindres carrés:

$$Z = B_2 Y_2 + U \quad (57)$$

Notons  $\| \| A \| \|$  la norme de Hilbert-Schmidt de l'opérateur symétrique  $A$  défini sur  $F$ . Alors :

$$\| \| A \| \|^2 = \| \mathbf{a} \|_{\text{IØI}}^2$$

où  $\mathbf{a} = \text{Vec}(A)$ .

Par suite le critère  $E(Z, Y)$  s'écrit

$$E(Z, Y) = \| \| U_Z - U_Y \| \|^2$$

La valeur de  $E(Z, Y)$  s'écrit donc en fonction du coefficient RV de corrélation entre opérateurs (cf. BONIFAS et al. (1984)) :

**Corollaire:**

$$E(Z, Y) = \| \| U_Z \| \|^2 (1 - RV^2(U_Z, U_Y)) \quad (58)$$

**Preuve:** Les équations classiques de la régression donnent

$$E(Z, Y) = \| \underline{w}_Z - \underline{w}_Y \|_{NØN}^2 = \| \underline{w}_Z \|_{NØN}^2 [1 - r^2(\underline{w}_Z, \underline{w}_Y)]$$

où, puisque les tenseurs  $\underline{w}_Z$  et  $\underline{w}_Y$  sont centrés par construction lorsque les configurations  $Z$  et  $Y_2$  sont supposées centrées

$$\begin{aligned} r(\underline{w}_Z, \underline{w}_Y) &= \frac{\langle \underline{w}_Z, \underline{w}_Y \rangle_{NØN}}{\left[ \| \underline{w}_Z \|_{NØN}^2 \| \underline{w}_Y \|_{NØN}^2 \right]^{\frac{1}{2}}} \\ &= \frac{\text{Trace}(W_Z N W_Y N)}{\left[ \text{Trace}(W_Z N) \text{Trace}(W_Y N) \right]^{\frac{1}{2}}} \end{aligned}$$

qui est le coefficient  $RV(U_Z, U_Y)$  étudié par ESCOUFIER (1973).

$$RV(U_Z, U_Y) = \frac{\langle\langle U_Z, U_Y \rangle\rangle}{\|U_Z\| \|U_Y\|} \quad \parallel$$

On a de plus

$$\widehat{w}_Z = {}^t(Y_2 \otimes Y_2) \widehat{m}_2 \Leftrightarrow \widehat{W}_Z = {}^t Y_2 \widehat{M}_2 Y_2 \quad (59)$$

L'équation (56) et la définition du projecteur  $P_2$  permettent donc de vérifier l'identité :

$$\begin{aligned} \widehat{W}_Z &= P_2 {}^t Z M Z {}^t P_2 \\ &= {}^t \widehat{Z} \widehat{M} \widehat{Z} = W_{\widehat{Z}} \end{aligned} \quad (59\text{-bis})$$

Par suite la configuration  $X$  est obtenue par l'analyse spectrale de l'opérateur  $\widehat{U}_Z = \widehat{W}_Z N$  ou de façon équivalente par Analyse en Composantes Principales du triplet  $(Z, \widehat{M}_2, N)$ . De plus, on déduit de la décomposition orthogonale des résidus

$$\begin{aligned} E(Z, Y) &= \|w_Z - w_Y\|_{N \otimes N}^2 = \|w_Z - \widehat{w}_Z\|_{N \otimes N}^2 + \|\widehat{w}_Z - w_Y\|_{N \otimes N}^2 \\ &= E(Z, \widehat{Z}) + E(\widehat{Z}, Y) \end{aligned} \quad (60)$$

où le terme d'intérêt, mesurant la dispersion résiduelle s'écrit :

**Corollaire:**

$$E(Z, \widehat{Z}) = \| (I - P_2 \otimes P_2) w_Z \|_{N \otimes N}^2 = \| (W_Z - P_2 W_Z {}^t P_2) N \|_{\parallel}^2$$

**Preuve :** On a en effet :

$$\begin{aligned} \widehat{w}_Z &= {}^t(Y_2 \otimes Y_2) \widehat{m}_2 = {}^t(Y_2 \otimes Y_2) V_2^{-1} Y_2 N \otimes V_2^{-1} Y_2 N w_Z \\ &= P_2 \otimes P_2 w_Z \end{aligned} \quad (61)$$

où  $P_2$  est le projecteur  $N$ -orthogonal sur le sous-espace  $F_2$  de  $F$  engendré par le tableau des contraintes  $Y_2$ :

$$P_2 = {}^t Y_2 V_2^{-1} Y_2 N = W_2 N \quad \parallel \quad (62)$$

**Remarque :** L'opérateur  $W_2$  déduit du tableau centré  $Y_2$  mesure les distances de Mahalanobis entre les unités statistiques évaluées à l'aide de  $Y_2$ .

De plus,  $(U_Z - \widehat{U}_Z)$  est la matrice du projecteur  $N$ -orthogonal sur le sous-espace complémentaire  $N$ -orthogonal à  $F_2$  dans  $F$  :

$$\begin{aligned}
 U_Z - \hat{U}_Z &= (W_Z - \hat{W}_Z)N \\
 &= (W_Z - W_2 N W_Z N W_2)N \\
 &= W_{Z.Y_2} N
 \end{aligned} \tag{63}$$

où  $W_{Z.Y_2}$  s'interprète comme la matrice des produits scalaires résiduels de  $Z$  lorsque  $Y_2$  est contrôlée, c'est-à-dire la matrice des produits scalaires des  $z_i$  hors variables instrumentales  $Y_2$ . De façon duale nous vérifions en utilisant l'équation (59-bis):

$$\hat{U}_Z = \hat{W}_Z N = {}^t \hat{Z} M \hat{Z} N \tag{64}$$

C'est dire que l'Analyse en Composantes Principales de  $(Z, \hat{M}_2, N)$  est une Analyse en Composantes Principales de  $(\hat{Z}, M, N)$  où  $\hat{Z}$  est déduit de  $Z$  par projection des variables constituant le tableau  $Z$  sur  $F_2$ . C'est donc une Analyse en Composantes Principales sur Variables Instrumentales. La solution construite en termes d'opérateurs fournit donc une solution identique au cas particulier de la méthode de CARROLL et al. (1976), caractérisant le cas d'une contrainte linéaire portant sur les seules unités statistiques.

**Remarque :**

$$\| \underline{W}_Z - \hat{\underline{W}}_Z \|_{N \otimes N}^2 = \| {}^t (Z \otimes Z) (\underline{m} - \hat{\underline{m}}_2) \|_{N \otimes N}^2 = \| {}^t Z (M - \hat{M}_2) Z N \| ^2$$

Aussi les équations (55) et (60) traduisent la propriété géométrique selon laquelle la décomposition  $M$ -orthogonale de l'espace de représentation

$$E = E_Y \oplus E_Y^\perp$$

correspond à une décomposition additive de la métrique ambiante:

$$M = \hat{M}_2 + (M - \hat{M}_2).$$

Cette correspondance est la base des travaux de TEBBOUNE (1981) sur l'Analyse en Composantes Principales sous contraintes linéaires.

L'application des résultats précédents à l'étude de données de dissimilarité permet de donner une caractérisation de la solution de l'Analyse en Coordonnées Principales sur Variables Instrumentales complémentaire des résultats fournis par CARROLL et al..

En effet considérons le triple  $(I, \delta_1(Z), P_I)$ , où  $\delta_1(Z)$  est une dissimilarité Euclidienne sur  $I$ . Considérons de plus l'élément  $Y_2 \in \mathcal{M}_{p_2, n}(\mathbb{R})$  définissant les contraintes linéaires. On suppose que  $Y_2$  engendre un sous-espace vectoriel de  $F_0$  de dimension  $k_2 \leq p_2$ . Alors on vérifie immédiatement

$$N = D_p$$

et en utilisant les équations (18) et (18-bis)

$$\begin{aligned} E(Z, X) &= \| \underline{w}_Z - \underline{w}_X \|_{D_P \otimes D_P}^2 = \text{STRAIN} \\ &= \frac{1}{4} \| \hat{\underline{d}}^2(Z) - \underline{d}^2(X) \|_{D_P Q_1 \otimes D_P Q_1}^2 \end{aligned}$$

Par suite on démontre :

**Proposition 4:** (a) la solution  $X$  du problème d'Analyse en Coordonnées Principales appliquée au triple  $(I, \delta_1(Z), P_1)$  sous les contraintes linéaires déterminées par la matrice  $Y_2$  ne dépend que de la dissimilarité Euclidienne  $\delta_1(Z)$ , de la pondération  $P_1$  et du projecteur  $D_P$ -orthogonal  $P_2$  sur le sous-espace  $F_2$  de  $F$  engendré par  $Y_2$ .

(b) On a de plus

$$E(Z, X) = \frac{1}{4} \| \hat{\underline{d}}^2(Z) - \underline{d}^2(\hat{Z}) \|_{D_P Q_1 \otimes D_P Q_1}^2 + \frac{1}{4} \| \underline{d}^2(\hat{Z}) - \underline{d}^2(X) \|_{D_P Q_1 \otimes D_P Q_1}^2$$

où  $X$  est la solution du problème d'Analyse en Coordonnées Principales appliqué au triple  $(I, \delta_1(\hat{Z}), P_1)$ .

**Preuve :** Le point (a) se déduit des résultats précédents. En effet l'expression de  $W_Z$  en fonction du tableau de dissimilarité Euclidienne  $\Delta(Z)$  et donc de la matrice  $H_Z$  associée s'écrit sous forme tensorielle

$$\underline{w}_Z = Q_1 \otimes Q_1 \underline{h}_Z$$

Par suite

$$\begin{aligned} \hat{\underline{w}}_Z &= P_2 \otimes P_2 (Q_1 \otimes Q_1 \underline{h}_Z) \\ &= P_2 Q_1 \otimes P_2 Q_1 \underline{h}_Z \end{aligned}$$

Par suite, si les variables définissant les contraintes sont centrées, on a  $P_2 Q_1 = P_2$  et donc

$$\hat{\underline{w}}_Z = P_2 \otimes P_2 \underline{h}_Z \quad \Leftrightarrow \quad \hat{W}_Z = P_2 \left( -\frac{1}{2} \Delta(Z) * \Delta(Z) \right) P_2$$

L'expression de  $E(Z, Y)$  se déduit alors de l'équation (60). ¶

$\hat{W}_Z$  s'exprime en fonction des dissimilarités  $\Delta(Z)$  observées et du projecteur  $P_2$ . On peut donc être tenté de traiter le cas général d'une dissimilarité, a priori non Euclidienne en réalisant l'Analyse en Coordonnées Principales à partir de l'analyse spectrale de l'opérateur  $\hat{W}_Z N$ . Le statut mathématique d'une telle analyse est cependant difficile à justifier autrement que par un argument d'analogie.

**Remarque :** L'étude du triple  $(I, \delta_1(Z), Q_{|x|})$ , sous contraintes linéaires déterminées par la matrice  $Y_2$  suit les mêmes voies pour une pondération  $Q_{|x|}$  de forme générale. Le lemme de Coxeter permet alors de munir  $F_0$  d'une métrique  $N$  non diagonale lorsque  $Q_{|x|} \neq P_1 \otimes P_1$ . Puisque le vecteur  $\underline{1}$  appartient au noyau de  $N$ , on a alors

$$W_Z = \frac{1}{2} \Delta(Z) * \Delta(Z)$$

et la configuration est centrée à l'origine.  $E(Z, X)$  généralise alors la forme (18) donnée par l'indice STRAIN et la solution  $X$  est construite à partir de l'analyse spectrale de l'opérateur  $\widehat{W}_Z N$ . Par suite la proposition 4 s'étend au cas d'une pondération générale

$$E(Z, X) = \frac{1}{4} \| \underline{d}^2(Z) - \underline{d}^2(X) \|_{N \otimes N}^2$$

puisque le projecteur  $N$ -orthogonal  $Q_1$  sur  $F_0$  est alors l'identité.

### III.1.3. UNE NOUVELLE APPROCHE EN TERMES DE ROTATIONS PROCRUSTES

Les contraintes traitées dans les deux paragraphes précédents imposent des contraintes d'appartenance à un sous-espace vectoriel de  $F \otimes E$ . Elles s'appliquent en particulier au cas de données collectées dans le cadre d'un plan d'expérience, telles que celles qu'envisagent NOMA et JOHNSON (1979) et CARROLL et al. (1976, 1980). Mais la référence à la méthodologie de l'analyse des courbes de croissance montre que leur domaine d'application est beaucoup plus vaste. Dans tous les cas, lorsque le modèle ajuste de façon satisfaisante les données, cette approche permet de retrouver des facteurs explicatifs suivant des axes en général distincts des axes principaux issus d'une analyse sans contraintes.

Nous étudions ici un type de contraintes un peu différent puisqu'il consiste à imposer des contraintes déterminées dans la tradition des rotations procrustes (cf. SIBSON (1978)). L'application d'une telle contrainte dans le cadre de l'analyse de tableaux de dissimilarité permet de traduire simplement des informations a priori portant sur la position relative des unités statistiques dans un espace affine de dimension connue. On recherche alors une transformation  $B_2$  appartenant au groupe des déplacements engendré par les translations, les rotations, les symétries et les homothéties globales. Les problèmes d'ajustement d'une translation optimale (c'est-à-dire de choix d'une origine) ne sont pas discutés ici. Usuellement, l'Analyse en Coordonnées Principales considère que la configuration reconstituée est naturellement centrée en retenant pour origine des axes le barycentre des points représentatifs des unités statistiques. Nous avons vu plus haut que le choix d'une

origine autre que le barycentre des points représentatifs, discuté dans GOWER (1982) et dans d'AUBIGNY (1984), conduit à une formule adaptée de  $W_Z$  à partir de  $\Delta(Z)$  ou de  $H(Z)$  et n'introduit aucune complexité nouvelle ici. On suppose donc dans la suite, sans perte de généralité, que les configurations  $Z$  et  $Y$  sont centrées en leur barycentre. Nous nous proposons ici de rechercher la matrice de rotation et/ou de symétrie optimale  $B$  parmi les matrices orthogonales, de sorte que la matrice  $\hat{Z} = BZ$  soit aussi proche que possible de la matrice des contraintes  $Y$ . Nous construisons une solution du problème qui étend les résultats de SIBSON (1978) au cadre Euclidien. La géométrie sur  $E$  et sur  $F$  est alors spécifiée par la donnée des métriques  $M$  et  $N$  fixées à priori par l'analyste. Le critère d'adéquation retenu s'écrit alors :

$$\text{Min}\{ G(BZ, Y) ; {}^tBMB = M \} \quad (65)$$

On retrouve d'ailleurs le critère utilisé par cet auteur en prenant pour  $M$  et  $N$  les métriques identité.

Nous montrons alors que la solution obtenue s'exprime du point de vue algébrique en fonction de  $\delta_I$ ,  $Q_{I, I}$  et  $Y$  uniquement et par suite s'applique dans le cas de données de dissimilarité.

Notons  $A^{1/2}$  l'unique matrice symétrique, semi-définie positive, racine carrée de la matrice symétrique semi-définie  $A$ . On vérifie simplement (SIBSON lemme 4.1) que toute matrice carrée  $C$  vérifie l'identité

$$\text{Trace}[({}^tCC)^{1/2}] = \text{Trace}[(C^tC)^{1/2}] \quad (66)$$

nous montrons de plus le lemme suivant:

**Lemme 2:** Soient  $A$  une matrice carrée et  $(A, M^{-1}, M)$  l'étude définie par l'opérateur de matrice  $A$ ,  $M$  une matrice symétrique définie positive et  $B$  une matrice carrée de même ordre que  $A$  et telle que  ${}^tBMB = M$ . Alors:

$$\text{Trace}({}^tBA) = \text{Trace}[(M^{1/2} {}^tA M^{-1} A M^{1/2})^{1/2}] \quad (67)$$

De plus l'égalité est réalisée lorsque

$${}^tBA = (M {}^tA M^{-1} A)^{1/2} \quad (68)$$

et cette équation admet toujours une solution  $M$ -orthogonale, donnée de façon unique lorsque  $A$  est régulière par

$$B = M^{-1} A M ({}^tA M^{-1} A M)^{-1/2} \quad (69)$$

**Preuve:** La décomposition en valeurs singulières de A s'écrit

$$A = UD\lambda{}^tV \quad {}^tUM^{-1}U=I \quad {}^tVMV=I$$

et l'on vérifie sans peine que:

$$\text{Trace}({}^tBA) = \text{Trace}({}^tV{}^tBUD\lambda)$$

B est un déplacement et donc  ${}^tBMB = M$ . Par suite la matrice BV est M-orthogonale, et les matrices  $K = M^{1/2}BV$  et  $L = M^{-1/2}UM ({}^tAM^{-1}AM)^{-1/2}$  sont orthogonales. Par suite:

$$\text{Trace}({}^tV{}^tBUD\lambda) = \text{Trace}({}^tKLD\lambda) = \sum_i^n \sum_j^n k_{ji}l_{ji}\lambda_i$$

et comme  $\lambda_i$  est positive ou nulle pour tout i, l'inégalité

$$\text{Trace}({}^tBA) \leq \text{Trace}(D\lambda)$$

résulte des deux propriétés

$$\begin{aligned} \sum_j^n k_{ji}l_{ji} &\leq \left[ \sum_j^n k_{ji}^2 \right]^{1/2} \cdot \left[ \sum_j^n l_{ji}^2 \right]^{1/2} \\ \sum_j^n k_{ji}^2 &= \sum_j^n l_{ji}^2 = 1 \end{aligned}$$

et l'égalité est réalisée lorsque  ${}^tV{}^tBUD\lambda = D\lambda$ . En effet

$$\left( \lambda_i \sum_j^n k_{ji}l_{ji} = \lambda_i \quad \forall i \right) \Leftrightarrow \left( \text{Trace}({}^tKLD\lambda) = \text{Trace}(D\lambda) \right)$$

mais si  $\delta_i^j$  est le symbol de Kronecker, on a

$$\text{si } \lambda_i = 0 \text{ alors } \lambda_i \sum_j^n k_{jm}l_{ji} = \lambda_i \delta_m^i \quad \forall m$$

$$\text{si } \lambda_i \neq 0 \text{ alors } \sum_j^n k_{ji}l_{ji} = 1 \Leftrightarrow \sum_j^n k_{ji} = \sum_j^n l_{ji}$$

c'est dire que

$$\lambda_i \sum_j^n k_{jm}l_{ji} = \lambda_i \delta_m^i \quad \forall m$$

et donc

$$({}^t\text{KLD}_\lambda = D_\lambda) \Leftrightarrow (\text{Trace}({}^t\text{KLD}_\lambda) = \text{Trace}(D_\lambda))$$

Maintenant

$$\begin{aligned} \text{Trace}(D_\lambda) &= \text{Trace}\left((D_\lambda^2)^{\frac{1}{2}}\right) = \text{Trace}\left({}^tD_\lambda {}^t\text{VMVD}_\lambda^{\frac{1}{2}}\right) \\ &= \text{Trace}\left({}^t(M^{1/2}\text{VD}_\lambda)(M^{1/2}\text{VD}_\lambda)^{\frac{1}{2}}\right) = \text{Trace}\left((M^{1/2}\text{VD}_\lambda^2 {}^t\text{VM}^{1/2})^{\frac{1}{2}}\right) \end{aligned}$$

On obtient donc l'inégalité (67) comme conséquence de l'équation

$$\text{VD}_\lambda^2 {}^t\text{V} = \text{VD}_\lambda {}^t\text{U M}^{-1} \text{U D}_\lambda {}^t\text{V} = {}^t\text{A M}^{-1} \text{A}$$

qui implique

$$\text{Trace}({}^t\text{BA}) = \text{Trace}(D_\lambda) = \text{Trace}\left[(M^{1/2} {}^t\text{A M}^{-1} \text{A M}^{1/2})^{1/2}\right]$$

D'autre part, l'égalité (68) est obtenue comme conséquence de l'égalité

$${}^t\text{V} {}^t\text{B U D}_\lambda = D_\lambda$$

celle-ci implique en effet

$${}^t\text{V} {}^t\text{B U D}_\lambda {}^t\text{V} = D_\lambda {}^t\text{V} \Leftrightarrow {}^t\text{V} {}^t\text{BA} = D_\lambda {}^t\text{V}$$

et par conséquent

$${}^t\text{V} {}^t\text{BA} = {}^t\text{VMVD}_\lambda {}^t\text{V} \Rightarrow {}^t\text{BA} = \text{MVD}_\lambda {}^t\text{V}$$

qui conduit à l'identité (68) cherchée

$${}^t\text{BA} = (\text{M} {}^t\text{A M}^{-1} \text{A})^{1/2}$$

puisque

$$(\text{MVD}_\lambda {}^t\text{V})^2 = \text{MVD}_\lambda^2 {}^t\text{V}$$

De plus

$${}^t\text{AB} = ({}^t\text{A M}^{-1} \text{A M})^{1/2} = {}^t\text{A M}^{-1} \text{A M} ({}^t\text{A M}^{-1} \text{A M})^{-1/2}$$

et donc

$$\text{B} = \text{M}^{-1} \text{A M} ({}^t\text{A M}^{-1} \text{A M})^{-1/2} \quad \parallel$$

La solution du problème posé en termes de rotations procrustes est fournie par la proposition suivante qui généralise le théorème 4.3 de SIBSON (1978):

**Proposition 5:** (a) Si Z et Y sont deux configurations centrées au barycentre, l'optimum de la fonction  $G(Z, BY)$  dans le groupe des déplacements sur l'espace Euclidien E est donné par toute transformation

$$B: Y \rightarrow BY$$

vérifiant l'équation

$$B({}^tV_{ZY} M V_{ZY} M)^{1/2} = V_{ZY} M \quad (70)$$

Cette équation admet toujours une solution M-orthogonale unique quand  $V_{ZY}$  est régulière, donnée par:

$$B = V_{ZY} M ({}^tV_{ZY} M V_{ZY} M)^{-1/2} \quad (71)$$

$$V_{ZY} = ZN {}^tY \quad (72)$$

(b) Le minimum de  $G(Z, BY)$  est alors

$$G^*(Z, BY) = \text{Trace}(V_Z M) + \text{Trace}(V_Y M) - 2 \text{Trace}[(M {}^tV_{ZY} M V_{ZY})^{1/2}] \quad (73)$$

$$\begin{aligned} &= \text{Trace}(W_Z N) + \text{Trace}(W_Y N) \\ &\quad - 2 \text{Trace}[\{(N^2 W_Z N^2)^{1/2} (N^2 W_Y N^2)^{1/2} (N^2 W_Z N^2)^{1/2}\}^{1/2}] \quad (73\text{-bis}) \end{aligned}$$

**Preuve:** Le point (a) est une conséquence directe du lemme 2 appliqué à la matrice

$$A = MZN {}^tY = M V_{ZY}$$

L'équation (73) s'obtient par substitution dans l'expression de  $G(Z, BY)$  du fait de l'inégalité (67). De plus (73-bis) résulte de l'application de l'identité (66) à la matrice

$$C = M^{\frac{1}{2}} Y N^{\frac{1}{2}} (N^{1/2} W_Z N^{1/2})^{\frac{1}{2}}$$

intervenant dans le troisième terme de droite de l'identité (73) symétrisée en

$$\text{Trace}\left((M^{1/2} {}^tV_{ZY} M V_{ZY} M^{1/2})^{\frac{1}{2}}\right)$$

On vérifie en effet que ces deux matrices dont on évalue la trace ont les mêmes valeurs propres non nulles avec le même ordre de multiplicité. ¶

La résolution du problème étendu au cadre Euclidien par l'introduction des métriques  $M$  et  $N$  qui réalise l'optimum  $G^*(Z, Y)$  traite le problème de rotation procruste lorsqu'on prend de plus en compte les homothéties globales. Ce problème n'offre cependant pas de complexité particulière puisque les résultats de SIBSON (1978) à ce sujet ne dépendent clairement pas d'un choix particulier de géométrie et s'appliquent donc directement au cadre général. On montre ainsi que le coefficient d'homothétie optimal réalisant le maximum de  $G^*(Z, \alpha B^* Y)$  où  $B^*$  est l'optimum construit précédemment est donné par:

$$\alpha = \frac{\text{Trace}(M {}^t V_{ZY} M V_{ZY})^{\frac{1}{2}}}{\text{Trace}(M V_Y)}$$

où  $V_Y = YN {}^t Y$ . On obtient donc une expression généralisant l'indice proposé par GOWER (1971) et repris par SIBSON (1978) dans un cadre intégrant des métriques non triviales, en appliquant la proposition 5:

$$G_s(Z, Y) = \text{Trace}(M V_Z) - \frac{\text{Trace}(M {}^t V_{ZY} M V_{ZY})}{\text{Trace}(M V_Y)} \quad (74)$$

$$= \text{Trace}(W_Z N) - \frac{\text{Trace}[(N^{1/2} W_Z N^{1/2})^{\frac{1}{2}} (N^{1/2} W_Y N^{1/2}) (N^{1/2} W_Z N^{1/2})^{\frac{1}{2}}]}{\text{Trace}(W_Y N)} \quad (75)$$

Les formules (74) et (75) expriment l'indice optimisé en fonction des seuls opérateurs  $W_Z$ ,  $W_Y$  et  $N$ . Par suite, ils sont directement utilisables sur des données de dissimilarité lorsque sont connues les matrices  $\Delta(Z)$  et  $D(Y)$ . On note cependant que leur application suppose que  $\Delta(Z)$  est une dissimilarité Euclidienne. La connaissance de la matrice  $Y$  même n'est pas nécessaire ici. Le problème des rotations procrustes peut donc s'exprimer aussi bien en termes de contraintes portant sur la configuration que de contraintes portant sur les distances étudiées dans le prochain chapitre.

Dans le cas d'une dissimilarité observée quelconque, la matrice  $W$  associée à  $\Delta$  n'est pas forcément gramienne. On rencontre alors les problèmes évoqués plus haut dans le cadre de l'Analyse en Coordonnées Principales. De façon complémentaire, bien que  $B$  s'exprime en fonction de  $V_{ZY}$ , l'opérateur  $\hat{Z} = {}^t B A$  qui approche au mieux  $Z$  au sens du critère des rotations procrustes s'écrit:

$$\hat{Z} = {}^t B A = ({}^t A M^{-1} A M)^{1/2} = (Y N W_Z N {}^t Y M)^{1/2}$$

Il est donc calculable en fonction de  $Y$ ,  $N$ ,  $M$  et  $W_Z$  et donc en fonction de  $Y$ ,  $N$ ,  $M$  et  $\Delta(Z)$ . On en déduit donc la proposition suivante.

**Proposition 6:** Soit  $(I, \delta_I, Q_{I \times I})$  un triplet formé d'un ensemble fini  $I$  de cardinal  $n$ , d'une dissimilarité Euclidienne  $\delta_I$  définie sur  $I$  et d'une pondération  $Q_{I \times I}$  définie sur  $I \times I$ . Soit une configuration cible des éléments de  $I$ , caractérisée soit par la donnée d'une matrice réelle  $Y$  d'ordre  $n \times r$  et de rang  $r$ , soit par la donnée d'une matrice de distances Euclidiennes  $D(Y)$  d'ordre  $n$  et de rang  $r$ , alors la solution du problème des rotations procrustes défini par la donnée de la configuration cible  $Y$  ou de  $D(Y)$  est totalement déterminée par la donnée de  $\delta_I$  lorsque les géométries retenues sur  $E$  et  $F$  sont décrites par des matrices  $M$  et  $N$  ne dépendant que de  $Q_{I \times I}$ .

Par suite le cadre hilbertien de l'Analyse en Coordonnées Principales permet de prendre en compte une grande variété de contraintes linéaires lorsque les données sont fournies sous forme d'un tableau de dissimilarité Euclidienne. L'analyse du tableau  $\hat{Z} = \alpha^* B^* Y$  permet alors de construire la solution  $X$  de dimension  $r$ . Les limites d'une telle méthodologie reposent cependant sur le rôle central joué par la matrice de produits scalaires  $W_Z$ , déduite des données  $\Delta(Z)$  par une transformation algébrique peu adaptée au traitement de données entachées d'erreur.

#### III.1.4. EXTENSION DE CES MODELES AU CAS DE CONTRAINTES STRUCTURELLES

La famille de contraintes traitées par CARROLL et al. s'écrit sous la forme la plus générale:

$$Z = A_1 {}^t A_2$$

$$A_1 = Y_1 B_1$$

$$A_2 = Y_2 B_2$$

où  $Y_1$  et  $Y_2$  sont des données du problème. Ce ne sont certes pas les contraintes les plus générales envisageables dans un cadre linéaire. L'expression  $Z = Y_1 B {}^t Y_2$  fournit en effet une équation de structure plus aisément reconnaissable sous la forme tensorielle  $z = (Y_2 \otimes Y_1) b$ . On peut donc se proposer d'estimer les termes  $(Y_1, Y_2, B)$  en imposant des contraintes de structure à chaque composante. L'interprétation en termes d'analyse de courbes de croissance, donnée plus haut, fournit alors un catalogue de structures particulières étudiées dans le domaine de la statistique multivariée. Le lecteur intéressé peut se reporter à GEISSER (1980) pour un panorama des modèles étudiés. Dans le domaine d'application de la psychométrie, l'introduction de contraintes structurelles est actuellement limité à des structures simples. Ainsi BLOXOM (1978) étudie le cas particulier de contraintes définies par des équations structurelles linéaires :

$$Z = Y_1 B$$

où  $B$  et  $Y_1$  sont à estimer, sous des contraintes de structure, adaptées au traitement du problème de Codage Multidimensionnel Multiple connu dans la littérature de langue anglaise sous l'intitulé Individual Differences Analysis (BLOXOM (1968), HORAN (1969), CARROLL et CHANG (1970), HARSHMAN (1970), CARROLL et ARABIE (1981), de LEEUW et HEISER (1980), SCHIFFMAN et al. (1981), TAKANE et al. (1977)). Dans ce cas, on dispose alors d'un ensemble fini de tableaux de dissimilarité:

$$\{\Delta^k, k=1, 2, \dots, K\}$$

dont on se propose de réaliser un plongement Euclidien dans un espace de faible dimension. La spécificité du modèle structurel est caractérisée par la recherche des représentations  $Z^k$  décrites sous la forme

$$Z^k = Y_k X$$

où la matrice  $X$  d'ordre  $r \times n$  s'interprète comme la représentation commune des unités statistiques dans un espace de dimension  $r$  et  $Y_k$  fournit la transformation spécifique imposée à la configuration de référence  $X$  de sorte que  $Z^k$  permette de reconstruire au mieux (au sens d'un critère de moindres carrés) la matrice de pseudo-produits scalaires  $W_Z$  associée aux dissimilarités données par le tableau  $\Delta^k$ . De plus, comme dans le cas de l'Analyse en Coordonnées Principales que cette classe de modèles prétend généraliser, l'utilisation d'un formalisme algébrique n'est pleinement justifiée que lorsque les dissimilarités  $\Delta^k$  sont Euclidiennes. Les problèmes difficiles d'identifiabilité semblent ne pas avoir été abordés dans le cas général d'un tableau  $Y_k$  quelconque de taille  $r \times r_k$ . Néanmoins, deux modèles structurels ont reçu une attention particulière dans les deux dernières décades :

(a) lorsque  $Y_k$  a la structure d'une matrice diagonale,  $Y_k = D_k$ , on retrouve les modèles d'Analyse des Différences Individuelles proposés initialement par BLOXOM (1968), HARSHMAN (1972), HORAN (1969) et CARROLL et CHANG (1970). On impose en général  $(D_k)_{ii} \geq 0$  de sorte que  $D_k$  s'interprète comme une métrique sur  $E$ , propre à chaque élément  $k$  de l'échantillon.  $(D_k)_{ii}$  traduit alors la pondération différentielle de l'axe  $i$  propre à l'individu  $k$ .

Associons à la matrice diagonale  $D_k$  le vecteur  $\underline{d}_k$  de  $i$ -ième terme  $(D_k)_{ii}$ . BLOXOM (1978) vérifie alors que

$${}^t \underline{w}^k(X) = {}^t \underline{d}_k {}^t \Lambda(X)(X \otimes I_n) = {}^t \underline{d}_k {}^t \Lambda(X)B(X)$$

où  ${}^t \Lambda(X)$  est une matrice d'ordre  $n \times p$  de  $j$ -ième ligne

$${}^t \lambda^j(X) = (0, \dots, {}^t \underline{x}^j, \dots, 0)$$

De sorte que les matrices regroupant les paramètres associés au  $k$  tableaux  $\Delta^k$  vérifient

$$W(X) = D {}^t\Lambda(X) B(X)$$

où  $W(X)$  (resp.  $D$ ) est la matrice d'ordre  $K \times n^2$  (resp.  $K \times p$ ) de  $k$ -ième ligne  ${}^t w^k(X)$  (resp.  ${}^t d_k$ ). On obtient donc une équation structurelle linéaire permettant d'estimer  $X$  et  $D$  à partir des données  $W$ . Notons de plus  $V_{W(X)}$  et  $V_D$  les matrices de covariance empirique déduites respectivement des tableaux  $W(X)$  et  $D$ , on vérifie alors que

$$V_{W(X)} = W(X)Q_1W(X) = {}^tB(X)\Lambda(X)V_D\Lambda(X)B(X)$$

où  $Q_1$  est le projecteur associé à une pondération  $P_K$  (cf. équation (17)).

BLOXOM (1978) remarque qu'on obtient ainsi un modèle de Structure de Covariance particulier appartenant à la classe des modèles

$$\Sigma = B (\Lambda\Phi\Lambda + \Psi^2) {}^tB + \theta^2$$

ajustés aux moindres carrés par JORESKOG (1970), avec ici

$$\Sigma = V_{W(X)}, \quad \Phi = V_D, \quad \Psi^2 = 0, \quad \theta^2 = 0.$$

Cette formulation algébrique conduit cependant à des algorithmes de résolution exigeants en place mémoire. C'est pourquoi chaque auteur a développé un algorithme spécifique. Bien que les propriétés statistiques des solutions ainsi construites ne soient pas connues, ces algorithmes sont très utilisés en psychométrie. Ces modèles sont implémentés dans de nombreux programmes du marché tels par exemple que INDSCAL et PARAFAC. SCHONEMANN (1972) présente un théorème d'existence de la solution pour des dissimilarités Euclidiennes et des diagnostics de falsifiabilité du modèle, alors que les problèmes liés d'identifiabilité et de dimension de l'espace de représentation posés par ces modèles sont étudiés par KRUSKAL (1977).

Un deuxième exemple de contraintes structurelles est fourni par l'analyse en facteurs communs et spécifiques d'un tableau de dissimilarité (CARROLL (1983)). Ce modèle, présenté dans l'introduction (ch. I) s'exprime clairement à l'aide de l'équation (7) comme un cas particulier du modèle de Structure de Covariance de JORESKOG (1970), bien que WEINSBERG et CARROLL (1984) proposent un algorithme spécifique. Il en est de même du cas particulier fourni par la résolution de la constante additive associée aux carrés des dissimilarités.

Remarquons de plus que le modèle de CARROLL (1983) peut simplement être étendu au traitement de données de Différences Individuelles. Considérons en effet le modèle défini par

$$Y^k = \begin{bmatrix} Z^k \\ D_{\underline{u}^k} \end{bmatrix} \quad \text{où } Z^k = D_k X$$

où  $D_{\underline{u}^k}$  (resp.  $D_k$ ) est une matrice diagonale d'ordre  $n$  (resp.  $r$ ). On admet ainsi que la spécificité  $u_i^k$  de chaque individu  $i$  peut dépendre de l'occasion  $k$  où elle est estimée. On vérifie alors que

$$Y^k = D_k Y$$

où l'on a posé

$$Y = \begin{bmatrix} X \\ I_n \end{bmatrix} \quad \text{et } \Delta^k = \begin{bmatrix} D_k & 0 \\ 0 & D_{\underline{u}^k} \end{bmatrix}$$

On obtient ainsi un modèle de Structure de Covariance particulier qui admet donc une solution  $X$  optimale en termes de moindres carrés, à l'aide de la procédure de JORESKOG (1970). Le cas particulier  $D_{\underline{u}^k} = c^k I_n$  conduit de façon corollaire à estimer simultanément les  $K$  constantes additives associées aux carrés des distances:

$$d_{ij}^2(Y^k) = d_{ij}^2(Z^k) + c^k \quad i \neq j$$

dans le cadre de l'Analyse des Différences Individuelles. On obtient alors un modèle complémentaire de celui qui est traité par BLOXOM (1978) et KRANE (1978) :

$$d_{ij}(Y^k) = d_{ij}(Z^k) + c^k \quad i \neq j$$

BLOXOM construit un algorithme spécifique à l'aide d'une méthode de gradient, alors que KRANE associe à ce problème une Structure de Covariance et optimise le critère aux moindres carrés

$$L(X) = \sum_{k=1}^K \sum_{i < j} [e_{ij}^k]^2$$

$$e_{ij}^k = \delta_{ij}^k - d_{ij}^k(X) - c^k$$

mis à l'honneur par COOPER (1972) dans le cas  $K=1$  pour résoudre ce problème de constante additive.

(b) lorsque  $Y_k = M_k$  où  $M_k$  est une pseudo-métrique de forme générale conduisant au modèle général qui caractérise chaque élément  $k$  par une géométrie Euclidienne spécifique définie sur un même espace vectoriel  $E$ . La construction d'une solution est en général itérative. Une caractérisation algébrique de la solution est donnée par SCHONEMANN

(1972) qui traite le problème comme un cas particulier du modèle de CARROLL et al. (1976). Cette solution adaptée au cas de données de dissimilarité Euclidienne non entachées d'erreurs ne s'applique pas dans le cas général. Dans ce cas, les propriétés mathématiques du modèle et ses propriétés d'identifiabilité et d'estimation des paramètres structuraux ou d'optimalité ont été peu étudiées. Les travaux de TEN BERGE (1984) établissent cependant les liens de certains sous-modèles avec le problème posé par la rotation simultanée de la configuration au sens du critère VARIMAX. L'étude des propriétés mathématiques des modèles descriptifs s'attachant à l'analyse d'échantillons de matrices aléatoires, entreprise dans le cadre de l'Analyse en Composantes Principales d'un Cube de données par JAFFRENOU (1978), MIZERE (1981) et POLIT (1986) invite à la prudence dans la mise en œuvre de modèles structurels correspondant au cas (b). Le grand nombre de paramètres à estimer, parmi lesquels le rang de  $M_k$  joue un rôle important et très peu maîtrisé, conduit en effet à douter de la stabilité des résultats mis en évidence par une approche aussi générale. La formulation des modèles de BLOXON (1974) et de KRANE (1978) dans le cadre de l'analyse des structures de covariance est limitée au cas de matrices  $M_k$  diagonales et ne traite donc pas le cas général. Cette formulation du problème offre cependant un champ d'étude prometteur, permettant en particulier de s'affranchir des aspects tautologiques de la formulation algébrique initiale du problème. KRANE (1978) formule le problème en termes d'ajustement aux moindres carrés des dissimilarités et propose de construire la solution à l'aide d'une méthode de gradient. Pour les mêmes raisons que précédemment, il nous semble préférable d'adopter la démarche algorithmique de LEEUW et HEISER (1977) qui proposent une généralisation de leur algorithme de sous gradient adaptée à cette situation (cf. d'AUBIGNY (1984)). Nous n'en poursuivons pas l'étude ici, notre exposé se limitant à la donnée d'une seule matrice de dissimilarité.

**Remarque :** Le cadre général d'une analyse de données de dissimilarité sous contraintes est exprimé en termes de mélange de contraintes externes définies par des matrices  $Y_1$  et  $Y_2$  et de contraintes structurelles. On est alors conduit à étudier les propriétés d'un modèle de courbes de croissance lorsque la matrice  $\Sigma$  est structurée (cf. § III.1.1). Ce problème a été peu étudié du point de vue de la statistique inférentielle. Cependant, RAO (1967) étudie le modèle particulier d'autocorrélation

$$\Sigma = \{\sigma^2 \rho^{|i-j|}\}_{i,j=1,\dots,p}$$

et surtout le modèle d'analyse factorielle

$$\Sigma = \Lambda \Phi \Lambda + \sigma^2 I$$

qui est un cas particulier du modèle de JORES-KOG. Ce modèle permet en théorie de traiter des problèmes statistiques liés aux modèles des Différences Individuelles sous contraintes linéaires. Mais aucune application pratique exploitant complètement cette opportunité ne nous semble avoir été publiée.

### III.2. CONTRAINTES ORDINALES PORTANT SUR LA CONFIGURATION X

Les contraintes étudiées dans les paragraphes précédents permettent d'enrichir très fortement la méthodologie du Codage Multidimensionnel. Elles ont cependant pour caractéristique commune de définir des ensembles de contraintes décrites en terme d'appartenance à des sous-espaces vectoriels  $\langle E_1, F_2 \rangle$  de l'espace  $F \otimes E$  décrit par les solutions du problème.

Nous traitons ici un deuxième axe de généralisation de ces méthodes. On se propose en effet d'étendre l'approche du problème en termes de Q-ACPVI. Dans le cadre linéaire l'étude du triple  $(Z^1 P_2, M, N)$  repose sur les propriétés mathématiques du projecteur orthogonal  $P_2$  sur un sous-espace vectoriel fermé  $F_2$  de  $F$ . Une telle approche est clairement généralisable au cas d'un ensemble de contraintes proximal quelconque  $K_2$  inclus dans  $F$ . Dans ce cas en effet, le projecteur  $P_2$  est défini (cf. ROCKAFFELLAR (1970)) et conduit à l'analyse des données induites  $\hat{Z} = Z^1 P_2$ , menée dans le cadre de l'étude  $(Z^1 P_2, M, N)$ , bien que  $K_2$  ne soit pas forcément décrit par la donnée de variables instrumentales. Dans un cadre linéaire, on mène de façon analogue une Analyse en Coordonnées Principales hors Variables Instrumentales, puisque les méthodes présentées reposent sur la connaissance du projecteur  $P_2$  associé à une décomposition de  $F_0$  en somme directe de deux sous espaces N-orthogonaux

$$F_0 = F_1 \oplus F_2$$

Dans le cas de contraintes d'appartenance à un ensemble proximal caractérisé par un cône convexe polyédrique fermé étudié dans la suite, la notion de cônes polaires (cf. ROCKAFELLAR 1970) généralise la notion d'espaces vectoriels supplémentaires orthogonaux et permet donc de traiter de façon analogue le cas d'une analyse hors contraintes instrumentales.

Nous traitons ici à titre d'exemple le cas de contraintes ordinales imposées aux coordonnées des unités statistiques sur les axes de représentation fournis par l'analyse. De telles contraintes structurelles se traduisent dans  $F \otimes F$  par un ensemble de contraintes défini par un cône convexe polyédrique fermé  $K_2$ . On traite ainsi dans toute sa généralité le problème posé par NOMA et JOHNSON (1979) qui recherchent une représentation bidimensionnelle des dissimilarités observées. Les données sont alors collectées dans le

cadre d'un plan d'expérience croisant deux facteurs et les auteurs proposent de décrire la surface réponse à l'aide d'une méthodologie de codage multidimensionnel. Les dissimilarités observées  $\delta_{ij}$  décrivent de façon approchée la surface réponse par ses propriétés métriques et l'on impose le choix d'un repère de coordonnées tel que l'ordre des modalités des facteurs soit respecté par les coordonnées des points représentatifs sur les deux axes de représentation.

Une telle stratégie présente deux caractéristiques originales. Tout d'abord, la surface réponse est alors décrite par ses propriétés topologiques déterminées par les distances interpoints et non plus par une expression analytique connue de façon approchée par la fonction liant une variable réponse à un ensemble de facteurs connus a priori. De plus, le recours à une transformation monotone de la variable réponse a pour but d'effacer d'éventuels termes d'interaction entre facteurs, selon une démarche adoptée auparavant par KRUSKAL (1965) dans le cadre plus classique de l'Analyse de la Variance appliquée à l'étude d'un plan factoriel croisant deux facteurs, pour lequel l'interprétation de la réponse en termes de distances n'est pas imposée. Nous donnons ainsi une base formelle à la démarche heuristique de NOMA et JOHNSON. Ecrivons en effet :

$$\hat{Z} = \varphi(Y_2)$$

où  $\varphi$  est une transformation définie dans la classe particulière des applications caractérisée par les transformations monotones croissantes  $m_j$ :

$$\hat{z}^j = m_j(y^j) \quad j = 1, 2, \dots, p \quad (76)$$

la notation abusive (76) traduit ici le respect des contraintes d'ordre

$$(y_i^j < y_l^j) \Rightarrow (\hat{z}_i^j < \hat{z}_l^j) \quad i, l \in I$$

Le modèle structurel n'est cependant pas complètement décrit par cette implication. En pratique en effet, on doit de plus distinguer différents types de contraintes ordinales selon qu'on admet ou non l'éventualité d'égalités. Ce problème technique complémentaire se traduit lors de la construction de l'ensemble des contraintes décrivant les relations (76) (cf. KRUSKAL (1964)). Malgré son importance concrète, ce problème est ignoré ici afin de centrer notre étude sur les principes généraux de traitement.

Nous traitons ici ce type de contraintes dans le cadre d'une  $\Sigma$ -stratégie (cf. CAILLEZ (1984)) afin d'une part de justifier la solution heuristique proposée par NOMA et JOHNSON et d'autre part de montrer la parenté liant ce problème à la classe des méthodes de Codage Optimal traitées en régression non linéaire (cf. YOUNG, de LEEUW et TAKANE (1978)), adaptées à l'étude de régresseurs mesurés selon des échelles mixtes (nominales, ordinales, numériques). YOUNG, de LEEUW et TAKANE (1978) montrent

que la formulation donnée en termes de contraintes traduites par des cônes polyédriques généralisent le cas de contraintes linéaires (définies par des régresseurs numériques), un sous-espace vectoriel étant un cône polyédrique particulier.

Remarquons tout d'abord que le problème posé par NOMA et JOHNSON généralise le modèle de CARROLL et al. (1976) lorsqu'on prend en compte les contraintes ordinales. L'équation (76) traduit alors une transformation des données qui modifie tant la structure d'association entre variables que la mesure des proximités inter-individuelles. On se place donc ici dans  $F \otimes E$  où  $E$  décrit l'ensemble des sous-espaces vectoriels de dimension  $p$  de  $F$  (NOMA et JOHNSON fixent  $p=2$ ).

Considérons les configurations centrées et notons  $\underline{a} = \text{Vec}(A)$  le vecteur de  $F_0 \otimes E$  associé à la matrice  $A$  de l'application linéaire

$$a : E^* \rightarrow F_0$$

dans les bases  $\{\underline{e}^{*j}, j=1, 2, \dots, p\}$  de  $E^*$  et  $\{\underline{f}_k, k=1, 2, \dots, n\}$  de  $F$ . L'ajustement du modèle (76) s'écrit:

$$\underline{z} = \hat{\underline{z}} + \underline{e} \quad \hat{\underline{z}}^j = m_j(Y^j) \quad j = 1, 2, \dots, p$$

On décompose alors l'espace vectoriel  $G = F_0 \otimes E$  de façon canonique en somme directe

$$G = \bigoplus_{j=1}^p G_j \quad G_j = F_0 \otimes e_j$$

où le sous espace vectoriel  $G_j$  de  $G$  a pour dimension  $(n-1)$  et permet de traduire dans  $G$  les contraintes associées à la variable  $Y^j$

$$\hat{\underline{z}}^j = m_j(Y^j) \quad \underline{z}^j = \hat{\underline{z}}^j + \underline{u}^j \quad (77)$$

$\hat{\underline{z}}^j$  est plongé dans  $G$  par l'injection canonique usuelle de  $F_0$  dans  $G_j$  définie par

$$\begin{aligned} \psi_j : F_0 &\rightarrow G_j \\ \underline{x} &\mapsto \psi_j(\underline{x}) = \underline{x} \otimes e_j \end{aligned}$$

On note alors son image par

$$\hat{\underline{z}}^j = \hat{\underline{z}}^j \otimes e_j$$

Notons de plus  $\hat{\underline{Z}}^j$  la matrice définie par

$$\hat{\underline{z}}^j = \text{Vec}(\hat{\underline{Z}}^j)$$

Seule la  $j$ -ème ligne de la matrice  $\hat{\underline{Z}}^j$  est non nulle et égale à  $\hat{\underline{z}}^j$ . Notons  $K_j$  le cône convexe polyédrique fermé décrivant dans  $G_j$  les contraintes monotones induites par la variable  $Y^j$ .

$K_j$  est l'image par  $\psi_j$  du cône polyédrique  $C_j$  décrivant les contraintes (77) dans  $F_0$  (cf. d'AUBIGNY (1975, 1979)) et l'on note en conséquence :

$$K_j = C_j \otimes e_j$$

Par définition, la solution  $\hat{z}^j$  du problème s'écrit :

$$\hat{z} = \sum_{j=1}^p \hat{z}^j \quad , \quad \hat{z}^j \in K_j$$

Par suite, la prise en compte simultanée des  $p$  contraintes de type (77), associées aux variables  $Y^j$ ,  $j=1, \dots, p$  se traduit alors dans  $G$  par la définition du cône convexe polyédrique fermé  $K$  inclus dans  $G$

$$K = \bigoplus_{j=1}^p K_j$$

**Remarque 1 :** Tout sous-espace vectoriel de  $F_0 \otimes E$  est un cas particulier de cône convexe polyédrique fermé  $K$  inclus dans  $F_0 \otimes E$ . Par suite, les contraintes linéaires traitées par CARROLL et al. (1976) et étendues dans les paragraphes précédents sont un cas particulier du problème traité ici.

La donnée de  $K$  permet d'exprimer la recherche d'une solution comme un problème particulier de programmation quadratique sous contraintes de type inégalité :

$$\text{Min } \{ \|z - t\|_{N \otimes M}^2 ; t \in K \}$$

que l'on peut résoudre de façon classique (cf. BOUGHAZI (1987), d'AUBIGNY et al. (1988)). Il est cependant préférable d'utiliser un algorithme adapté à la structure particulière de  $K$  comme somme directe de cônes convexes polyédriques fermés. On peut par exemple utiliser l'algorithme des projections successives de DYKSTRA (1983), bien que ses propriétés numériques soient critiquées par d'AUBIGNY et al. (1988).

Nous présentons ici un algorithme différent traité dans le cas  $M=I$  par d'AUBIGNY (1979) que nous étendons directement au cas d'une métrique  $M$  décrite sur  $E$  par une matrice diagonale de poids  $D_m$  vérifiant  $m_j > 0$  et  $\sum m_j = 1$ . Dans ce cas, en effet, la norme induite sur  $F \otimes E$  s'écrit :

$$\|z - t\|_{N \otimes D_m}^2 = \sum_{j=1}^p m_j \|z^j - t^j\|_N^2 \quad (78)$$

où  $m_j \langle x, y \rangle_N$  définit un produit scalaire dans  $G^j$ . Par suite les résultats de d'AUBIGNY (1979) s'étendent de façon immédiate et permettent de construire le tenseur solution  $\hat{z}$  à partir de la résolution des problèmes de programmation quadratique

$$\text{Min } \{ \|z^j - t^j\|_N^2 ; t^j \in C_j \}$$

Dans le cas où  $C_j$  est défini comme un cône de contraintes décrivant les relations de monotonie (76), on résoud ce problème à l'aide d'un algorithme usuel de régression isotone. Ces résultats fournissent en particulier un indice d'association monotone multiple permettant de juger de la qualité globale de l'ajustement de  $t$  à  $z$ . De plus, puisque le problème se traduit géométriquement en termes de projection  $N \otimes D_m$ -orthogonale sur le cône polyédrique  $K$ , la solution optimale  $t^*$  appartient à l'une des faces de  $K$  et donc au sous-espace vectoriel de  $F$  support de cette face. Par suite, la solution  $t^*$  est interprétable a posteriori comme la solution d'un problème de régression linéaire multiple, dont la qualité globale est évaluée à l'aide des outils de diagnostics décrits au paragraphe III.1.1. On peut en particulier évaluer, dans le cas présent, le bras de levier de chaque contrainte et dégager les contraintes influentes. Ces outils, utilisés conjointement à une analyse des résidus dans la régression, apportent une connaissance complète de la stabilité de la solution.

**Proposition 7:** Soit  $P_j$  le projecteur  $N$ -orthogonal sur le cône polyédrique  $C_j$  inclus dans  $F$  défini par

$$t^j \in C_j \Leftrightarrow t^j = m_j(y^j)$$

où  $m_j$  est une fonction monotone vérifiant

$$(y_i^j \leq y_k^j) \Rightarrow (t_i^j \leq t_k^j)$$

La solution du problème

$$\text{Min } \{ \text{Trace } ({}^t(Z - T)D_m(Z - T)N) ; t^j \in C_j \quad j = 1, 2, \dots, p \}$$

est donnée par

$$\hat{Z} = {}^t \{ P_1 y^1, \dots, P_p y^p \} \quad (79)$$

**Preuve :** L'expression (78) de la fonction objectif permet de décomposer l'analyse en fonction des  $G_j$ . La solution s'écrit dans  $G_j$  en fonction du projecteur orthogonal  $\Pi_j$  sur le cône polyédrique  $K_j$ :

$$\hat{z}^j = \Pi_j y^j \Leftrightarrow \hat{z}^j = \Pi_j y^j \quad (80)$$

$$\Pi_j : G \rightarrow K \quad P_j : F_0 \rightarrow C_j$$

où  $\hat{z}^j = \hat{z}^j \otimes \mathbf{e}_j$ . On construit ainsi la solution  $\hat{z} \in G$ :

$$\hat{z} = \sum_{j=1}^p \hat{z}^j = \sum_{j=1}^p \hat{z}^j \otimes \mathbf{e}_j$$

Soit, en identifiant les projecteurs à leurs matrices de représentation dans les bases définies plus haut:

$$\begin{aligned} \hat{z} &= \sum_{j=1}^p \Pi_j \mathbf{y}^j = \sum_{j=1}^p P_j \mathbf{y}^j \otimes \mathbf{e}_j \\ &= [P_1, \dots, P_p] \mathbf{y} \end{aligned}$$

Par suite  $\hat{z}$  représente dans  $G = F_0 \otimes E$  la matrice

$$\hat{Z} = [P_1 \mathbf{y}^1, \dots, P_p \mathbf{y}^p]$$

associée à la configuration recherchée.  $\square$

Une telle analyse se prête cependant à une visualisation des données dans le seul cas où  $p = 1, 2$  ou  $3$ , puisque la non corrélation des variables reconstruites  $P_j \mathbf{y}^j$  n'est pas assurée. Dans le cas  $p$  quelconque, on peut alors recourir à l'Analyse en Composantes Principales de  $(\hat{Z}, M, N)$  dans le but de construire une représentation résumée. Bien que les axes principaux de coordonnées ne respectent pas en général les contraintes ordinales du plan d'expérience, on peut utiliser ces représentations approchées (du fait des projections) des variables reconstruites sur l'espace principal retenu afin d'exploiter l'information fournie par la structure du plan, à l'aide de techniques de variables supplémentaires.

**Remarque :** La contrainte de monotonie s'exprime matriciellement par la contrainte de structure (79) dans laquelle les coordonnées de  $P_j$  ne sont pas connues a priori. De plus, on peut réécrire  $\hat{Z}$  sous la forme

$$\hat{Z} = [P_1, \dots, P_p] \tilde{Y} \quad (81)$$

où  $\tilde{Y}$  est une matrice  $n \times p$  de  $j$ -ème colonne  $\mathbf{y}^j \otimes \mathbf{e}_j$  connue, dont la structure particulière est due au traitement séparé des contraintes associées à chaque variable facteur.

Remarquons que cette formalisation du problème décrit la structure de l'ensemble des contraintes dans l'espace  $G$  et non dans  $F$ . L'extension du cadre formel de l'ACPVI au

cas de contraintes définissant un sous-ensemble proximal de  $F$  n'est donc pas suffisante pour traiter ce problème. Construire une transformation monotone de chacune des variables initiales influe sur leur structure de corrélation mais aussi sur leur positions relatives dans  $E$ . De plus, le problème n'est pas exprimable, à notre connaissance, sous la forme simple

$$\hat{Z} = Y_1 Z' Y_2$$

ce qui nécessiterait une décomposition de  $P = [P_1, \dots, P_p]$  sous la forme  $P = Y_1 \otimes Y_2$ , c'est-à-dire une hypothèse de décomposabilité. Une telle hypothèse se traduirait géométriquement par une formalisation de l'ensemble des contraintes sous la forme  $K = K_2 \otimes K_1$  où  $K_2$  (resp  $K_1$ ) est un sous-ensemble proximal de  $F_0$  (resp  $E$ ) ou bien, comme dans le cas linéaire, à une identification près  $K = \varphi(K_2 \otimes K_1) = \langle K_2, K_1 \rangle$ . A notre connaissance, une telle formulation n'est pas possible, rendant inadaptés à la situation présente les théorèmes de convergence d'algorithmes généralement donnés en référence pour justifier leur utilisation dans le cadre des méthodes de Codage Multidimensionnel.

### III.3. OPTIMISATION DE CRITERES D'ECART ENTRE DISSIMILARITES

Dans le paragraphe précédent, nous avons traité le problème de NOMA et JOHNSON dans le cas où la donnée est un tableau de données de profil  $Y$ . Dans leur expérience, la donnée est une matrice de dissimilarité. Lorsque celle-ci est Euclidienne, l'utilisation de l'ACPVI permet de construire la représentation  $Z$  que l'on ajuste ensuite aux contraintes. Mais la satisfaction du modèle de NOMA et JOHNSON impose de plus une contrainte de rang pour  $Z$  qui n'est pas prise en compte par cette démarche. Par suite, il paraît plus raisonnable de construire une configuration  $Z$  à l'aide d'une méthode de Codage Multidimensionnel, en itérant deux étapes

- (a) construction d'une configuration  $Z^{(v)}$ ,
- (b) ajustement des contraintes de plan  $\hat{Z}^{(v)}$  et recalcul des distances ajustées.

La mise en œuvre d'une telle approche est discutée plus bas, à l'aide de l'algorithme de de LEEUW et HEISER. Elle s'inscrit en effet dans le cadre des méthodes optimisant un critère d'écart entre tableaux de dissimilarité.

Plus généralement, dans le paragraphe III.1, l'introduction de contraintes portant sur la configuration  $X$  est menée en termes d'optimisation des fonctions critères  $G(Z, Y)$ ,

$E(Z, Y)$  et d'indices dérivés. Nous avons montré que la solution s'exprime dans les deux cas en fonction de l'opérateur  $U_Z = W_Z N$ . La matrice  $W_Z$  est alors déduite de  $\Delta(Z)$  par la transformation de Torgerson lorsque  $N = D_p$ . Dans le cadre d'une analyse sous contraintes linéaires décrites dans F, l'Analyse en Coordonnées Principales appliquée aux tableaux de dissimilarité Euclidienne met en œuvre des outils analogues à ceux qui sont fournis par l'ACPVI pour l'analyse de données de profils. Cependant, le problème se complique dans deux situations :

D'une part, lorsque la métrique  $N$  n'est pas diagonale. Par exemple lorsque  $Q_{|x|}$  définit sur  $I$  une pondération qui n'est pas décomposable sous la forme  $q_{ij} = p_i p_j$ . La projection  $N_Q$ -orthogonale sur le supplémentaire  $N_Q$ -orthogonal à la droite des constantes dans  $F$  perd alors beaucoup de son caractère intuitif. La matrice des produits scalaires  $W_Z$  déduite de  $\Delta(Z)$  par la projection

$$w_Z = Q_N \otimes Q_N h_Z$$

où

$$Q_N = I - \frac{1}{\mathbf{1}' N \mathbf{1}} \mathbf{1} \mathbf{1}' N$$

correspond alors à une opération de centrage de la configuration (inconnue)  $Z$  par rapport à rapport au point

$$\underline{s} = \frac{\mathbf{1}' Z N \mathbf{1}}{\mathbf{1}' N \mathbf{1}}$$

qui ne s'interprète plus en terme de barycentre

D'autre part, lors de l'étude de données de dissimilarité quelconques, lorsque celles-ci sont soit non Euclidiennes soit des distances Euclidiennes bruitées par des erreurs de mesure, le passage de  $\Delta$  à  $W$  pose de gros problèmes de modélisation statistique (on parle souvent à ce propos de méthodes négligeant les erreurs) et conduit le plus souvent à une matrice  $W$  non gramienne.

On préfère alors optimiser des fonctions critères mesurant directement l'écart entre dissimilarités et distances reconstruites. Nous avons vu au chapitre II que le critère le plus proche de  $E(Z, Y)$  et exprimé en termes de dissimilarité est défini par l'indice SSTRESS. Il mesure l'écart entre deux dissimilarités par la quantité :

$$\begin{aligned} SS(Z, Y) &= \sum_{i < j} q_{ij} (\delta_{ij}^2 - d_{ij}^2(Y))^2 = \| \underline{\delta}^{*2} - \underline{d}^{*2} \|_{D_Q}^2 \\ &= \text{Trace} ((\Psi_Z - \Psi_Y) D_Q (\Psi_Z - \Psi_Y)) \end{aligned}$$

$$\begin{aligned}
&= \text{Trace} (T(W_Z - W_Y) {}^t D_Q T(W_Z - W_Y) {}^t T) \\
&= \text{Trace} (Q_0 (W_Z - W_Y) N_Q (W_Z - W_Y)) \\
&= \| \underline{w}_Z - \underline{w}_Y \|_{N_Q \otimes Q_0}^2
\end{aligned}$$

Rappelons que nous avons noté  $Q_0$  le projecteur sur l'orthogonal de la droite des constantes de  $F$  muni de la métrique identité et  $N_Q$  est la métrique induite sur  $F_0$  par la pondération  $Q_{|x|}$ .

$$N_Q = D_Q - Q$$

On obtient ainsi un critère mesurant l'écart entre les tenseurs  $\underline{w}_Z$  et  $\underline{w}_Y$  au sens des moindres carrés généralisés. Mais puisqu'en général  $N_Q \neq Q_0$  lorsque la métrique  $Q$  est non triviale, c'est-à-dire lorsqu'il existe deux paires  $(i,j)$  et  $(k,l)$  pour lesquelles  $q_{ij} \neq q_{kl}$ , la norme utilisée est distincte de la norme de Hilbert Schmidt définie en III.2 pour mesurer l'écart entre les opérateurs  $W_Z N$  et  $W_Y N$ .

Par suite la construction de la solution nécessite la mise en œuvre d'une procédure itérative. Il en est de même de l'optimisation des critères  $\psi(\alpha, \beta, r)$  proposés par OBENCHAIN et en particulier de l'indice de STRESS

$$S(Z, Y) = \| \underline{d}(Z) - \underline{d}(X) \|_{D_Q}^2$$

présenté au chapitre II. L'itération proposée par GUTTMAN (1968) puis justifiée et étendue par de LEEUW et HEISER (1977, 1980) consiste alors à itérer les trois étapes suivantes jusqu'à convergence

(1) Pour une configuration  $X^{(v)}$  donnée, on évalue comme précédemment (25) la matrice

$${}^t Z^{(v+1)} = N_Q N(X^{(v)}) {}^t X^{(v)}$$

On obtient ainsi une configuration de rang au plus égal à celui de  $X^{(v)}$ , qui ne respecte pas les contraintes en général.

(2) Respect des contraintes : on projette les vecteurs de cette configuration sur le sous-espace de  $F$  défini par les contraintes :

$${}^t Y^{(v+1)} = P_2 {}^t Z^{(v+1)}$$

(3) Normalisation

$${}^tX^{(v+1)} = \frac{{}^tY^{(v+1)}}{\|{}^tY^{(v+1)}\|_{N_Q}}$$

Les contraintes sont alors prises en compte dans l'étape (2). Ces auteurs notent la parenté de cette itération avec la méthode de la puissance itérée utilisée en ACP sans contrainte. L'étape (2) de l'algorithme ci-dessus est une modification de l'algorithme initial de de LEEUW et HEISER(1977) présentée par MEULMAN et HEISER (1984), et montre les liens unissant cette approche à la logique de l'ACPVI, par comparaison aux résultats présentés en III.2.1. On peut ainsi clairement interpréter l'approche de MEULMAN et HEISER comme l'extension au cadre non linéaire d'un algorithme d'ACPVI, caractérisé par le projecteur  $N_Q$ -orthogonal  $P_2$  sur le sous-ensemble proximal  $K_2$  inclus dans  $F_0$ . L'étude menée dans le cas du modèle général de CARROLL et al. (1976) ou dans le cas du modèle de NOMA et JOHNSON (1979) montre que le traitement de modèles généraux de contraintes nécessite de généraliser l'étape (2) à une classe plus large de contraintes exprimées dans  $F \otimes F$  et traitées à l'aide du formalisme tensoriel.

Dans le cas de contraintes de la forme

$$y \in K_2 \otimes K_1$$

où  $K_1$  (resp.  $K_2$ ) est un sous-ensemble proximal de  $E$  (resp.  $F$ ), on obtient une classe de problèmes généralisant directement l'approche de CARROLL et al. (1976) et le respect des contraintes impose de réaliser la phase (2) décrite en fonction des projecteurs orthogonaux  $P_1$  et  $P_2$  sur les cônes de contraintes  $K_1$  et  $K_2$  par :

$$Y^{(v+1)} = P_1 Z^{(v+1)} {}^tP_2 \Leftrightarrow y^{(v+1)} = P_2 \otimes P_1 z^{(v+1)}$$

On est ainsi conduit à compléter l'étape (2) de l'algorithme par une projection dans  $E$ .

Lorsqu'on impose des contraintes du type traité par NOMA et JOHNSON ou plus généralement dans le cas de contraintes de structure, il convient de substituer à l'étape (2) une procédure de projection sur un sous-espace proximal  $K$  de  $F_0 \otimes E$ . On écrit alors

$$y^{(v+1)} = P_k z^{(v+1)}$$

où le projecteur  $P_k$  n'est pas forcément décomposable en  $P_2 \otimes P_1$ . La nature géométrique du problème reste inchangée mais sa résolution algorithmique peut être moins aisée et nécessite une étude cas par cas. La convergence de cette procédure algorithmique n'a pas été étudiée dans un cadre aussi général.

## CHAPITRE IV

### ANALYSE DU TRIPLE $(I, \delta_I, Q_{kl})$ SOUS CONTRAINTES RELATIONNELLES

Les contraintes prises en compte dans les paragraphes précédents portent sur la configuration résultant de l'analyse, c'est à dire la matrice  $X$  des coordonnées des points représentant les unités statistiques dans  $E$ . Nous avons vu par exemple qu'une Analyse en Coordonnées Principales sur Variables Instrumentales permet de construire une configuration  $X$  respectant l'information a priori décrite dans un cadre linéaire par une matrice  $Y_2$ .

Cependant de nombreuses méthodes de collecte de l'information conduisent à exprimer les contraintes traduisant des connaissances a priori, en termes de distances ou de dissimilarités. Les raisons sont alors identiques à celles qui président à la mesure des dissimilarités observées  $\delta_{ij}$  de préférence à des données de profil fournies par les mesures  $Z_{ij}$ . Les contraintes définissent alors une relation binaire sur  $I$  et l'on interprète les dissimilarités comme des valuations des arêtes du graphe  $G = (I, U)$  de cette relation. Nous nous proposons donc d'étudier ici ce problème. Tout d'abord nous reformulons dans le cadre de l'analyse des tableaux de dissimilarités, les travaux de LEBART (1969, 1973) et de BANET et LEBART (1984) dans le domaine voisin de l'Analyse en Composantes Locales. Ces auteurs étudient le cas de contraintes spatiales qui sont alors décrites par un graphe de contiguïté. Nous étendons ensuite cette approche au cas de graphes valués associés à une relation floue d'appartenance, puis nous étudions le cas général de contraintes définies par la donnée d'une ou de plusieurs matrices de dissimilarité.

#### IV.1. ANALYSE EN COORDONNEES PRINCIPALES

Les contraintes prises en compte dans les paragraphes précédents portent directement sur la configuration recherchée  $X$ . La matrice des contraintes  $Y$  définit cependant une structure de proximités Euclidiennes entre les unités statistiques, déterminée par la matrice  $D(Y)$  de terme général

$$d_{ik}(Y) = \left[ \sum_{j=1}^p (y_i^j - y_k^j)^2 \right]^{1/2}$$

Nous interprétons ici cette contrainte comme une relation valuée particulière. Bien entendu, tout indice de dissimilarité n'étant pas nécessairement Euclidien, la classe des contraintes définies par une relation valuée conduit à une méthodologie plus générale que l'Analyse en Coordonnées Principales sur (ou hors) Variables Instrumentales exposée plus haut.

#### IV.1.1. LE CAS DES CONTRAINTES DE GRAPHE

Considérons l'étude du triple  $(I, \delta_I, Q_{I \times I})$  dans laquelle on se propose d'étudier le plongement Euclidien du tableau de dissimilarité  $\Delta$  sous une contrainte déterminée par une relation binaire  $R$  définie sur  $I$ .

##### IV.1.1.1. Description métrique de l'ensemble des contraintes

Soit  $G = (I, U)$  un graphe non orienté de sommets les unités statistiques  $i$  de  $I$ . On note  $u_{ij}$  l'arête de  $U$  joignant les sommets  $i$  et  $j$  du graphe. Le graphe  $G$  traduit ici les contraintes imposées lors de l'analyse d'un tableau de dissimilarité  $\Delta$ . De façon classique, la donnée du graphe  $G$  est équivalente à la donnée d'une relation binaire  $R$  définie sur  $I$  traduisant par exemple des contraintes de sériation temporelle (LE FOLL (1979)) ou encore des contraintes de contiguité spatiale (LEBART (1969, 1984)). Ainsi lorsque  $I$  est l'ensemble des départements français, KENDALL (1971) prend en compte la relation

$$i R l \Leftrightarrow i \text{ et } l \text{ sont deux départements contigus.}$$

Un troisième exemple de contraintes décrites par une relation binaire particulière et d'un emploi très fréquent en statistique intervient en Analyse Discriminante.  $G$  est alors un graphe  $k$ -partite et décrit la relation d'équivalence associée à une partition de  $I$  en  $k$  classes fixées a priori. HUBERT (1982) et d'AUBIGNY (1988) montrent que la prise en compte de telles contraintes de structure de plan permettent d'étendre la méthodologie de l'Analyse de la Variance au cas où la variable réponse mesure les dissimilarités entre unités statistiques. D'un point de vue exploratoire, l'introduction de contraintes de graphe permet d'étendre de façon importante les méthodes d'analyse des données à la description de structures non linéaires. Elle ne se limite d'ailleurs pas à la prise en compte de structures agrégatives. Ainsi, en présence d'un effet Guttman ou horseshoe, couramment obtenu dans les problèmes de sériation (cf. KENDALL (1971) pour un

exemple en archéologie), cette méthodologie permet de décrire de façon discrétisée les géodésiques de la variété différentiable décrite par les données de dissimilarité. La prise en compte du graphe ainsi construit permet de représenter les données soit le long de cette variété, soit dans son complémentaire. On retrouve dans le cadre de l'analyse des tableaux de dissimilarité une technique analogue à l'Analyse des Correspondances Hors Tendances proposée par HILL et GAUCH (1980). TER BRAAK (1985, 1986) produit une étude détaillée de cette méthode et de ses applications en Ecologie.

On considère alors la matrice de connexion  $Q$  associée au graphe  $G$ .  $Q$  est une matrice carrée d'ordre  $n$  de terme générique

$$q_{ij} = \begin{cases} 1 & \text{si } i R j \\ 0 & \text{sinon} \end{cases}$$

La matrice  $Q$  est symétrique lorsque le graphe  $G$  est non orienté et nous supposons dans un premier temps que la relation  $R$  est antiréflexive.  $G$  est alors un graphe simple et  $Q$  est à diagonale nulle. Le graphe  $G$  définit une structure métrique sur  $I$ . En effet posons les définitions suivantes (cf. BERGE (1973))

**Définition 1 :** On appelle chemin sur un graphe  $G$ , toute suite d'arêtes  $(u_1, \dots, u_k)$  telle que pour tout  $s \in [2, k]$ ,  $u_{s-1}$  et  $u_s$  admettent un sommet en commun. La longueur du chemin est le nombre d'arêtes qui le composent.

**Définition 2 :** On appelle distance entre deux sommets  $i$  et  $j$  d'un graphe non orienté  $G$ , la longueur  $d_{ij}$  du plus court chemin joignant  $i$  à  $j$ .

On vérifie sans peine que cette définition est consistante avec les axiomes définissant une distance:

$$\begin{aligned} d_{ij} &\geq 0 \\ d_{ij} &= d_{ji} \\ d_{ij} &\leq d_{ik} + d_{kj} \end{aligned}$$

Cependant la distance ainsi construite n'est pas Euclidienne en général. Considérons en effet la matrice  $S = Q + I$  où  $I$  est la matrice unité d'ordre  $n$  et  $S$  est la matrice de connexité associée au graphe  $G$  complété par l'ajout d'une boucle en chaque sommet.  $S$  s'interprète comme une matrice de pseudo produits scalaires. On est donc confronté au problème classique de la constante additive permettant d'associer à  $S$  une matrice de Gram. Soit donc  $-\lambda^*$  la plus petite valeur propre de  $S$ . Posons alors  $S^* = S + \lambda^* I$ . GUTTMAN montre alors le lemme suivant

**Lemme 1 :** *Si il existe une arête (i,j) telle que  $q_{ij}=1$ , alors S admet au moins une valeur propre négative.*

**Preuve :** Soit  $-\lambda^*$  la plus petite valeur propre de S.  $S^*$  est une matrice symétrique. Elle est gramienne si et seulement si le déterminant de tout mineur principal d'ordre deux de  $S^*$  est positif ou nul. En particulier pour le mineur (i,j)

$$\det \begin{pmatrix} 1+\lambda^* & 1 \\ 1 & 1+\lambda^* \end{pmatrix} = \lambda^* (\lambda^* + 2) \geq 0$$

et donc  $\lambda^* \geq 0$ , les valeurs propres de Q étant de module inférieur à 1. ¶

Notons de plus  $R_i$  l'ensemble des sommets du graphe G adjacents au sommet i et  $H_i$  son complémentaire :

$$\begin{array}{lll} j \in R_i & \text{si } (i,j) \in U & (\Leftrightarrow q_{ij} = 1) \\ j \in H_i & \text{si } (i,j) \notin U & (\Leftrightarrow q_{ij} = 0) \end{array}$$

et imposons à toute représentation métrique de G de vérifier la condition de pertinence

$$(2) \quad d_{ij} \leq d_{ik} \quad \text{pour tout } j \in R_i \text{ et } k \in H_i \quad (\forall i, i \in I)$$

GUTTMAN (1965) montre alors la non vacuité de la définition 2, à l'aide du théorème d'existence

**Théorème 1:** *Si  $\text{card}(I) = n$ , il existe un espace Euclidien de dimension m avec  $m \leq n-1$ , dans lequel G est représentable, de sorte à vérifier (1) et (2).*

**Preuve :** En effet les valeurs propres de  $S^* = Q + (\lambda^* + 1)I$  sont positives ou nulles et au moins une d'entre elles est nulle. Par suite  $m \leq n-1$ . Posons donc

$$S^* = {}^t X X \quad X \in \mathcal{M}_{m,n}(\mathbb{R})$$

alors pour tout  $(i,j) \in I \times I$

$$\begin{aligned} d_{ij}^2(X) &= \sum_s (x_i^s - x_j^s)^2 = s_{ii}^* + s_{jj}^* - 2s_{ij}^* \\ &= 2(\lambda^* + 1 - s_{ij}) \end{aligned}$$

Par suite, la contrainte (2) est vérifiée puisque

$$\begin{aligned} s_{ij} = 1 &\Rightarrow d_{ij}(X) = \sqrt{2\lambda^*} \\ s_{ij} = 0 &\Rightarrow d_{ij}(X) = \sqrt{2(\lambda^* + 1)} \quad \parallel \end{aligned}$$

**Définition 3 (GUTTMAN) :** On appelle *dimension linéaire du graphe G* la valeur minimale  $r$  d'un espace permettant une telle représentation Euclidienne de  $I$ .

Nous interprétons ici le formalisme algébrique présenté au chapitre II comme une description de la dualité entre sommets interprétés comme éléments de  $F$  et arêtes d'un graphe, interprétées comme éléments de  $F \wedge F$ . L'application linéaire  $t$  de matrice  $T$  définit alors une isométrie partielle entre l'espace de représentation des unités statistiques  $F_0$  et l'espace de représentation des couples  $F \wedge F$ . Par suite nous interprétons  $Q$  dans ce schéma linéaire, comme une pondération définie sur  $I \times I$  et nous associons à  $Q$  une matrice diagonale  $D_Q$  définissant une (semi-)métrique sur  $F \wedge F$ . On définit ainsi une (semi-)métrique sur  $F_0$  par la transformation  $T$  :

$$N_Q = D_Q \cdot Q \quad q_{i+} = \sum_{k \neq i}^n q_{ik} \quad D_Q = \text{diag}(q_{i+})$$

La matrice  $D_Q$  a pour terme diagonal la quantité  $q_{i+}$  qui mesure le nombre de sommets appartenant à  $R_i$ , c'est à dire le nombre d'unités statistiques en relation par  $R$  avec l'unité statistique  $i$ . Par suite  $D_Q$  s'interprète sur le graphe  $G$  comme la matrice des demi-degrés extérieurs (BERGE (1973)) des sommets du graphe.

Le cas d'un graphe complet correspond alors aux identités démontrées plus haut

$$q_{ij}^1 = 1 \quad \forall (i,j), \quad i,j \in I \quad i \neq j$$

$$N_{Q^1} = (nI_n - \mathbf{1}_n \mathbf{1}_n^t) = nQ_0$$

Par suite, puisque  $F \wedge F$  est de dimension  $n(n-1)/2$ , on peut choisir sans perte de généralité la normalisation suivante du graphe simple complet défini sur  $I$ :

$$D_{Q^1} = \frac{2}{n(n-1)} I$$

où  $q = n(n-1)/2$  est le nombre d'arêtes du graphe. Elle conduit à

$$N_{Q^1} = \frac{2}{n-1} Q_0$$

La norme de tout vecteur  $\underline{x}$  de  $F$  dans cette métrique est alors associée à la formule de l'écart-type corrigé de la variable  $X$  puisque la projection de  $\underline{x}$  sur  $F_0$  est la variable centrée de terme général  $(x_i - \bar{x})$  et donc

$$\|\underline{x}\|_{N_{Q^1}}^2 = \frac{2}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (82)$$

En l'absence de risques d'ambiguïté, on note  $N$  la métrique  $N_{Q1}$  dans la suite.

Désignons par  $\bar{G}$  le graphe complémentaire de  $G$  défini sur  $I$ . La donnée de  $\bar{G}$  permet de définir une matrice  $\bar{Q}$  de terme général :

$$\bar{q}_{ij} = \begin{cases} 1 & \text{si et seulement si } q_{ij} = 0 \\ 0 & \text{si et seulement si } q_{ij} = 1 \end{cases} \quad (83)$$

$\bar{G}$  est en effet associé à la relation  $\bar{R}$  complémentaire de la relation  $R$  et vérifie

$$\begin{aligned} I &= D_Q + D_{\bar{Q}} \Leftrightarrow Q_0 = N_Q + (N - N_Q) = N_Q + N_{\bar{Q}} \\ D_{Q1} &= \frac{2}{n(n-1)} (D_Q + D_{\bar{Q}}) \Leftrightarrow N = \frac{2}{n-1} (N_Q + N_{\bar{Q}}) \end{aligned} \quad (84)$$

On obtient ainsi la décomposition de la métrique  $N$  comme somme d'une métrique induite par la relation  $R$  et d'une métrique induite par la relation complémentaire  $\bar{R}$ .

**Remarque :** Lorsque la relation  $R$  est une relation d'équivalence sur  $I$ ,  $Q$  définit un indice de similarité entre unités statistiques.  $\bar{Q}$  s'interprète alors comme un indice de dissimilarité.  $\bar{Q}$  vérifie d'ailleurs les axiomes d'une distance lorsqu'on pose  $q_{ii} = 1$  pour tout  $i$ , c'est à dire lorsque  $R$  est une relation réflexive.

On déduit de (84) la décomposition additive de la variance :

$$\frac{n-1}{2} \text{Var}(X) = \underline{x} N \underline{x} = \underline{x} N_Q \underline{x} + \underline{x} N_{\bar{Q}} \underline{x} \quad (85)$$

qui fournit une généralisation de la décomposition usuelle de la variance totale comme somme de la variance intra-classe et de la variance inter-classe induite par le graphe associé à une relation de partition :

**Proposition 1:** *Pour toute variable  $X$  représentée par le vecteur  $\underline{x} \in F$ , la variance empirique de  $X$  associée à une relation  $R$  se décompose en:*

$$\begin{aligned} \text{Variance de } X &= \text{Variance de } X \text{ sur } R + \text{Variance de } X \text{ hors } R \\ s^2(X) &= s_R^2(X) + s_H^2(X) \end{aligned} \quad (85 \text{ bis})$$

**Définition 4:** *Etant donnée une relation binaire  $R$  définie sur  $I$ , on appelle variance locale de la variable  $X$  la quantité  $s_R^2(X)$ .*

L'information apportée par  $R$  sur les variations de  $X$  est en particulier mesurée par le rapport

$$C(X) = \frac{s_R^2(X)}{s^2(X)} \quad (86)$$

$C(X)$  s'interprète comme le coefficient de contiguïté de GEARY lorsque  $R$  est une relation de contiguïté spatiale (cf. LEBART (1969,1973)). Cet auteur montre que  $C(X)$  mesure l'écart entre deux estimations de la variance de  $X$  lorsque les unités statistiques sont distribuées selon un aléa uniforme sur le graphe  $G$ . Le coefficient de contiguïté dans la direction  $\underline{x}$ ,  $C(\underline{x})$  s'écrit sous forme d'un rapport de Rayleigh

$$C(\underline{x}) = \frac{\|\underline{x}\|_{N_C}^2}{\|\underline{x}\|_N^2} \quad (87)$$

et son optimisation conduit classiquement à la recherche des éléments propres de l'opérateur  $N^{-1}N_Q$  (où  $N^{-1}$  est une inverse généralisée de  $N$ ), c'est à dire des solutions de

$$\begin{aligned} N^{-1}N_Q \underline{x} &= \lambda \underline{x} \\ \|\underline{x}\|_N^2 &= 1 \end{aligned} \quad (88)$$

afin d'analyser la structure du sous-espace de  $F$  défini par les contraintes relationnelles.

Cette façon de procéder peut être comparée à l'analyse spectrale du tableau de données  $Q$ . On est alors conduit à résoudre les équations aux valeurs propres

$$\frac{1}{4} Q^2 F_s = \lambda_s F_s \quad (89)$$

BENZECRI (1973) et GREENACRE (1978) montrent que les équations de transition associées à ce problème s'écrivent (puisque  $Q$  est symétrique)

$$\frac{1}{2} Q F_s = \epsilon_s \sqrt{\lambda_s} F_s \quad (90)$$

On dit (BENZECRI (1973)) que  $F_s$  est un facteur direct (resp. inverse) si  $\epsilon_s = +1$  (resp.  $\epsilon_s = -1$ ). Les facteurs inverses correspondent aux valeurs propres négatives de  $Q$ . GUTTMAN élimine celles-ci à l'aide d'une méthodologie classique d'estimation de la constante additive  $-\lambda^*$ . Une deuxième approche consiste à rajouter  $b$  boucles à chaque sommet du graphe  $G$  de sorte que la matrice  $Q_b = Q + bI$  soit semi-définie positive. Cependant le lemme de COXETER, cité plus haut, offre une approche alternative que nous rapprochons de l'analyse de  $G$  en termes d'Analyse des Correspondances proposée par BENZECRI (1973) et LEBART (1984).

Nous avons vu ci-dessus que  $Q$  est alors interprétée comme la matrice d'une correspondance mathématique que l'on analyse alors de façon classique en termes d'Analyse des Correspondances.

Notons

$$q = \sum_i q_{i.} = \sum_i \sum_j q_{ij}$$

$q$  est égal à deux fois le nombre d'arêtes du graphe. On munit alors  $F$  de la métrique diagonale  $\frac{1}{q}D_q$ , de sorte que les deux premiers moments empiriques de toute variable  $\underline{x} \in F$  s'écrivent

$$\bar{x} = \frac{1}{q} \sum_i q_i x_i \quad s^2(X) = \frac{1}{q} \sum_i q_i (x_i - \bar{x})^2$$

La minimisation de l'indice de Geary associé à la correspondance  $Q$  conduit alors aux équations aux valeurs propres :

$$N_Q \underline{x} = \nu D_q \underline{x} \quad \Leftrightarrow \quad D_q^{-1} Q \underline{x} = (1-\nu) \underline{x} \quad (91)$$

où les vecteurs propres sont  $D_q$ -normés. Par comparaison aux équations de transition (90), écrivons les équations de l'Analyse des Correspondances appliquée à la matrice  $Q$  sous la forme

$$D_q^{-1} Q F_s = \epsilon_s \sqrt{\lambda_s} F_s \quad (92)$$

(92) montre que  $C(\underline{x})$  est minimisé par le vecteur propre direct associé à la plus grande valeur propre issue de cette analyse.

Les équations (90) et (92) fournissent les mêmes vecteurs propres dans le seul cas où  $G$  est un graphe régulier, c'est à dire tel que  $D_Q = r I_n$ . Dans ce cas, chaque sommet est de degré extérieur  $r$ . Les vecteurs propres de  $D_Q^{-1} N_Q$  associés à des valeurs propres non nulles solutions de (91) définissent un sous-espace vectoriel  $F_2$  de  $F$ . La donnée de  $F_2$  permet de traduire les analyses sous contraintes relationnelles dans un cadre linéaire analogue à celui de l'ACPVI présenté au chapitre II.

**Définition 5:** Soit la matrice  $X \in \mathcal{M}_{p,n}(\mathbb{R})$  et  $N_Q$  la (semi-)métrique induite sur  $F$  par une contrainte relationnelle de matrice  $Q$ , on appelle matrice de covariances locales empiriques du tableau  $X$ , la matrice :

$$V_G = \frac{1}{q} X N_Q {}^t X \quad (93)$$

Etant données les variables instrumentales regroupées dans la matrice  $Y$ , la matrice des covariances partielles de  $X$  est définie par:

$$V_p = \frac{1}{n} X(I - {}^tY(YN{}^tY)^{-1}YN)X$$

où l'on munit  $F$  de la métrique  $N = \frac{1}{q} D_Q$  afin de comparer les deux méthodes.  $N$  généralise simplement la matrice  $N_{Q1}$  dans le cas d'un graphe simple pondéré par  $Q_{|x|}$ .

L'approche précédente ne coïncide pas en général avec l'Analyse en Coordonnées Principales hors Variables Instrumentales. Les deux méthodes sont équivalentes lorsque

$$N_Q = \frac{q}{n} (I - {}^tY(YN{}^tY)^{-1}YN)$$

Supposons le graphe  $G$  régulier. Si chaque sommet est adjacent à  $k$  autres sommets du graphe, on a  $D_Q = kI$  et donc la matrice  $N_Q = k(I - \frac{1}{k} Q)$  admet pour décomposition spectrale

$$\frac{1}{k} Q = {}^tF \Lambda F \quad F N {}^tF = I$$

Supposons de plus que  $Q$  admette des valeurs propres non nulles distinctes et posons  $Y = F_{[r]}$  où  $F_{[r]}$  est l'élément de  $\mathcal{M}_{r,n}(\mathbb{R})$  ayant pour lignes les vecteurs propres associés aux  $r$  plus grandes valeurs propres de  $Q$ . Alors

$${}^tY(YN{}^tY)^{-1}YN = \frac{r}{q} {}^tF_{[r]} F_{[r]}$$

Cette expression coïncide avec la décomposition spectrale de  $\frac{1}{k} Q$  uniquement lorsque  $r = \text{rang}(Q)$  et lorsque  $\Lambda = I$ . Ceci n'est possible que si  $G$  est le graphe d'une partition de  $I$  en  $k$  classes, c'est à dire dans le cas de contraintes classiquement traitées en Analyse Discriminante Linéaire. Nous pouvons donc parler dans ce cas d'Analyse en Coordonnées Principales Discriminantes du tableau de dissimilarité  $\Delta$ . Comme l'Analyse en Coordonnées Principales, elle n'est pleinement justifiée que si  $\delta$  est une dissimilarité Euclidienne.

**Remarque :** Une comparaison analogue de l'Analyse en Coordonnées Principales sur Variables Instrumentales et Analyse en Coordonnées Principales sous contraintes de distances peut de plus être menée à l'aide de la matrice  $\bar{Q}$  associée à la relation complémentaire  $\bar{R}$  puisque  $\bar{Q}$  définit une distance sur  $I$ . Cette propriété montre que les

contraintes relationnelles agissent sur les propriétés métriques de la solution. On peut donc parler de façon équivalente de contraintes métriques.

Les approches présentées plus haut permettent en théorie de mettre en œuvre une analyse multidimensionnelle sous contraintes relationnelles. Cependant la mise en œuvre concrète de cette démarche présente quelques difficultés. En particulier, comme dans le cas de la régression sur composantes principales, le fait de ne retenir que  $r$  parmi les  $(n-1)$  dimensions de l'espace engendré par la matrice des contraintes conduit à des résultats erronés lorsque les facteurs associés aux petites valeurs propres sont fortement corrélés avec la (ou les) variable(s) réponse(s).

On peut donc être tenté de retenir une dimension  $r$  égale au rang de la matrice  $N^T N_Q$ . On s'expose alors aux artefacts connus de l'Analyse en Composantes Principales d'un graphe. En effet, si le graphe  $G$  admet un plongement Euclidien dans un espace de dimension  $q$ , le choix optimal de  $r$  est l'entier inconnu  $q$ . Ainsi un graphe planaire nécessite  $q=2$  dimensions. BENZECRI (1973) montre qu'alors les  $n-q$  axes suivants sont en général des fonctions polynomiales des  $q$  axes fondamentaux. Cet auteur montre de plus que cette particularité résulte du fait que les opérateurs soumis à l'analyse décrivent des équations aux différences du second ordre, résultant de la discrétisation d'équations différentielles laplaciennes (cf. GUTTMAN (1954), BENZECRI (1973)). Les vecteurs propres de ces opérateurs de Laplace appartiennent alors à une famille de polynômes orthogonaux (par exemple de Legendre, d'Hermite, ...). On décrit ainsi de façon approchée une variété différentiable engendrée par les contraintes. Cette situation étend donc au cadre non linéaire la situation rencontrée en ACPVI, où les contraintes définissent un sous-espace vectoriel de  $F$ . Les deux méthodes diffèrent cependant, puisque la variété différentiable est connue par ses propriétés métriques induites par les contraintes. Les distances inter-points nous renseignent par exemple sur la courbure de l'espace (cf. d'AUBIGNY (1984)). Au contraire, les contraintes de configuration utilisent dans un cadre algébrique la structure d'espace vectoriel normé sous-jacente de  $F$ .

La surface réponse, solution d'une équation différentielle, n'est décrite que de façon approchée. D'une part, le graphe  $G$  ne fournit qu'une information discrétisée sur les structures métriques qui la caractérisent. D'autre part, le recours à une méthodologie algébrique fournit la dimension linéaire de l'espace vectoriel permettant de la décrire et non sa dimension topologique, c'est à dire la dimension minimale de l'espace des paramètres nécessaires à la description de cette surface. On parle aussi à ce sujet, de dimension intrinsèque (QUINQUETON (1981)).

Bien que de nombreux graphes particuliers aient été étudiés en terme de représentation Euclidienne par BENZECRI (1973), BASTIN et al. (1980), LEBART (1969, 1984), IWATSUBO (1983), l'utilisation de ces travaux à des fins d'analyse d'un tableau de

profil X ou d'un tableau de données de dissimilarité sous contraintes est quasi inexistante et reste un domaine de recherche ouvert.

#### IV.1.1.2. L'Analyse en Coordonnées Principales sous contraintes de graphe

Soit R une relation binaire définie sur I. On note Q la matrice de connexion du graphe G associé à R. Q définit une géométrie de matrice  $N_Q$  sur  $F_0$ , qu'il est naturel de prendre en compte dans l'analyse du tableau de dissimilarité  $\Delta$ .

Tout d'abord, nous étudions le cas particulier caractérisé par les conditions :

- (1)  $\Delta$  est une dissimilarité Euclidienne  $\Delta = \Delta(Z)$ ,
- (2) R est une relation d'équivalence.

On se place dans l'espace de représentation  $F = (\mathbb{R}^n, D_q)$  où la matrice diagonale  $D_q$  est induite par R. On construit alors comme précédemment les matrices :

$$H = -\frac{1}{2} \Delta * \Delta$$

$$W = Q_1 H Q_1 \quad \text{rang}(W) = p$$

où  $Q_1$  est le projecteur  $D_q$ -orthogonal sur  $F_0$ , sous-espace complémentaire  $D_q$ -orthogonal de la droite des constantes dans F. La matrice Z qui factorise  $W_Z$  en  $W_Z = Z'Z$  est alors centrée au barycentre des points représentatifs de I pondérés par les demi-degrés extérieurs des sommets correspondants du graphe de la relation. D'autre part, nous venons de montrer que le sous espace vectoriel de F engendré par les contraintes est décrit par les vecteurs propres  $D_q$ -orthonormés de l'opérateur  $D_q^{-1} Q$ . Ils permettent de construire le projecteur  $P_2$ , et donc de résoudre le problème d'Analyse en Coordonnées Principales sous contraintes dans toute sa généralité. De plus, les axes principaux de l'analyse s'interprètent comme des axes discriminants, de sorte que l'on peut parler d'une méthode d'Analyse en Coordonnées Discriminantes.

Lorsque la condition (2) n'est pas vérifiée, les résultats précédents conduisent à résoudre le problème d'Analyse en Coordonnées Principales Locales, que nous définissons comme la recherche des axes de représentations associés à la factorisation de l'opérateur  $V_G$ . En effet, munissons E de la métrique identité:  $M = I$ . Puisque Z factorise  $W_Z$ , on a:

$$\text{Trace}(V_G) = \frac{1}{q} \text{Trace}(Z N_Q Z) = \frac{1}{q} \text{Trace}(W_Z N_Q)$$

On résoud donc les équations aux valeurs propres:

$$W_Z N_Q \underline{c} = \lambda \underline{c} \quad {}^t \underline{c}^r N_Q \underline{c}^s = 1 \text{ si } r = s, \text{ et } 0 \text{ sinon.}$$

Cette méthode trouve une justification partielle seulement à l'aide du corollaire 2 (Ch. II-1-1). On vérifie en effet le lemme suivant:

**Lemme 2:** *L'opérateur  $D_q^{-1} Q$  est  $D_q$ -symétrique, et de plus*

$${}^t Q_1 N_Q = N_Q$$

**Preuve:** Le premier point résulte de l'égalité

$${}^t (D_q^{-1} Q) D_q = Q$$

D'autre part,

$${}^t Q_1 N_q = {}^t Q_1 D_q (I - D_q^{-1} Q) = D_q Q_1 (I - D_q^{-1} Q)$$

et puisque  ${}^t \underline{1} D_q = \underline{1} Q$ , on vérifie

$$\begin{aligned} Q_1 (I - D_q^{-1} Q) &= (I - \underline{1} {}^t \underline{1} D_q) (I - D_q^{-1} Q) \\ &= I - \underline{1} {}^t \underline{1} D_q - D_q^{-1} Q + \underline{1} {}^t \underline{1} Q = I - D_q^{-1} Q \quad \parallel \end{aligned}$$

Ces propriétés ne permettent cependant pas d'interpréter ce problème comme une ACPVI particulière puisqu'en général, les opérateurs  $D_q^{-1} Q$  et  $N_Q$  ne sont pas idempotents. La solution vérifie cependant la proposition:

**Proposition 2:** *Si la configuration  $Z$  est centrée par rapport à la pondération  $D_q$ , la solution  $X$  réalise le minimum de la fonction*

$$E(Z, Y) = \text{Trace}((W_Z - W_Y) N_Q (W_Z - W_Y) N_Q)$$

*parmi les configurations  $Y$  de rang  $r$ . De plus, à l'optimum:*

$$E(Z, X) = \frac{1}{4} \| \underline{\delta}^2(Z) - \underline{d}^2(X) \|_{N_Q \circ N_Q}^2 \quad (98)$$

**Preuve:** Le premier point résulte des propriétés d'optimalité classiques en Analyse en Composantes Principales, puisqu'ici seule la métrique change. Le point 2 résulte alors du lemme 2  $\parallel$

Dans  $F \wedge F$  ce problème se traduit, à l'aide de l'isométrie  $T$ , par la relation

$$\text{Trace}(W N_Q) = \text{Trace}(W T D_Q {}^t T) = \text{Trace}({}^t T W T D_Q) = \text{Trace}({}^t Y D_Q)$$

c'est à dire

$$\text{Trace}(W_N Q) = \sum_{i,l} q_{il} \delta_{il}^2 = \sum_{(i,l) \in U} \delta_{il}^2 \quad (98 \text{ bis})$$

Par suite, l'approximation interne de  $W_Z N_Q$  par un opérateur de rang  $r$   $W_X N_Q$ , vérifie

$$\begin{aligned} G(Z,X) &= \text{Trace}((W_Z - W_X)N_Q) \\ &= \sum_{i < j} \sum q_{ij} (\delta_{ij}^2(Z) - d_{ij}^2(X)) \\ &= \sum_{(i,j) \in U} \delta_{ij}^2(Z) - \sum_{(i,j) \in U} d_{ij}^2(X) \end{aligned}$$

Le tableau  $X$  ayant pour lignes les vecteurs propres associés aux  $r$  plus grandes valeurs propres de  $W_Z N_Q$  ( $r \leq \text{rang}(W_Z N_Q)$ ) maximise donc en moyenne les carrés des distances inter-points le long du graphe  $G$ . De plus la contrainte d'orthonormalisation des vecteurs propres par rapport à la métrique  $N_Q$  se traduit dans  $F \wedge F$  par la relation

$${}^t c^r N_Q c^s = \sum_{i < j} q_{ij} (c_i^r - c_i^s) (c_j^r - c_j^s) = \sum_{(i,j) \in U} (c_i^r - c_i^s) (c_j^r - c_j^s)$$

qui montre que l'on impose aux vecteurs propres des contraintes de non corrélation et de variance unité le long du graphe. Dans le cas de l'Analyse en Coordonnées Discriminantes, on retrouve les contraintes portant sur les variances et corrélations intra-classes.

Cette analyse repose sur le caractère Euclidien de la dissimilarité  $\Delta$ . Lorsque  $\Delta$  est quelconque,  $W$  n'est pas nécessairement gramienne. L'Analyse en Coordonnées Principales sous contraintes de graphe présente donc les mêmes difficultés de justification théorique et de mise en œuvre sur des données réelles que l'Analyse en Coordonnées Principales sans contraintes.

**Remarque 1:** L'équation (98 bis) est à la base de l'extension de la méthodologie de l'Analyse de la Variance au traitement de variables réponse mesurées sous forme de données de dissimilarité (HUBERT (1982), d'AUBIGNY ((1988)).

**Remarque 2 :** Les calculs qui précèdent supposent que  $M$  est la métrique identité. Leur extension au cas d'une métrique quelconque est cependant immédiat et conduit à factoriser l'opérateur  $V_G M$ . On construit alors une factorisation de la matrice de produits scalaires associée à  $\Delta(Z)$  de la forme

$$W_Z = {}^t Z M Z$$

Cette possibilité permet en particulier de traiter le cas de contraintes mixtes, caractérisées par la donnée d'une relation  $R$  et d'un tableau de mesures externes  $Y_2$ . On choisit alors pour métrique  $M$  la matrice  $\widehat{M}_2$ , définie au ch. II-1-1.

#### IV.1.2. EXTENSION AU CAS DE RELATIONS VALUÉES

L'étude des contraintes de graphe s'étend de façon naturelle au cas de contraintes décrites par la donnée de graphes valués. Chaque arête  $u_{ij}$  du graphe  $G$  est alors munie d'une valeur que l'on interprète selon le contexte d'application, soit comme une mesure de similarité ou de ressemblance  $s_{ij}$ , soit de façon duale comme une mesure de dissimilarité ou de dissemblance  $\gamma_{ij}$ .

Nous étudions d'abord le cas particulier d'une relation d'équivalence floue qui généralise de façon directe l'Analyse en Coordonnées Discriminantes associées à des contraintes de partition.

Une relation  $R$  définie sur l'ensemble  $I$  et à valeur dans l'ensemble  $J$  ou de façon équivalente une relation définie sur  $I \times J$  est définie par la donnée d'un sous-ensemble de  $I \times J$ . Elle est caractérisée par une fonction d'appartenance à la relation qui associe à toute paire  $(i,j) \in I \times J$  son intensité d'appartenance à la relation  $R$   $\mu_R(i,j)$ . Lorsque  $\mu_R$  est à valeurs dans  $\{0,1\}$  on parle de relation nette. Dans le cas contraire, on parle de relation floue. Sans perte de généralité nous supposons dans la suite que  $\mu_R$  est à valeurs dans  $[0,1]$ . On note respectivement par  $\wedge$  et  $\vee$  les opérations de minimum et de maximum.

**Définitions 6 (ZADEH (1971)) :** *Le domaine de la relation  $R$  est le sous-ensemble flou  $\text{dom}(R)$  de  $I$  défini par*

$$\mu_{\text{dom}(R)}(i) = \vee \{ \mu_R(i,j) ; j \in J \} \quad \forall i \in I$$

*L'image de la relation  $R$  est le sous ensemble flou  $\text{ran}(R)$  de  $J$  défini par*

$$\mu_{\text{ran}(R)}(j) = \vee \{ \mu_R(i,j) ; i \in I \} \quad \forall j \in J$$

*Le support de la relation floue  $R$  est le sous ensemble net de  $I \times J$  sur lequel  $\mu_R(i,j) > 0$ .*

On munit alors l'ensemble des relations définies sur  $I \times J$  des opérations d'union et d'intersection définies respectivement par

$$R \cup S : \mu_{R \cup S} = \mu_R \vee \mu_S$$

$$R \cap S : \mu_{R \cap S} = \mu_R \wedge \mu_S$$

La relation complémentaire de  $R$  est la relation  $\bar{R}$  définie par

$$\mu_{\bar{R}} = 1 - \mu_R$$

De plus la composition de deux relations  $R \subset I \times J$  et  $S \subset J \times K$  est la relation  $RoS$  définie par

$$\mu_{RoS}(i,k) = \vee \{ \mu_R(i,j) \wedge \mu_S(j,k) ; j \in J \} \quad \forall (i,k), (i,k) \in I \times K$$

**Définition 7:** Soit  $\alpha \in [0,1]$ . On appelle composante de niveau  $\alpha$  de la relation  $R$ , la relation nette  $R_\alpha$  définie par

$$R_\alpha = \{ (i,j) \in I \times J ; \mu_R(i,j) \geq \alpha \}$$

La famille  $\{R_\alpha, \alpha \in [0,1]\}$  définit une suite emboîtée de relations nettes :

$$\alpha_1 \geq \alpha_2 \quad \Rightarrow \quad R_{\alpha_1} \subset R_{\alpha_2}$$

Par suite on construit la décomposition de  $R$  suivante :

**Proposition 3 (ZADEH (1971) page 181) :** Toute relation floue définie sur  $I \cup J$  admet la décomposition

$$R = \bigcup \{ \alpha R_\alpha ; 0 < \alpha \leq 1 \}$$

où  $\alpha R_\alpha$  est l'ensemble net défini par

$$\mu_{\alpha R_\alpha} = \begin{cases} \alpha & \text{si } (i,j) \in R_\alpha \\ 0 & \text{sinon} \end{cases}$$

ZADEH (1971) montre qu'à chaque ensemble  $R_\alpha$  peut être associé un  $\alpha$  unique. De plus, lorsque  $I$  et  $J$  sont des ensembles finis, il existe un entier  $K$  tel que  $R$  s'écrit

$$R = \bigcup \{ \alpha_k R_{\alpha_k} ; k = 1, \dots, K \}$$

**Définition 8:** On appelle similarité floue toute relation  $S$  floue définie sur  $I \times I$  telle que

(a)  $S$  est réflexive :  $\mu_S(i,i) = 1 \quad \forall i \in \text{dom}(S)$

(b)  $S$  est symétrique :  $\mu_S(i,j) = \mu_S(j,i) \quad \forall i,j \in \text{dom}(S)$

(c)  $S$  est transitive : SoS c'est à dire  $\mu_S(i,j) \geq \vee \{ \mu_S(i,k) \wedge \mu_S(k,j) ; k \in I \}$

L'intérêt des relations de similarité floue réside dans les propriétés déduites de la condition de transitivité. Considérons en effet la relation floue complémentaire d'une relation de similarité floue.

**Proposition 4:** *La relation floue complémentaire d'une relation de similarité floue définit une dissimilarité ultramétrique.*

**Preuve :** Notons  $\gamma_{ij}$  la quantité  $\mu_{\bar{S}}(i,j) = 1 - \mu_S(i,j) \forall i,j \in I \times I$ .  $\gamma$  s'interprète comme une mesure de dissimilarité ultramétrique puisque la condition de transitivité (c) implique

$$1 - \gamma_{ik} \geq \vee \{ (1 - \gamma_{ij}) \wedge (1 - \gamma_{jk}) ; j \in I \}$$

mais puisque

$$(1 - \gamma_{ij}) \wedge (1 - \gamma_{jk}) = 1 - (\gamma_{ij} \vee \gamma_{jk}) \quad \forall i,j,k \in I$$

on obtient l'inégalité ultramétrique

$$\gamma_{ik} \leq \gamma_{ij} \vee \gamma_{jk} \quad \forall j \in I \quad \parallel$$

Par suite, la similarité floue  $S$  définit une contrainte d'appartenance à une hiérarchie indicée et fournit donc une généralisation de l'analyse en coordonnées discriminantes puisque  $S$  définit une hiérarchie de partitions emboîtées et non plus une partition nette.

De plus le théorème de HOLMAN (1972) (cf. aussi LINGOES (1971), d'AUBIGNY (1975)) assure que  $\Gamma$  est alors une dissimilarité Euclidienne nécessitant  $(n-1)$  dimensions pour être représentée. L'Analyse en Coordonnées Principales d'un tableau  $\Delta$  sous les contraintes définies par  $S$  fournit alors une méthode de représentation des dissimilarités  $\Delta = (\delta_{ij})$  de sorte à respecter les contraintes typologiques définies par la dissimilarité ultramétrique  $\Gamma$ . Il suffit pour cela de mener l'analyse dans l'espace vectoriel muni de la (semi-)métrique  $N_Q$  associée à la matrice de connexion "valuée"  $Q$  définie par

$$q_{ij} = \mu_S(i,j) \quad \forall (i,j), (i,j) \in I \times I$$

On peut de plus étudier la stabilité des représentations Euclidiennes obtenues en menant l'Analyse en Coordonnées Principales de  $\Delta$  sous les contraintes définies par les sections de l'arbre de classification correspondant aux partitions de  $I$  définies par les relations nettes  $R_\alpha$  intervenant dans la décomposition de  $R$ . Enfin le plongement Euclidien de  $\Delta$  sous la contrainte associée à la valuation  $\gamma = \mu_{\bar{R}}$  permet d'analyser l'information apportée par  $\Delta$  complémentaire à la structure ultramétrique décrite par  $q = \mu_R$ .

Dans le cas général, on dispose d'une relation floue valuée associée à une matrice de dissimilarité  $Q$ , a priori quelconque, et donc en général non Euclidienne, que l'on prend en compte dans l'analyse du tableau observé  $\Delta$  en fixant sur  $F_0$  une métrique induite par la matrice  $N_Q$ . Par analogie au cas des contraintes de graphe  $G$ , on résoud l'équation aux valeurs propres :

$$WN_{Qx} = \lambda x \quad {}^t x N_{Qx} = 1 \quad (99)$$

Remarquons cependant que la complémentarité entre similarités et dissimilarités est alors difficile à décrire à l'aide d'une formule générale (cf. FICHET (1985), GOWER (1985), LE CALVE (1985)).

Même si la normalisation à 1 de S est naturelle et ne limite en rien la généralité de la méthode, le choix  $q_{ij} = 1 - s_{ij}$  ne s'impose en effet nullement si S n'est pas une relation de similarité floue, puisqu'il ne s'appuie sur aucun résultat général d'optimalité. En pratique, le choix d'une transformation du type

$$s_{ij} = \tau(q_{ij})$$

répond à des considérations non mathématiques bien qu'il influe sur les résultats de l'analyse lorsque la classe des transformations retenues ne respecte pas les propriétés d'invariance de la méthode de plongement Euclidien retenue (cf. BENINEL (1987)). Nous n'aborderons pas ici ce problème et proposons d'appliquer l'analyse spectrale de  $WN_Q$ , décrite pour une relation nette, au cas d'une dissimilarité générale Q. Le lemme de COXETER (cf. supra) assure en effet que  $N_Q$  définit une forme quadratique sur  $F_0$  pour toute contrainte qui s'exprime sous la forme d'une matrice de dissimilarité. A l'opposé, la présentation de l'ACPVI en termes d'opérateurs, décrite en III.1.2 n'est pas applicable dans le cas de contraintes générales. dans ce cas en effet, la relation évaluée R définit une dissimilarité  $\Delta_R$  qui n'est pas Euclidienne en général,  $\Gamma \neq \Gamma(Y)$  de sorte que la matrice  $W_R$  associée à  $\Gamma$  par la transformation

$$W_R = Q_1 H_R Q_1 \\ H_R = -\frac{1}{2} \Gamma * \Gamma$$

n'est pas gramienne pour une relation R quelconque. Par suite la difficulté du problème est liée au calcul du projecteur  $P_R$ , noté  $P_2$  dans l'équation (62)

On a en effet

$$P_2 = W_2 N = {}^t Y_2 V_2 Y_2 \\ V_2 = Y_2 N {}^t Y_2$$

La donnée d'une relation floue définie par une mesure de comptage sur  $I \times I$  requiert une mention spéciale. On peut l'interpréter soit en termes de graphe quelconque (c'est à dire non simple) de sommets I (BERGE (1973)), soit comme une correspondance symétrique définie sur I. Cette situation se rencontre en particulier lors de l'étude d'une relation de confusion entre unités statistiques. Dans ce cas, la normalisation de Q par la somme des

pondérations  $q$  permet d'interpréter  $1/q Q$  comme une matrice de fréquences relatives et  $D_q$  comme la matrice diagonale des marges de  $1/qQ$ . Par suite les matrices

$$N_Q = \frac{1}{q} (D_q - Q) \text{ et } D_q^{-1} N_Q = \frac{1}{q} (I - D_q^{-1} Q)$$

admettent pour vecteurs propres les solutions de l'équation

$$\frac{1}{q} D_q^{-1} Q \underline{E}^s = \lambda_s \underline{E}^s$$

c'est à dire les facteurs de l'Analyse des Correspondances appliquée au tableau de contingence  $1/qQ$ . Allons plus loin et remarquons qu'alors la relation floue complémentaire  $\bar{R}$  s'interprète aussi en termes de fréquence et

$$\bar{q}_{ij} = 1 - \frac{1}{q} q_{ij}$$

Supposons alors qu'il existe une partition de  $I$  en  $J \cup K$ , de telle sorte que

$$q_{ii'} = 0 \text{ si } (i \in J \text{ et } i' \in K)$$

$Q_{I \times I}$  définit deux composantes connexes sur le graphe  $G$  de la relation  $R$ , et par suite la décomposition spectrale de  $W_Z N_Q$  généralise la problématique de l'Analyse en Coordonnées Discriminantes.

De façon complémentaire, l'analyse de  $\Delta(Z)$  sous la contrainte  $\bar{R}$  est intéressante lorsque l'on suppose que

$$q_{ii'} = 1 \text{ si } (i \in J \text{ et } i' \in K) \quad (\Leftrightarrow \bar{q}_{ii'} = 0 \text{ si } (i, i' \in J \text{ ou } i, i' \in K))$$

puisqu'alors,  $\bar{Q}$  définit une correspondance entre  $J$  et  $K$ , et

$$D_{\bar{q}} = \begin{bmatrix} D_J & 0 \\ 0 & D_K \end{bmatrix}, \quad \bar{Q} = \begin{bmatrix} 0 & \bar{Q}_{JK} \\ \bar{Q}_{JK} & 0 \end{bmatrix}$$

où  $D_J$  est la matrice diagonale associée à la marge  $J$  du tableau de contingence  $\bar{Q}_{JK}$ .  $D_K$  se définit de façon analogue à partir la marge  $K$  du tableau. On reconnaît alors en l'analyse spectrale de l'opérateur  $D_{\bar{q}}^{-1} \bar{Q}$  les équations de l'Analyse des Correspondances binaires dans le cadre d'une  $\Sigma$ -stratégie. Lorsque l'on étend cette démarche au cas d'une partition

de I en p classes, on retrouve de la même façon les équations de l'Analyse des Correspondances Multiples.

## IV.2. CODAGE MULTIDIMENSIONNEL

Lorsque l'application des méthodes d'Analyse en Coordonnées Principales sous contraintes est rendue impossible soit par le caractère non Euclidien des dissimilarités observées  $\Delta$ , soit par les particularités propres à la matrice de dissimilarité  $\Gamma$  définissant les contraintes, on traite le problème en termes d'approximation du tableau  $\Delta$  par un tableau de distance Euclidienne. Cette approche s'applique aux méthodes de Codage Multidimensionnel fondées sur l'optimisation de la fonction STRESS, de la fonction SSTRESS ou plus généralement lorsqu'on optimise une fonction critère de la forme

$$\psi(\alpha, 2, d(X)) = \sum_i^n \sum_j^n q_{ij} (\delta_{ij}^\alpha - d_{ij}^\alpha(X))^2$$

A titre d'exemple, notre exposé est limité au cas  $\alpha = 1$  (cas du STRESS).

### IV.2.1. CAS D'UNE CONTRAINTTE DÉFINIE PAR UNE MATRICE DE DISSIMILARITÉ

Soient  $\Delta$  et  $\Gamma$  les matrices de dissimilarité définissant respectivement les observations et les contraintes relationnelles. La matrice  $D(X)$  des distances associées à la solution  $X$  doit minimiser simultanément les fonctions objectif

$$L_o = \sum_{i < j}^n \sum_{i < j}^n q_{ij} (\delta_{ij} - d_{ij}(X))^2 = \|\delta - \underline{d}(X)\|_{D_Q}^2$$

$$L_R = \sum_{i < j}^n \sum_{i < j}^n q_{ij} (\gamma_{ij} - d_{ij}(X))^2 = \|\gamma - \underline{d}(X)\|_{D_Q}^2$$

BORG et LINGOES (1980) proposent de résoudre ce problème en utilisant une méthode numérique de pénalisation. On minimise alors la fonction

$$L(X) = \lambda_1 L_o + \lambda_2 L_R \quad \text{où } \lambda_1 + \lambda_2 = 1 \quad \lambda_1, \lambda_2 \in [0, 1]$$

à l'aide d'un algorithme de gradient. Là encore, les problèmes de différentiabilité de  $L(X)$  en fonction de  $X$  aux points  $X$  pour lesquels  $d_{ij}(X) = 0$  pour un couple  $(i, j)$  au moins, posent des difficultés algorithmiques ignorées par ces auteurs. Cependant, la structure additive de  $L(X)$  assure que l'extension de l'algorithme de sous-gradient de de LEEUW et

HEISER (1977) à l'optimisation de  $L(X)$  ne pose pas de problèmes particuliers de mise en œuvre. Les inconnues du problème traité par BORG et LINGOES (1980) sont les éléments de la matrice  $X$ . On suppose alors la dimension  $r$  de l'espace de représentation fixée et les auteurs ne fournissent pas de méthode optimale de détermination des coefficients  $\lambda_1$  et  $\lambda_2$ . L'itération proposée consiste en effet à indexer ces coefficients en fonction de l'itération  $v$ . Ils construisent ainsi une suite  $\lambda_2^{(v)}$  croissant vers 1, afin de donner une importance croissante au respect des contraintes au cours du développement de l'itération. La stratégie adoptée par BORG et LINGOES privilégie donc le respect des contraintes au détriment de l'ajustement des observations. Une telle logique peut clairement être inversée en permutant le rôle des coefficients  $\lambda$ . Nous préférons développer une méthode qui détermine la qualité relative de l'ajustement de  $\Delta$  et  $\Gamma$  conjointement à la construction de la solution  $X$ .

Nous nous proposons de plus de formuler le problème dans un cadre plus général. Nous considérons en effet le cas de contraintes définies par deux cônes convexes polyédriques fermés  $K_1$  et  $K_2$ . Nous notons alors  $\underline{\delta}_1 = \underline{\delta}$  et  $\underline{\delta}_2 = \underline{\gamma}$  et nous supposons que  $\underline{\delta}_1 \in K_1$  et  $\underline{\delta}_2 \in K_2$ .

On étend ainsi l'approche de BORG et LINGOES au cas du Codage Multidimensionnel Ordinal ( $K_1$  décrit l'ensemble des transformation monotones de  $\underline{\delta}$ ), sous contraintes ordinales ( $K_2$  décrit l'ensemble des transformation monotones de  $\underline{\gamma}$ ). Le cas de dissimilarités observées Euclidiennes (resp. de contraintes linéaires) réduit alors  $K_1$  (resp.  $K_2$ ) à une droite ou une demi-droite dans  $\mathbb{R}^m$ ,  $m = n(n-1)/2$ . Ces deux cas particuliers montrent que cette formulation du problème recouvre un grand nombre de situations concrètes.

**Lemme 3:**

$$L(X) = \sum_{j=1}^2 \lambda_j \|\underline{\delta}_j - \underline{d}(X)\|_{D_Q}^2 = \sum_{j=1}^2 \lambda_j \|\underline{\delta}_j\|_{D_Q}^2 - \sum_{j=1}^2 \lambda_j \|\underline{\delta}_j\|_{D_Q}^2 + \sum_{j=1}^2 \lambda_j \|\underline{\delta}_j - \underline{d}(X)\|_{D_Q}^2 \quad (101)$$

**Preuve :** Cette identité résulte de façon directe du développement de la fonction

$$\text{STRESS} = \|\underline{\delta} - \underline{d}(X)\|_{D_Q}^2 = \|\underline{\delta}\|_{D_Q}^2 + \|\underline{d}(X)\|_{D_Q}^2 - 2 \langle \underline{\delta}, \underline{d}(X) \rangle_{D_Q}$$

et de la contrainte

$$\lambda_1 + \lambda_2 = 1 \quad \square$$

Nous nous proposons ici de déterminer  $\underline{d}(X)$  comme solution du problème

$$(P) \text{ Min } \left\{ \left\| \sum_{j=1}^2 \lambda_j \underline{\delta}_j - \underline{d}(X) \right\|_{D_Q}^2; \lambda_j \in [0,1], \lambda_1 + \lambda_2 = 1, \underline{\delta}_j \in K_j, \underline{d}(X) \in \mathfrak{D}(X) \right\} \quad (102)$$

où  $\mathfrak{D}(X)$  est l'ensemble des matrices de distance Euclidienne d'ordre  $n$  associées à une représentation de  $I$  dans un espace vectoriel de dimension  $r$  fixée a priori.

La résolution de (P) pose un problème distinct de la minimisation de  $L(X)$  proposée par BORG et LINGOES (1980). La détermination des coefficients  $\lambda_1$  et  $\lambda_2$  par l'analyse permet alors de juger du poids relatif des données et des contraintes dans la construction de la représentation  $X$ . D'autre part, lorsque nous supposons  $\underline{d}(X)$  fixée, on est amené à résoudre le problème de Régression Multiple Optimale (P1) défini par

$$(P1) \text{ Min } \left\{ \left\| \sum_{j=1}^2 \lambda_j \underline{\delta}_j - \underline{d}(X) \right\|_{D_Q}^2, \underline{\delta}_j \in K_j, j = 1, 2 \right\} \quad (103)$$

que l'on résoud de façon désormais classique à l'aide d'un algorithme de moindres carrés alternés (YOUNG, de LEEUW et TAKANE (1976)). On peut donc imaginer de résoudre le problème (P) en alternant deux phases d'optimisation :

(1) Pour  $\underline{d}(X)$  fixé, résoudre le problème (P1)

(2) Pour  $\underline{\delta} = \sum_{j=1}^2 \lambda_j \underline{\delta}_j$  fixé, résoudre le problème (P2)

$$(P2) \text{ Min } \left\{ \left\| \underline{\delta} - \underline{d}(X) \right\|_{D_Q}^2; \underline{d}(X) \in \mathfrak{D}(X) \right\} \quad (104)$$

On détermine ainsi sous la contrainte technique  $\lambda_1 + \lambda_2 = 1$ , la valeur optimale des paramètres  $\lambda_j$  que l'on peut ensuite utiliser dans la formulation de BORG et LINGOES.

Nous étudions dans le paragraphe suivant les méthodes de résolution du problème de façon plus détaillée, dans le cadre de l'Analyse Canonique. Remarquons que les méthodes usuelles utilisées dans un cadre linéaire conduisent alors à prendre  $\lambda_1 = \lambda_2 = \frac{1}{2}$ .

#### IV.2.1.1. Analyse Canonique de deux convexes fermés

Nous montrons tout d'abord que le problème (P1) consiste soit à rechercher l'intersection de deux cônes convexes fermés  $K_1$  et  $K_2$  de  $\mathbb{R}^m$  dans le cas d'une intersection non vide soit, dans le cas contraire, à calculer la distance entre ces deux cônes. Considérons en effet le problème suivant :

(P') trouver  $(x_1^*, x_2^*) \in K_1 \times K_2$  tel que

$$\frac{1}{2} \|x_1^* - x_2^*\|_{D_Q}^2 = \text{Min} \left\{ \frac{1}{2} \|x_1 - x_2\|_{D_Q}^2 ; (x_1, x_2) \in K_1 \times K_2 \right\} \quad (104)$$

Nous disposons alors de la définition classique de la distance entre deux convexes comme la distance minimum séparant deux points appartenant respectivement à chaque convexe:

$$d(K_1, K_2) = \text{Min} \{ d(x_1, x_2) ; x_j \in K_j, j = 1, 2 \}$$

Cette définition conduit à noter l'ensemble des solutions du problème (P1) par

$$\mathcal{Y} = \{ (x_1^*, x_2^*) \in K_1 \times K_2 ; d(x_1^*, x_2^*) = d(K_1, K_2) \}$$

qui s'écrit dans le cas  $K_1 \cap K_2 \neq \emptyset$  :

$$\mathcal{Y} = \{ (x_1^*, x_2^*) \in \mathbb{R}^m \times \mathbb{R}^m ; x^* \in K_1 \cap K_2 \}$$

**Proposition 5 :** (PIERRA (1976)): *Les solutions du problème (P') vérifient les trois propriétés équivalentes :*

- (1)  $(x_1^*, x_2^*) \in K_1 \times K_2 ; d(x_1^*, x_2^*) = d(K_1, K_2)$   
où  $d(.,.)$  est la distance Euclidienne associée à la norme définie par  $D_Q$ .
- (2)  $(x_1^*, x_2^*) \in K_1 \times K_2 ; d(x_1^*, x_2^*) = d(K_1, x_2^*) = d(x_1^*, K_2)$
- (3)  $x_1^* = P_{K_1}(x_2^*)$  et  $x_2^* = P_{K_2}(x_1^*)$

La preuve est donnée par PIERRA (1976). Elle est immédiate en remarquant que  $P_{K_j}$  est le projecteur  $D_Q$ -orthogonal sur le convexe fermé  $K_j$ .

L'interprétation statistique du problème (P') est immédiate lorsque  $K_1$  et  $K_2$  sont deux cônes convexes polyédriques fermés inclus dans  $F$ . Le problème (P') décrit alors l'Analyse Canonique de deux cônes polyédriques fermés et généralise donc l'Analyse Canonique de deux sous espaces vectoriels fermés (DAUXOIS et POUSSE (1976)).

**Remarque :** Du point de vue probabiliste, les cônes de contraintes  $K_j$  sont le plus couramment associés à des variables ordinales  $X^j$ , définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ . d'AUBIGNY (1975) montre que lorsque  $X^j$  est supposée de carré intégrable, le  $\sigma$ -treillis  $\mathcal{B}_j$  inclus dans  $\mathcal{A}$ , induit par  $X^j$  définit dans  $L_{\mathbb{R}}^2(\Omega, \mathcal{A}, P)$  un cône polyédrique convexe fermé  $K_j$ . Ainsi donc, lorsqu'on travaille dans  $F = L_{\mathbb{R}}^2(\Omega, \mathcal{A}, P)$ , la présente formulation, qui n'impose pas de contraintes de dimension, offre une méthodologie algorithmique d'Analyse Canonique de deux  $\sigma$ -treillis. On généralise ainsi

L'Analyse Canonique de deux tribus décrite par DAUXOIS et POUSSE (1976) au cas d'ensembles de parties de  $\Omega$  non clos par passage au complémentaire.

L'Analyse Canonique de deux cônes convexes polyédriques a fait l'objet de nombreuses publications en dimension finie, du fait de son intérêt propre en analyse des données (cf. par exemple VAN DER BURG et de LEEUW (1983), TENENHAUS (1987) et VEDERGALL (1986)). Elle permet en particulier de définir des indices empiriques de liaison entre variables ordinales en termes de sup-correlation, dont la contrepartie théorique est discutée par d'AUBIGNY (1979).

La proposition 5 a pour conséquence immédiate le corollaire suivant

**Corollaire :** *Le problème (P') est équivalent aux deux problèmes suivants :*

i) trouver  $x_1^* \in K_1$  tel que

$$(P3) \quad \frac{1}{2} \|x_1^* - P_{K_2}(x_1^*)\|_{D_Q}^2 = \text{Min} \left\{ \frac{1}{2} \|x - P_{K_2}(x)\|_{D_Q}^2 ; x \in K_1 \right\}$$

ii) trouver  $x_2^* \in K_2$  tel que

$$(P4) \quad \frac{1}{2} \|x_2^* - P_{K_1}(x_2^*)\|_{D_Q}^2 = \text{Min} \left\{ \frac{1}{2} \|y - P_{K_1}(y)\|_{D_Q}^2 ; y \in K_2 \right\}$$

La résolution du problème (P3) peut utiliser classiquement un algorithme de gradient projeté :

(0) initialisation :  $k = 0 : x^0 \in K_1$

(1) itération  $k$ . On est en  $x^k$  :

calculer le gradient  $\nabla f(x^k)$  où  $f(x) = \frac{1}{2} \|x - P_{K_2}(x)\|_{D_Q}^2$

calculer  $y^k = P_{K_1}(x^k - \nabla f(x^k))$

définir le déplacement  $d^k = y^k - x^k$

(2) test de convergence

si  $d^k \neq 0$  poser  $x^{k+1} = x^k + \alpha_k d^k \quad \alpha_k > 0 \quad \forall k, k > 0$

faire  $k = k + 1$  et aller en (1)

sinon arrêt.

Remarquons que  $\nabla f(x) = (x - P_{K_2}(x))$  et donc l'itération s'écrit

$$x^{k+1} = x^k + \alpha_k [P_{K_1}(P_{K_2}(x^k)) - x^k] \quad (105)$$

on a donc l'optimum

$$x^* = \alpha_* P_{K_1} \circ P_{K_2}(x^*) \quad (106)$$

Cette expression classique fournit la solution du problème d'Analyse Canonique Linéaire, dans le cas où les  $K_j$  ( $j=1,2$ ) sont des sous-espaces vectoriels engendrés respectivement par deux tableaux de données  $X^j$ . Nous avons ainsi généralisé la formulation usuelle des équations de l'Analyse Canonique Linéaire de deux ensembles de variables quantitatives au sens de la  $\Pi$ -stratégie (cf. CAILLEZ et PAGES (1976)) au cadre de l'analyse de variables statistiques définissant des sous-ensembles des convexes fermés de  $F$ . Dans le cadre linéaire usuel,  $K_j$  est un sous-espace vectoriel de l'espace Euclidien  $F = \mathbb{R}^m$  muni de la métrique  $D_Q$  pour  $j=1,2$ . L'algorithme du gradient projeté se réduit alors à la recherche des éléments propres de l'opérateur  $P_{K_1} \circ P_{K_2}$  à l'aide d'un algorithme de puissance itérée.

BOUGHAZI (1987) montre que l'ensemble des solutions  $\mathcal{Y}$  est convexe, fermé et non vide si l'un des  $K_j$  est fermé. Il montre de plus le

**Théorème 2** (BOUGHAZI (1987) page 127) :

- Si  $K_1 \cap K_2 \neq \emptyset$ , les deux suites  $(x^k)$  et  $(y^k)$  définies par

$$y^k = P_{K_2}(x^{k-1}) \quad (107)$$

$$x^k = P_{K_1}(y^k) \quad (108)$$

convergent vers  $x^* = y^* \in K_1 \cap K_2$

- si  $K_1 \cap K_2 = \emptyset$  et l'un des  $K_j$  est borné, la suite  $(x^k, y^k)$  converge vers l'élément  $(x^*, y^*)$  de  $K_1 \times K_2$  et

$$\lim_{k \rightarrow \infty} \|x^k - y^k\|_{D_Q}^2 = d(K_1, K_2) \quad (109)$$

Notons enfin que lorsque l'un des deux cônes est une variété linéaire, on retrouve un problème classique d'analyse numérique pour lequel le choix du paramètre de relaxation  $\alpha_k$  définissant l'itération (105) peut être fixé de façon optimale (cf. BOUGHAZI (1987)), afin d'assurer une vitesse de convergence satisfaisante de l'algorithme.

La solution du problème (P') fournit un indice de compatibilité entre les données et les contraintes dans un cadre autorisant les transformations des données et des contraintes, les transformations admissibles étant déterminées respectivement par les cônes  $K_1$  et  $K_2$ . On mesure alors un écart. Par contraste, le problème (P3) (resp. (P4)) mesure la perte d'information sur les données observées (resp. les contraintes) encourue lors de la prise

en compte des contraintes (resp. des dissimilarités observées). On retrouve ainsi dans un cadre formel plus général les outils développés pour interpréter les résultats d'une Analyse Canonique Linéaire.

De plus, puisque les solutions  $x^*$  et  $y^*$  se correspondent par projection lorsque les cônes  $K_1$  et  $K_2$  sont disjoints, ils appartiennent alors à une face de leur cône respectif, et donc à leur sous-espace vectoriel support. La solution peut donc être interprétée a posteriori comme la solution d'un problème d'Analyse Canonique de ces supports. Le dépouillement des résultats de l'analyse est par suite enrichi par la consultation des diagnostics usuels utilisés en Analyse Canonique Linéaire.

#### IV.2.1.2. $\Sigma$ -stratégie de résolution du problème d'Analyse Canonique de deux convexes fermés

##### IV.2.1.2.1 Formulation du problème

La formulation précédente montre les liens existant entre une optique d'Analyse Canonique de deux convexes fermés et l'étude des relations entre données de dissimilarité observées sous des contraintes relationnelles décrites par un tableau de dissimilarité. Le problème (P') se place dans une logique d'analyse Canonique et non à proprement parler d'ACPVI. Le rôle symétrique joué ici par les données et les contraintes permet d'autre part d'exprimer le problème d'optimisation sous une forme adaptée au développement d'algorithmes plus satisfaisants que les méthodes de gradient, compte tenu des problèmes de différentiabilité de la fonction critère indiqués plus haut. Nous reformulons ici le problème dans le cadre d'une  $\Sigma$ -stratégie (cf. CAILLEZ et PAGES (1976)) afin de proposer un algorithme résolutif.

Posons  $\mathcal{C} = K_1 \times K_2$  et  $\mathcal{D} = \{ (x, y) \in \mathbb{R}^m \times \mathbb{R}^m ; x=y \}$ . On écrit le problème (P1) dans l'espace produit  $F \times F$  sous la forme équivalente

(P1') trouver  $(\mathcal{X}^*, \mathcal{Y}^*) \in \mathcal{C} \times \mathcal{D}$  tels que

$$\frac{1}{2} \|\mathcal{X}^* - \mathcal{Y}^*\|^2 = \text{Min} \left\{ \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|^2 ; (\mathcal{X}, \mathcal{Y}) \in \mathcal{C} \times \mathcal{D} \right\} \quad (110)$$

où  $\mathcal{X} = (x_1, x_2)$  et  $\mathcal{Y} = (y, y)$

Le problème (P1') fait intervenir la norme  $\|\cdot\|$  définie sur  $F \times F$  à partir de la norme Euclidienne  $\|\cdot\|_{D_Q}$  induite sur  $F$  par la métrique  $D_Q$  :

$$\|x\|^2 = \sum_j \lambda_j \|x_j\|_{D_Q}^2 \quad \sum \lambda_j = 1 \quad \lambda_j \in [0,1]$$

de sorte que le problème (P1') équivalent au problème (P1) s'écrit encore sous la forme:

(P1'') trouver  $(x_1^*, x_2^*, y^*) \in K_1 \times K_2 \times \mathbb{R}^m$  tel que

$$\frac{1}{2} \sum_j \lambda_j \|x_j^* - y^*\|_{D_Q}^2 = \text{Min} \left\{ \frac{1}{2} \sum_j \lambda_j \|x_j - y\|_{D_Q}^2 ; (x_1, x_2, y) \in K_1 \times K_2 \times \mathbb{R}^m \right\}$$

Posons  $u = (x_1, x_2, y)$  et  $f(u) = \frac{1}{2} \sum_j \lambda_j \|x_j - y\|_{D_Q}^2$ .  $f$  est une fonction réelle définie sur  $[\mathbb{R}^m]^3$ . (P1'') s'écrit alors

$$\text{Min} \{ f(u) ; u \in K_1 \times K_2 \times \mathbb{R}^m \}$$

La forme (P1'') permet de retrouver l'expression (101) de  $L(X)$  utilisée par BORG et LINGOES (1980) pour le problème de Codage Multidimensionnel de la matrice de dissimilarité  $\Delta$  sous les contraintes définies par une dissimilarité  $\Gamma$ . En posant  $F = F \wedge F$ ,  $m = n(n-1)/2$  et  $x_j = \delta_j$ , on montre ainsi cette fois dans le cadre d'une  $\Sigma$ -stratégie que la solution du problème (P1) est fournie par l'Analyse Canonique des cônes convexes fermés engendrés par  $\Delta$  et  $\Gamma$  respectivement. La quantité optimisée par une telle démarche

$$\frac{1}{2} \sum_j \lambda_j \|\delta_j - t\|_{D_Q}^2$$

fournit une disparité  $t$  telle que :

(a) Si  $K_1 \cap K_2 \neq \emptyset$ ,  $t$  est une fonction définie sur  $I \times I$  qui vérifie les contraintes  $K_1$  et  $K_2$ . En particulier si les deux ensembles  $K_j$  sont des cônes de fonctions monotones de  $\delta_j$ ,  $t$  est une fonction de deux variables, monotone par rapport à chacune d'elles,

(b)  $t$  fournit un intermédiaire de calcul à partir duquel on construit la solution  $d(X)$  recherchée qui réalise le minimum de la fonction critère

$$SS(X) = \|t - d(X)\|_{D_Q}^2$$

à l'aide d'un algorithme de Codage Multidimensionnel usuel, tel que celui de de LEEUW et HEISER (1977) présenté plus haut.

**Remarque :** On constate que sous cette forme, la fonction  $SS(X)$  généralise la fonction STRESS utilisée dans le cas du Codage Multidimensionnel sans contraintes puisqu'ici l'indice de disparité  $t$  s'exprime comme fonction des deux variables  $\delta_1$  et  $\delta_2$ .

#### IV.2.1.2.2 Résolution du problème

On doit construire  $\underline{u}^* \in K_1 \times K_2 \times \mathbb{R}^m$  tel que  $f(\underline{u}^*) = \text{Min} \{ f(u) ; u \in K_1 \times K_2 \times \mathbb{R}^m \}$ . La méthode de résolution consiste à construire une suite  $\underline{u}^0, \dots, \underline{u}^k, \dots$  d'éléments de  $(\mathbb{R}^m)^3$  convergeant vers la solution  $\underline{u}^*$  du problème (P1").

Posons

$$\begin{aligned}\underline{u}^k &= (\underline{x}_1^k, \underline{x}_2^k, \underline{y}^k) \\ f_{1,k}(\underline{x}_1) &= f(\underline{x}_1, \underline{x}_2^k, \underline{y}^k) \\ f_{2,k}(\underline{x}_2^k) &= f(\underline{x}_1^{k+1}, \underline{x}_2, \underline{y}^k) \\ f_{3,k}(\underline{y}^k) &= f(\underline{x}_1^{k+1}, \underline{x}_2^{k+1}, \underline{y})\end{aligned}$$

On passe alors de  $\underline{u}^k$  à  $\underline{u}^{k+1}$  en effectuant une seule itération d'une méthode destinée à résoudre chacun des problèmes

$$\text{Min} \{ f_{j,k}(\underline{x}_j) ; \underline{x}_j \in K_j \} \quad j = 1, 2, 3$$

où  $\underline{x}_3 = \underline{y}$  et  $K_3 = \mathbb{R}^m$ . Pour  $j=1$ , on écrit

$$f_{1,k}(\underline{x}_1) = \frac{1}{2} [\lambda_1 \|\underline{x}_1 - \underline{y}^k\|_{D_Q}^2 + \lambda_2 \|\underline{x}_2^k - \underline{y}^k\|_{D_Q}^2] \quad \underline{x}_1 \in K_1 \quad (111)$$

Pour  $j=2$ , on écrit de même

$$f_{2,k}(\underline{x}_2) = \frac{1}{2} [\lambda_1 \|\underline{x}_1^{k+1} - \underline{y}^k\|_{D_Q}^2 + \lambda_2 \|\underline{x}_2 - \underline{y}^k\|_{D_Q}^2] \quad \underline{x}_2 \in K_2 \quad (112)$$

De plus  $\underline{y}^{k+1}$  est solution du problème de minimisation SANS contraintes de la fonction

$$f_{3,k}(\underline{y}) = \frac{1}{2} [\lambda_1 \|\underline{x}_1^{k+1} - \underline{y}\|_{D_Q}^2 + \lambda_2 \|\underline{x}_2^{k+1} - \underline{y}\|_{D_Q}^2] \quad \underline{y} \in \mathbb{R}^m \quad (113)$$

Par suite

$$\begin{aligned}\nabla f_{3,k}(\underline{y}^{k+1}) &= \lambda_1 (\underline{x}_1^{k+1} - \underline{y}^{k+1}) + \lambda_2 (\underline{x}_2^{k+1} - \underline{y}^{k+1}) = 0 \\ \Leftrightarrow \underline{y}^{k+1} &= \lambda_1 \underline{x}_1^{k+1} + \lambda_2 \underline{x}_2^{k+1} = \lambda_1 P_{K_1}(\underline{y}^k) + \lambda_2 P_{K_2}(\underline{y}^k)\end{aligned}$$

c'est dire que  $\underline{y}^{k+1}$  est le barycentre pour les poids  $\lambda_j$  des projections  $P_{K_j}(\underline{y}^k)$ .

**Théorème 3** (BOUGHAZI (1987) page 135) :

- Si  $K_1 \cap K_2 \neq \emptyset$ , les trois suites  $(\underline{x}_1^k)$ ,  $(\underline{x}_2^k)$  et  $(\underline{y}^k)$  définies par l'algorithme ci-dessus convergent vers  $\underline{x}_1^* = \underline{x}_2^* = \underline{y}^* \in K_1 \cap K_2$
- si  $K_1 \cap K_2 = \emptyset$  et si l'un des deux cônes  $K_j$  est borné, les suites  $(\underline{x}_1^k)$ ,  $(\underline{x}_2^k)$  et  $(\underline{y}^k)$  convergent respectivement vers  $P_{K_1}(\underline{y}^*)$ ,  $P_{K_2}(\underline{y}^*)$  et  $\underline{y}^* \in \mathbb{R}^m$  de sorte que

$$\lim_{k \rightarrow \infty} \| \underline{x}_1^k - \underline{x}_2^k \|_{D_Q}^2 = d(K_1, K_2) = d(P_{K_1}(\underline{y}^*), P_{K_2}(\underline{y}^*)) \quad (114)$$

**Remarque :**  $\underline{y}^*$  est défini comme le barycentre des projections  $P_{K_j}(\underline{y}^*)$ . Par suite, en Codage Multidimensionnel sous contraintes, la disparité  $\underline{t}^*$  correspondant à l'optimum vérifie la décomposition additive du critère optimisé

$$\frac{1}{2} \sum_{j=1}^2 \lambda_j \| \underline{\delta}_j^* - \underline{d}(X) \|_{D_Q}^2 = \frac{1}{2} \sum_{j=1}^2 \lambda_j \| \underline{\delta}_j^* - \underline{t}^* \|_{D_Q}^2 + \| \underline{t}^* - \underline{d}(X) \|_{D_Q}^2$$

puisque

$$\sum_j \lambda_j \langle \underline{\delta}_j^* - \underline{t}^*, \underline{t}^* - \underline{d}(X) \rangle_{D_Q} = \langle \sum_j \lambda_j \underline{\delta}_j^* - \underline{t}^*, \underline{t}^* - \underline{d}(X) \rangle_{D_Q}$$

Les quantités

$$\frac{1}{2} \lambda_j \| \underline{\delta}_j^* - \underline{t}^* \|_{D_Q}^2 + \| \underline{t}^* - \underline{d}(X) \|_{D_Q}^2 \quad j = 1, 2$$

peuvent alors être utilisées pour juger de la qualité de la solution et les comparer l'une à l'autre pour juger de leur importance relative dans l'élaboration de la solution.

L'algorithme de résolution proposé ici opère donc en deux phases jusqu'à la convergence

- (1) construction de la disparité  $\underline{t}^*$  décrivant au mieux la proximité entre les données observées et les contraintes,
- (2) construction de la représentation  $\underline{d}(X)$  par un algorithme usuel de codage multidimensionnel.

**Remarque :** la formulation précédente utilise les propriétés géométriques de l'espace  $F \wedge F$ . Par suite il s'applique clairement au cas du STRESS, du SSTRESS ou de l'optimisation de la fonction  $\psi(\alpha, 2, \underline{d}(X)) = \sum \sum q_{ij} (\delta_{ij}^\alpha - d_{ij}^\alpha(X))^2$  puisque seule la structure des ensembles  $K_j$  est prise en compte dans les théorèmes fondant cette approche.

#### IV.2.2. CAS D'UNE CONTRAINTÉ DÉFINIE PAR PLUSIEURS MATRICES DE DISSIMILARITÉ

L'intérêt d'une formulation du problème dans le cadre d'une  $\Sigma$ -stratégie est de fournir un cadre naturel d'extension de la démarche suivie dans le cas de deux ensembles  $K_j$  au cas

de  $p$  ensembles  $K_j$ . Deux cas particuliers présentent un intérêt pratique immédiat. Tout d'abord, en Analyse Ordinale des Différences Individuelles, on dispose d'un échantillon de  $m$  tableaux de dissimilarités  $\{\Delta^j, j = 1, \dots, m\}$  que l'on se propose de représenter par une matrice de configuration  $X$  de rang  $r$  minimal, de sorte que la matrice des distances Euclidiennes  $D(X)$  associée, représente au mieux l'échantillon analysé. Lorsque les  $\Delta^j$  sont d'échelle d'intervalles, l'approche présentée ici justifie le choix du barycentre des  $\Delta^j$  comme pseudo-donnée approchée par  $\underline{d}(X)$ . Ce choix est adopté par de LEEUW (1980) et HEISER et STOOP (1986). Lorsque les  $\Delta^j$  sont d'échelle ordinale, ou bien d'échelles mixtes (certaines sont ordinales, d'autres nominales ou bien d'intervalles), la méthode présentée ici fournit un cadre rigoureux de choix d'une disparité de référence, qui manquait à notre connaissance à la méthodologie de l'analyse d'un échantillon de dissimilarités  $\{\Delta^j, j = 1, \dots, m\}$ . De plus, l'analogie au cadre de l'Analyse Canonique Linéaire généralisée n'est pas dépourvue d'intérêt dans l'évaluation du potentiel et surtout des limites d'une approche fondée sur l'introduction de contraintes induites par un tableau de dissimilarité.

La deuxième justification d'une telle approche est fondée sur l'étude de données particulières non prises en compte par les méthodes existantes. Le problème de Codage Multidimensionnel Ordinal est en général posé en termes de contraintes monotones nettes. On suppose en effet donnée une relation d'ordre nette entre paires. Une extension simple de cette méthode consiste à prendre en compte une contrainte d'ordre flou, suivant la définition suivante:

**Définition 9** (ZADEH (1971)) : *Une relation floue  $P$  est un ordre flou si elle est transitive.  $P$  est un ordre partiel flou si et seulement si  $P$  est reflexive, transitive et antisymétrique, c'est à dire telle que*

$$\mu_P(i,j) > 0 \text{ et } \mu_P(j,i) > 0 \Rightarrow (i=j) \quad \forall i,j \in I$$

Cet auteur vérifie alors la propriété de décomposition:

**Proposition 6** : *Soit  $P = \sum \alpha P_\alpha$ ,  $\alpha \in ]0,1[$  une décomposition d'une relation d'ordre partiel flou définie sur  $I \times I$ . Alors  $P_\alpha$  est une relation d'ordre net définie sur  $I \times I$ . Réciproquement si les  $P_\alpha$  ( $\alpha \in ]0,1[$ ) forment une suite emboîtée de relations d'ordre partiel net distinctes telles que*

- (a)  $\alpha_1 > \alpha_2 \Leftrightarrow P_{\alpha_1} \subset P_{\alpha_2}$
- (b)  $P_1 \neq \emptyset$  et  $\text{dom}P_1 = \text{dom}P_\alpha$

alors pour tout choix des  $\alpha$  dans  $]0,1[$ ,  $P$  est une relation d'ordre partiel flou.

L'étude qui suit montre que l'on peut étendre de façon simple l'approche usuelle des problèmes de Codage Multidimensionnel Ordinal au cas d'ordres partiels flous.

Enfin, dans le cadre de l'analyse d'un tableau de dissimilarité sous des contraintes définies par la donnée de  $(p-1)$  tableaux de dissimilarité  $\{\Delta^j, j = 2, 3, \dots, p\}$ , l'extension de la méthode de pénalisation proposée par BORG et LINGOES dans le cas  $p=2$ , conduit à minimiser la fonction

$$L(X) = \frac{1}{2} \sum_{j=1}^2 \lambda_j \| \delta_j - d(X) \|_{D_Q}^2 \quad \lambda_j \in [0,1] \quad \sum_{j=1}^2 \lambda_j = 1$$

et comme dans le cas  $p=2$ , cette démarche pose des problèmes difficiles de gestion des coefficients  $\lambda_j$ . Ces auteurs posent la contrainte simplificatrice

$$\lambda_j = \text{constante}, \quad j=2, 3, \dots, p$$

En fait, la valeur optimale des coefficients  $\lambda$  peut être déterminée à l'aide d'une Régression Multiple Optimale, puis utilisée par l'algorithme étudié ci dessous. Notons cependant à nouveau que dans le cas d'une Analyse Canonique Multiple menée dans un cadre linéaire, on choisit implicitement  $\lambda_j = \frac{1}{p}$ .

#### IV.2.2.1. Analyse canonique de $p$ convexes fermés ( $p \geq 2$ )

Considérons  $p$  sous ensembles convexes fermés  $K_j$  de l'espace Euclidien  $F = (\mathbb{R}^m, D_Q)$ , décrivant les données observées et les contraintes et utilisons un intermédiaire de calcul  $T$ , que nous appelons une matrice de disparité et qui s'interprète comme une mesure composite de la dissemblance des unités statistiques. Le problème peut se formuler dans les termes suivants:

$$(Q1) \quad \text{Trouver la disparité } \mathbf{t} \in F \text{ telle que } \mathbf{t} \in \bigcap_{j=1}^p K_j$$

On munit alors l'espace vectoriel  $\mathcal{H} = [\mathbb{R}^m]^p$  du produit scalaire

$$\langle\langle u, v \rangle\rangle = \sum_{j=1}^p \lambda_j \langle u^j, v^j \rangle_{D_Q} \quad \sum_{j=1}^p \lambda_j = 1 \quad \lambda_j \in [0,1] \quad (116)$$

et l'on note  $K = \bigcap_{j=1}^p K_j$ ,  $\mathcal{C} = K_1 \times K_2 \times \dots \times K_p$  et  $\mathcal{X} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)$  où  $\underline{x}_j \in F$  pour tout  $j = 1, \dots, p$ .

Introduisons de plus le convexe diagonal

$$\mathcal{D} = \{ \mathbf{y}; \mathbf{y} = (\underline{y}, \dots, \underline{y}) \in \mathcal{H} \text{ où } \underline{y} \in F \}.$$

**Proposition 7:**

1) si  $\underline{y} \in \mathcal{H}$  alors  $\mathbf{y} \in \mathcal{C} \cap \mathcal{D}$

2) la projection de  $\mathbf{y} \in \mathcal{D}$  sur le convexe  $\mathcal{C}$  a pour composantes

$$P_{\mathcal{C}}(\mathbf{y}) = (P_{K_1}(\underline{y}), P_{K_2}(\underline{y}), \dots, P_{K_p}(\underline{y})) \quad (117)$$

3) la projection du vecteur  $\mathcal{X} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)$  sur  $\mathcal{D}$  a pour composantes le barycentre des  $p$  points  $\underline{x}_j$ ,  $j = 1, \dots, p$ .

$$P_{\mathcal{D}}(\mathcal{X}) = \left( \sum_j \lambda_j \underline{x}_j, \sum_j \lambda_j \underline{x}_j, \dots, \sum_j \lambda_j \underline{x}_j \right) \quad (118)$$

**Preuve :** 1) résulte directement de la définition de  $\mathcal{D}$ . Pour 2), soit  $\mathbf{y} \in \mathcal{D}$ , alors la projection sur un convexe fermé vérifie

$$\begin{aligned} \|\| P_{\mathcal{C}}(\mathbf{y}) - \mathbf{y} \|\|^2 &= \inf \{ \|\| \mathbf{y} - \mathcal{X} \|\|^2 ; \mathcal{X} \in \mathcal{C} \} \\ &= \inf \left\{ \sum_j \lambda_j \|\| \underline{x}_j - \underline{y} \|\|_{D_Q}^2 ; \underline{x}_j \in K_j, j = 1, 2, \dots, p \right\} \\ &= \sum_j \lambda_j \|\| P_{K_j}(\underline{y}) - \underline{y} \|\|_{D_Q}^2 \end{aligned}$$

La solution optimale  $\mathcal{X}^* = (\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_p^*)$  est donc

$$\mathcal{X}^* = (\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_p^*) = (P_{K_1}(\underline{y}^*), P_{K_2}(\underline{y}^*), \dots, P_{K_p}(\underline{y}^*)) = P_{\mathcal{C}}(\mathbf{y}^*)$$

Pour 3), soit  $\mathcal{X} \in \mathcal{C}$ , comme en 2), on écrit

$$\begin{aligned} \|\| P_{\mathcal{D}}(\mathcal{X}) - \mathcal{X} \|\|^2 &= \inf \{ \|\| \mathbf{y} - \mathcal{X} \|\|^2 ; \mathbf{y} \in \mathcal{D} \} \\ &= \inf \left\{ \sum_j \lambda_j \|\| \underline{x}_j - \underline{y} \|\|_{D_Q}^2 ; \underline{y} \in F \right\} \end{aligned}$$

On doit donc minimiser sur tout l'espace la fonction convexe

$$f(\underline{y}) = \sum_j \lambda_j \|\| \underline{x}_j - \underline{y} \|\|_{D_Q}^2$$

$\underline{y}$  est solution si et seulement si le gradient de  $f(\underline{y})$  est nul, c'est à dire si et seulement si

$$\sum_j \lambda_j (\underline{x}_j - \underline{y}) = 0 \Leftrightarrow \underline{y} = \sum_j \lambda_j \underline{x}_j$$

ce qui établit le résultat 3).  $\square$

#### IV.2.2.2. Résolution du problème

Cette formulation du problème généralise immédiatement les résultats établis dans le cas  $p = 2$ . On vérifie en effet le théorème (BOUGHAZI (1987) page 138):

**Théorème 4:** Sous l'hypothèse  $K = \bigcap_{j=1}^p K_j \neq \emptyset$ , les deux problèmes suivants sont équivalents :

(Q1) : trouver  $\underline{y} \in F$  tel que  $\underline{y} \in K$ ,

(Q2) : trouver  $(\underline{x}^*, \underline{y}^*) \in \mathcal{E} \cap \mathcal{D}$  réalisant le minimum de

$$\| \underline{x}^* - \underline{y}^* \|^2 = \text{Min} \{ \| \underline{x} - \underline{y} \|^2 ; (\underline{x}, \underline{y}) \in \mathcal{E} \times \mathcal{D} \} \quad (119)$$

On vérifie de plus comme précédemment l'équivalence des trois propriétés

- 1)  $(\underline{x}^*, \underline{y}^*) \in \mathcal{E} \cap \mathcal{D} \quad d(\underline{x}^*, \underline{y}^*) = d(\mathcal{E}, \mathcal{D})$
- 2)  $(\underline{x}^*, \underline{y}^*) \in \mathcal{E} \cap \mathcal{D} \quad d(\underline{x}^*, \underline{y}^*) = d(\underline{x}^*, \mathcal{D}) = d(\mathcal{E}, \underline{y}^*)$
- 3)  $(\underline{x}^*, \underline{y}^*) \in \mathcal{E} \cap \mathcal{D} \quad \underline{y}^* = P_{\mathcal{E}}(\underline{x}^*)$

Cette dernière caractérisation (propriété 3)) s'exprime dans  $F$ , sous la forme

$$\begin{aligned} \underline{x}_p^* &= P_{K_j}(\underline{y}^*) & \underline{x}^* &= (\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_p^*) \\ \underline{y}^* &= \sum_j \lambda_j \underline{x}_j^* & \underline{y}^* &= (\underline{y}, \dots, \underline{y}) \end{aligned} \quad (120)$$

L'équivalence des deux propriétés 1) et 2) précédentes permet alors de montrer l'équivalence entre les problèmes traités dans la proposition 7 et les problèmes

$$(Q3) : \text{Min} \left\{ \frac{1}{2} \| \underline{x} - P_{\mathcal{D}}(\underline{x}) \|^2 ; \underline{x} \in \mathcal{E} \right\}$$

$$(Q4) : \text{Min} \left\{ \frac{1}{2} \| \underline{y} - P_{\mathcal{E}}(\underline{y}) \|^2 ; \underline{y} \in \mathcal{D} \right\}$$

Par suite l'application d'un algorithme de gradient projeté à l'un des problèmes (Q3) ou (Q4) permet de construire la solution. Notons qu'alors, pour le problème (Q4) l'itéré  $\underline{y}^{k+1}$  est donné par l'équation

$$\begin{aligned} \mathbf{y}^{k+1} &= \mathbf{y}^k + \alpha_k [P_{\mathcal{D}} P_{\mathcal{C}}(\mathbf{y}^k) - \mathbf{y}^k] & \alpha_k > 0 \\ \Leftrightarrow \mathbf{y}^{k+1} &= \mathbf{y}^k + \alpha_k \left[ \sum_j \lambda_j P_{K_j}(\mathbf{y}^k) - \mathbf{y}^k \right] & \sum_j \lambda_j = 1 \quad \lambda_j > 0 \end{aligned}$$

De plus, à l'optimum,  $\mathbf{y}^*$  est le barycentre de ses projections sur les ensembles de contraintes  $K_j$  :

$$\mathbf{y}^* = \sum_j \lambda_j P_{K_j}(\mathbf{y}^*) \quad (122)$$

D'autre part, l'expression de l'itération du gradient projeté appliqué au problème (Q3) s'écrit

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \alpha_k [P_{K_i}(\sum_j \lambda_j \mathbf{x}_j^k) - \mathbf{x}_i^k] \quad \sum_j \lambda_j = 1 \quad \lambda_j > 0$$

et à l'optimum

$$\mathbf{x}_i^* = P_{K_i}(\sum_j \lambda_j \mathbf{x}_j^*) \quad (123)$$

De sorte que l'on est conduit à chaque itération à résoudre des problèmes relativement simples de calcul de la projection  $D_Q$ -orthogonale d'un vecteur sur une partie convexe fermée de  $F$ .

De nombreux algorithmes peuvent être mis en œuvre pour résoudre sous l'une des formes équivalentes (Q1) à (Q4). Ceux-ci sont étudiés en détail et comparés entre eux par BOUGHAZI (1987) auquel nous renvoyons le lecteur intéressé. Notons cependant qu'en particulier, du fait des problèmes de différentiabilité de la fonction objectif aux points  $X$  pour lesquels il existe  $i \neq j$  tels que  $d_{ij}(X) = 0$ , cet auteur décrit une méthode de sous-gradient directement applicable au problème traité ici, dont l'expérimentation numérique est en cours.

Comme dans le cas d'une seule contrainte, on construit donc la disparité  $\mathfrak{t}$  à partir de laquelle un algorithme usuel de Codage Multidimensionnel est appliqué pour construire la représentation Euclidienne  $d(X)$  optimale, au sens du critère aux moindres carrés :

$$S = \frac{1}{2} \left( \sum_j \lambda_j \|\delta_j^* - d(X)\|_{D_Q}^2 \right)$$

où les poids  $\lambda_j$  sont fixés par l'utilisateur.

La validité de la démarche précédente est assurée pour des ensembles de contraintes  $K_j$  très généraux. Nous nous sommes limité à des cônes convexes polyédriques fermés, bien que les algorithmes de programmation convexe mis en œuvre restent justifiés lorsque les ensembles  $K_j$  sont des convexes fermés quelconques et plus généralement des sous-ensembles proximaux de  $F$  (cf. ROCKAFELLAR (1970)). L'approche décrite ici permet donc de retrouver comme cas particuliers tous les types de contraintes prises en compte par BENTLER et WEEKS (1978) SKARABIS (1978) et BORG et LINGOES (1980). La formulation de ces auteurs est cependant critiquable puisqu'elle fait intervenir le gradient de  $\underline{d}(X)$  alors que cette fonction de  $X$  n'est pas différentiable en zéro.

Notre formulation décompose le problème en deux étapes. Les distances  $\underline{d}(X)$  n'interviennent qu'à la deuxième étape, au travers d'un algorithme classique de Codage Multidimensionnel qui, comme l'algorithme de sous-gradient développé par de LEEUW et HEISER et décrit plus haut, présente lui une formulation correcte du problème puisque  $\underline{d}(X)$  est sous différentiable pour la classe des fonctions de distance usuellement utilisées dans les problèmes de Codage Multidimensionnel.

#### IV.2.3. DÉPOUILLEMENT DES RÉSULTATS

La formulation du problème utilisée dans les paragraphes précédents permet de plus de juger a posteriori de la qualité de la solution construite. L'ajustement de chaque composante est évalué à l'aide des fonctions d'écart

$$S_j = \frac{\|\underline{\delta}_j^* - \underline{d}(X)\|_{DQ}^2}{S} \quad \text{et} \quad S_j^* = \lambda_j S_j \quad (124)$$

qui mesurent la contribution des dissimilarités  $\underline{\delta}_j^*$  à la construction de  $\underline{d}(X)$ .  $S_j$  est proportionnel au STRESS qui mesure le qualité de l'ajustement d'une représentation Euclidienne  $X$  aux dissimilarités contenues dans le tableau  $\Delta^j$  en l'absence d'autres informations. L'importance relative des quantités  $S_j$  fournit en particulier un outil d'analyse permettant de fixer a posteriori les valeurs des pondérations  $\lambda_j$  de façon satisfaisante en fonction des objectifs de l'étude.

La comparaison des quantités figurant au numérateur de  $S_j$  à la valeur du STRESS obtenue lors de l'analyse sans contraintes de  $\Delta^1$ , pour un même nombre de dimensions demandées, permet d'évaluer l'influence des contraintes sur la qualité de l'ajustement. Cependant le caractère non linéaire du problème impose de réaliser ce Codage Multidimensionnel sans contraintes pour évaluer le STRESS.

D'autre part, le tenseur des distances Euclidiennes reconstruites  $\underline{d}(X)$  est construit en tant que meilleure approximation de  $\underline{t}$  et non des  $\underline{\delta}_j^*$ . On est donc amené à construire la projection  $\underline{t}_j^*$  de  $\underline{d}(X)$  sur  $K_j$  afin de juger de la contribution de  $\underline{d}(X)$  à la description des proximités  $\Delta^j$ . On obtient ainsi un diagramme de Shepard associé à chaque tableau  $\Delta^j$  qui permet d'étudier la transformation des données  $\Delta^j$  imposée lors de la construction de la solution.

Le réalisme de ces deux procédures d'évaluation de la solution diminue cependant vite lorsque  $p$  croît, du fait de la masse imposante de calculs qu'elles supposent.

Enfin, le critère  $S$  fournit un outil analogue à la variance pour juger de la qualité globale de la solution. Ainsi:

(a) la quantité

$$E = \frac{1}{2} \left( \sum_j \lambda_j \| \underline{\delta}_j^* - \underline{t}^* \|_{DQ}^2 \right) \quad (125)$$

mesure la cohérence moyenne entre les contraintes et les dissimilarités observées. Cet indice est nul si  $K \neq \emptyset$  et il est d'autant plus grand que les éléments  $K_j$  diffèrent, rendant la résolution du problème peu justifiée. De plus, du fait de l'expression de  $\underline{t}$ , cet indice s'écrit

$$E = \frac{1}{2} \sum_j \lambda_j \| \underline{\delta}_j^* \|_{DQ}^2 - \| \underline{t}^* \|_{DQ}^2$$

De plus, si pour tout  $j$  ( $j = 1, 2, \dots, p$ )  $K_j$  est le cône polyédrique fermé défini par une variable ordinale  $X^j$  ( le cas des matrices de dissimilarité ordinale est un cas particulier) mesurée sur un ensemble fini  $I$  ( $\text{card}(I) = n$ ), alors le problème posé en Analyse Canonique de  $p$  cônes polyédriques de  $F = (\mathbb{R}^n, N)$  généralise les méthodes usuelles d'analyse de la concordance entre juges (identifiés aux variables ici). En effet, les indices de concordance existants reposent sur la détermination de codages a priori des éléments de  $I$ . Ainsi, le coefficient de concordance de KENDALL et BADINGTON SMITH (cf. KENDALL (1962), DAVID (1988)) repose sur le codage par les rangs. Ici, les codages optimaux monotones  $\underline{\delta}_j^*$  sont construits de façon à minimiser la quantité

$$\frac{1}{2} \left( \sum_j \lambda_j \| \underline{\delta}_j^* - \underline{t}^* \|_{DQ}^2 \right)$$

On peut alors comparer la valeur du coefficient de concordance  $W$ , qui s'exprime en particulier en fonction de la moyenne des coefficients de corrélation de SPEARMAN

calculés sur les  $\binom{p}{2}$  paires possibles de variables, sous la forme (KENDALL (1962) page 95)

$$W = \frac{1 + (p - 1) \bar{r}_s}{p} \quad (126)$$

On peut donc calculer ainsi la moyenne  $\bar{r}^*$  des  $\binom{p}{2}$  corrélations linéaires  $r(\delta_j^*, \delta_k^*)$  et le coefficient de concordance optimal associé  $W^*$ .  $W^*$  fournit un indicateur global de concordance entre les contraintes et les données. Si cet indice est fort,  $\underline{t}^*$  est alors un bon résumé de l'information ordinale portée par les  $\Delta^j$ .

Une telle approche doit de plus être rapprochée des indices développés pour l'étude des données de préférence par d'AUBIGNY (1979). Lorsque la valeur de l'indice E est élevée, l'Analyse en Composantes Principales ordinaire du tableau des  $\delta_j^*$  permet de détecter les arêtes (i,l) qui sont responsables du désaccord. On mène une telle analyse soit à l'aide de codages a priori (cf. d'AUBIGNY (1979)), soit en termes de codage optimal (cf. YOUNG, TAKANE et de LEEUW (1978), TENENHAUS (1977)). Le codage fourni par les coordonnées des unités statistiques sur la première composante principale produit alors un résumé  $\underline{t}'$  distinct de  $\underline{t}^*$  puisqu'alors les coefficients  $\alpha_j$  de la combinaison linéaire

$$\underline{t}' = \sum_j \alpha_j \delta_j^*$$

résultent de l'analyse. L'apparition éventuelle de coefficients  $\alpha_j$  négatifs traduit alors une incohérence ordinale entre les contraintes et les données. Dans le cas contraire,  $\underline{t}'$  s'interprète comme un indice de dissimilarité composite. Il est associé à une solution admissible du problème précédent

$$\underline{t} = \sum_j \lambda_j \delta_j^*$$

où l'on pose

$$\lambda_j = \frac{\alpha_j}{\sum_j \alpha_j}$$

$\underline{t}$  est ainsi décrit comme combinaison convexe des  $\delta_j^*$  et la consultation des coefficients  $\lambda_j$  permet de déterminer les coefficients  $\lambda_j$  pris en compte dans l'Analyse Canonique des p convexes fermés  $K_j$ , en fonction des objectifs de l'étude.

(b) le rapport

$$\frac{\sum_j \lambda_j \|\delta_j^* - \underline{d}(X)\|_{DQ}^2}{\sum_j \lambda_j \|\delta_j^* - \underline{t}^*\|_{DQ}^2} = 1 - 2 \frac{\|\underline{t}^* - \underline{d}(X)\|_{DQ}^2}{\sum_j \lambda_j \|\delta_j^* - \underline{t}^*\|_{DQ}^2} \quad (127)$$

fournit un indicateur complémentaire de l'indice E. Il mesure de la qualité de la configuration X pour représenter les unités statistiques dans un espace vectoriel de dimension r fixée. Il fournit en effet une mesure de la qualité de D(X) comme résumé des tableaux  $\Delta^j$ . Il permet en effet de comparer la qualité des approximations des données dues au plongement Euclidien des unités statistiques dans un espace vectoriel de dimension r à partir de l'information fournie par la dissimilarité composite  $\underline{t}^*$ , à l'écart quadratique moyen résultant de l'imposition simultanée de l'ensemble des contraintes. A niveau de cohérence fixé des contraintes, mesuré par E, le rapport (127) mesure le degré d'Euclidéanité moyen des tableaux  $\Delta^j$  et donc celui du tableau T\*.

On peut enfin partitionner le numérateur du rapport (127), de façon à juger de la qualité de la reconstitution des données observées d'une part et des contraintes d'autre part à l'aide de  $\underline{d}(X)$ .



## REFERENCES

- d'AUBIGNY, G. (1975). Description Statistique des Données Ordinales: Analyse Multidimensionnelle. Thèse de 3ème cycle. Université de Grenoble I.
- d'AUBIGNY, G. (1979). " A Probabilistic Look at Ordinal Methods In Data Analysis." In P. MANDL et M. HUSKOVA, Eds., Proceedings of the Second Prague Symposium on Asymptotic Statistics. North Holland, Amsterdam: 23-40.
- d'AUBIGNY, G. (1981). " Analyse Factorielle et Visualisation des Préférences." In P. BATTEAU et al., Eds., Analyse et Agrégation des préférences. Economica, Paris: 177-212.
- d'AUBIGNY, G. (1984). " Un Modèle Algébrique pour l'Analyse des données de dissimilarité." Troisièmes Journées Franco-Soviétiques de Statistique et Analyse des Données. Paris. Tech. Rep. IMSS (1987), Grenoble.
- d'AUBIGNY, G. (1988). " The Additive Decomposition of Some Entropy Functions and (Constrained-) Ordination Methods." Papier présenté à la , XIV-th International Biometric Conference. Namur.
- d'AUBIGNY, G., POLIT, M.O. (1988). " Some Optimality Properties of the Generalization of the TUCKER Method to the Analysis of n-way Tables with Specified Metrics." In L. COPPI, ed., Proceedings of the MULTIWAY 88 Conference. Rome.
- d'AUBIGNY, G., PHAM DINH, T., YASSINE, A. (1988). " A Conditional Conjugate Gradient Method for Solving Isotonic and Concave Regression Problems." Computational Statistics and Data Analysis. ( à paraître).
- BALESTRA, P. (1975). La Dérivation Matricielle. Sirey, Paris.
- BANET, T.A. et L. LEBART. (1984). "Local and Partial Principal Component Analysis and Correspondence Analysis." COMPSTAT. Vienna : Physica Verlag.
- BASTIN, C., BENZECRI, J.-P., BOURGARIT, C., CAZES, P. (1980). " Pratique de l'Analyse des Données." Dunod. Paris.
- BEALS, R. et D.H. KRANTZ. (1967). "Metrics and Geodesics Induced by Order Relations." Mathematische Zeitschrift **101** : 285-298.
- BEALS, R., D.H. KRANTZ et A. TVERSKY. (1968). "Foundations of Multidimensional Scaling." Psychological review **75** : 127-142.
- BENINEL, F. (1987). "Problèmes de Représentations Sphériques des Tableaux de Dissimilarité." Thèse de Troisième Cycle, Univ. Rennes I.
- BENTLER, P.M. et D.G. WEEKS. (1978). "Restricted Multidimensional Scaling Models." J. of Mathematical Psychology Vol **17**, No. **2** : 138-151.
- BENZECRI, J.-P. et al. (1973). " L'Analyse des Données." Dunod. Paris.

- BERGE, C. (1973). "Graphes et Hypergraphes." Dunod. Paris.
- BLOXOM, B. (1968). "Individual Differences in Multidimensional Scaling." Research Bulletin 68-45, Educational Testing Service. Princeton.
- BLOXOM, B. (1974). "An Alternative Method for Fitting a Model of Individual Differences in Multidimensional Scaling." Psychometrika 93, 3: 365-367.
- BLOXOM, B. (1978). "Constrained Multidimensional Scaling in n Spaces." Psychometrika 43: 397-408.
- BONIFAS, L., Y. ESCOUFIER, P.L. GONZALEZ et R. SABATIER. (1984). "Choix de Variables en Analyse en Composantes Principales." Revue de Statistique Appliquée Vol XXXII, No. 2 : 5-15.
- BORG, I. et J.C. LINGOES. (1980). "A Model and Algorithm for Multidimensional Scaling with external constraints on the distances." Psychometrika 45 : 25-38.
- BOUGHAZI, M.A. (1987). Contribution à l'étude des algorithmes d'Optimisation en Analyse des Données. Thèse de 3ème cycle. Université de Grenoble I.
- BRADU, D. (1984). "Response Surface Model Diagnosis in Two-way Tables." Communications in Statist., Theor. Meth. 13, 24: 3059-3106.
- BROSSIER, G. (1986). "Problèmes de Représentations de Données par des Arbres." Thèse d'Etat, Université de Rennes I.
- CAELLI, T. (1981). "On Generating Spatial Configurations with Identical Interpoint Distance Distributions." In ROBINSON, R.W., SOUTHERN, G.W., WALLIS, W.D. EDS. Lecture Notes In Mathematics. Springer Verlag. New York.
- CAILLIEZ, F. (1983). "The Analytical Solution of the Additive Constant Problem." Psychometrika 48 : 305-310.
- CAILLIEZ, F. (1984). Analyse des Données. Les Presses de l'Université de Montréal. Montréal.
- CAILLIEZ, F., PAGES, J.P. (1976). Introduction à l'Analyse des Données. SMASH. 9 rue DUBAN, Paris.
- CARROLL, J.D. (1983) . "Common Dimension Analysis." Paper presented at the Multidimensional Data Analysis Workshop. Paris.
- CARROLL, J.D. et P. ARABIE (1980). "Multidimensional Scaling." In M.R. ROSENZWEIG and L.W. PORTER, eds. Annual Review of Psychology 31 : 607-649.
- CARROLL, J.D. et J.J. CHANG. (1970). "Analysis of Individual Differences in Multidimensional Scaling via a N-way Generalization of 'ECKART-YOUNG' Decomposition." Psychometrika 35 : 283-319.
- CARROLL, J.D., P.E. GREEN et F.J. CARMONE. (1976). "CANDELINC (CANonical Decomposition with Linear Constraints). A New Method for Multidimensional Analysis with Constrained Solutions." Paper presented at the International Congress of Psychology, Paris, France.
- CARROLL, J.D., P.E. GREEN et F.J. CARMONE. (1980). "CANDELINC : a General Approach to Multidimensional Scaling of Many-way Arrays with Linear Constraints on Parameters." Psychometrika 45 : 3-24.

- CARROLL, J.D. et S. PRUZANSKI. (1977). " MULTILINC : MULTIWAY CANDELINC (CANonical DEcomposition with LINear Constraints)." Paper presented at the American Psychological Association Meeting, San Francisco.
- CHATTERJEE, S., HADI, A.S. (1986). " Influencial Observations, High Leverage Points, and Outliers in Linear Regression." Staistical Science **1** , No 3: 379-416.
- COLLEDGE, R.G. et G. RUSHTON. (1972). " Multidimensional Scaling : Review and Geographical Applications." Geographical technical papers series No. 10. Association of American Geographers. Washington.
- COOK, R.D., WEISBERG, S. (1982). Residuals and Influence in Regression. Chapman and Hall, New York.
- COOPER, L.G. (1972). " A New Solution to the Additive Constant Problem in Metric Multidimensional Scaling." Psychometrika **37** : 311-322.
- COXETER, H.S.M. (1948). Regular polytopes. London : Methnen and Co.
- COXON, A.P.M. et C.L. JONES. (1974 a). " Occupational Similarities : Subjective Aspects of Social Stratification." Quality and Quantity **22** : 369-384.
- COXON, A.P.M. et C.L. JONES. (1974 b). " Problems in Selection of Occupational Titles." The Sociological Review **22** : 369-384.
- CRIPPEN, G.M. (1977). " A Novel Approach to Calculations of Conformation : Distance Geometry." J. of Computational Physics **24** : 96-107.
- CRIPPEN, G.M. (1978). " Rapid Calculation of Coordinates from Distance Measures." J. of Computational Physics **26** : 449-452.
- CROQUETTE, A. (1980). " Quelques Résultats Synthétiques en Analyse Des Données Multidimmsionnelles: Optimalité et Métriques à Effets Relationnels." Thèse de Troisième Cycle, Université Paul Sabatier, Toulouse.
- DANIELS, H.E. (1944). " The Relation Between Measures of Correlation in the Universe of Sample Permutations." Biometrika **33** : 129-135.
- DAUXOIS, J., POUSSE, A. (1976). " Les Analyses Factorielles en Calcul des Probabilités et en Statistique: Essais d'étude synthétique." Thèse d'Etat, Université Paul Sabatier, Toulouse.
- DAVID, H.A. (1988). The Method of Paired Comparisons. (Second edition). Charles Griffin and Co. London.
- DAVIDSON, M.L. (1983). Multidimensional Scaling. J. Wiley and Sons. New York.
- DENT, W.T. (1980). " On Restricted Estimation in Linear Models." J. of Econometrics, Vol **12**, No 1: 49-58.
- DROESLER, J. (1981 a). " The Empirical Validity of Multidimensional Scaling." In Multidimensional Data Representations : When and Why. I. BORG, Ed., Ann Arbor, Michigan : Mathesis Press, pp 627-651.
- DROESLER, J. (1981 b). " Foundations of Multidimensional Scaling in Cayley - Klein Geometry." In Multidimensional Data Representations : When and Why. I. BORG, Ed., Ann Arbor, Michigan : Mathesis Press, pp 652-678.

- DYKSTRA, R.L. (1983). " An Algorithm for Restricted Least Squares Regression." J. Amer. Statist. Assoc. 78, Vol. 78: 837-842.
- EL HADRI, K. (1988). Contribution Algorithmique à l'Analyse des Données de dissimilarité sous contraintes de lissage. Thèse (à paraître). Université de Grenoble II.
- EKMAN, G. (1954). " Dimensions of Color Vision." J. of Psychology, 38: 467-474.
- ESCOUFIER, Y. (1973). " Le Traitement des Variables Vectorielles." Biometrics 29 : 751-760.
- FASHAM, M.J.R. (1977). " A Comparison of Non Metric Multidimensional Scaling, Principal Components Analysis, and Reciprocal Averaging for the Ordination of Simulated Coenoclines and Coenoplanes." Ecology 58: 551-561.
- FERLIGOJ, A., BATAGELJ, V. (1982). " Clustering with Relational Constraint." Psychometrika 47 : 413-426.
- FERLIGOJ, A., BATAGELJ, V. (1983). " Some Types of Clustering with Relational Constraints." Psychometrika 48 : 541-552.
- FICHET, B. (1983). " Analyse Factorielle sur Tableaux de Dissimilarité: Application aux données sur Signe de Présence-Absence en Médecine." Thèse d'état, Univ. Marseilles.
- FICHET, B. (1985). " Distances and Euclidean Distances for Presence-Absence Characters and their Application to Factor Analysis. In de LEEUW et al. eds., Multidimensional Data Analysis. (1986) DSWO Press, Leiden: 23-46.
- FORGAS, J.P., E. KAGEN et D. FREY. (1977). " The Cognitive Representation of Political Personalities. A Cross-cultural Comparison." International J. of Psychology 12 : 19-30.
- FUNIA, S., A. HOROWITZ et F.W. YOUNG. (1976). " The Perceived Structure of American Ethnic Groups : the Use of Multidimensional Scaling Stereotype Research." Sociometry 39 : 116-130.
- GAUCH, H.G. (1982). Multivariate Analysis in Community Ecology. Cambridge University Press, Cambridge.
- GAUCH, H.G., WHITTAKER, R.H., SINGER, S.B. (1981). " A Comparative Study of Non Metric Ordinations." J. of Ecology 69: 135-152.
- GAUCH, H.G., WHITTAKER, R.H., WENTWORTH, T.R. (1977). " A Comparative Study of Reciprocal Averaging and Other Ordination Techniques." J. of Ecology 65: 157-174.
- McGEE, V.E. (1968). " Multidimensional Scaling of N sets of Similarity Measures: A Nonmetric Individual Differences Approach." Multivariate Behavioral Research, 3 : 233-248.
- McGEE, V.E. (1969). " Speech Perception and the Development of Appropriate Nonmetric Analytical Techniques." Tech. Rep., Darmouth College, Hanover, 66p.
- GEISSER, S. (1980). " Growth Curve Analysis." In P.R. KRISHNAIAH, Ed., Handbook of Statistics, Vol. 1, North Holland. 89-115.
- GILBERT, E.N. (1974). "Distorsion in Maps." SIAM Review 16 : 47-62.

- GLESER, L.J., OLKIN, I. (1970). "Linear Models in Multivariate Analysis." In Essays In Probability and Statistics. BOSE, R.C. et al., Eds., The Univ. of North Carolina Press, Chapel Hill: 267-292.
- GORDON, A.D. (1973 a). "Classification in the Presence of Constraints." in Analyse des données et Informatique, R.TOMASSONE ed.INRIA, Le Chesnay: 161-171.
- GORDON, A.D. (1973 b). "Methods of Constrained Classification." Biometrics, 29: 821-827.
- GOWER, J.C. (1966). "Some Distance Properties of Latent Roots and Vectors Used in Multivariate Analysis." Biometrika 53 : 325-338.
- GOWER, J.C. (1971). "Statistical Methods of Comparing Different Multivariate Analyses of the same Data." In F.R. HODSON, D.G.KENDALL and P. TAUTU eds., Mathematics in the Archaeological and the Historical Sciences. Edimburgh University Press: 138-149.
- GOWER, J.C. (1985). "Euclidean distance matrices." In de LEEUW et al. eds., Multidimensional Data Analysis. (1986) DSWO Press, Leiden: 67-80.
- GOWER, J.C. (1982). "Multivariate Analysis: Ordination Multidimensional Scaling and Allied Topics." Séminaire Scientifique OTAN (ASI 82/62) Montreal, Canada, publié in W. LEDERMANN ed. (1984), Handbook of Applicable Mathematics, Vol. VI, Part B: 727-782.
- GREEN, P.E., CARMONE, F.J. (1970). Multidimensional Scaling and Related Techniques in Marketing Analysis. Allyn and Bacon, Boston.
- GREENACRE, M.J. (1978). Some Objective Methods of Graphical Display of a Data Matrix. Special Report, University of South Africa, Pretoria.
- GREGSON, R.A.M. (1975). Psychometrics of Similarity. Academic Press, New York.
- GREUB, W.H. (1967). Multilinear Algebra. Springer Verlag, New York.
- GUTTMAN, L. (1954). "The Principal Components of Scalable Attitudes" In LAZARFELD, P.F., Ed., Mathematical Thinking in the Social Sciences: 216-348.
- GUTTMAN, L. (1965). "A Definition of Dimensionality and Distance for Graphs." In LINGOES et al., eds., Geometric Representations of Relational Data. (1977). Mathesis Press, Ann Arbor: 707-712.
- GUTTMAN, L. (1968). "A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points." Psychometrika 33 : 469-504.
- HANHAM, R.Q. (1976). "Factorial Ecology in Space and Time." Environment and Planning 8 : 823-828.
- HARDWICK, J. (1986). "2R Diagnostics." BMDP Communications. Vol. 19, No 2: 12-14.
- HARSHMAN, R.A. (1972 ). "Determination and Proof of Minimum Uniqueness Conditions for PARAFAC 1." U.C.L.A. Working Papers in Phonetics 22.
- HEISER, W.J. (1981). "Unfolding Analysis of Proximity Data." Ph. D. Thesis, University of Leiden.

- HEISER, W.J. (1985). " Undesired Nonlinearities in Nonlinear Multivariate Analysis." In E. Diday ed., Data Analysis and Informatics. North Holland, Amsterdam .
- HEISER, W.J., de LEEUW, J. (1981). " Multidimensional Mapping of Preference Data." Math.et Sci. Humaines 19 : 39-96.
- HEISER, W.J., MEULMAN, J. (1983). " Analyzing Rectangular Tables by Joint and Constrained Multidimensional Scaling." Journal of Econometrics 22 : 139-167.
- HEISER, W.J., STOOP, I. (1986). " Explicit Smacof Algorithms for Individual Differences Scaling." Recherche Report RR-86-14.
- HENDERSON, H.V. et S.R. SEARLE. (1979). " Vec and Vech Operators for Matrices with Some Uses in Jacobian and Multivariate Statistics." La Revue Canadienne de Statistique Vol 7, No. 1 : 65-81.
- HILL, M.O., GAUCH, H.G. (1980). " Detrended Correspondence Analysis, An Improved Ordination Method." Vegetatio 42: 47-58.
- HOAGLIN, D.C., WELSCH, R.E. (1978). " The Hat Matrix in Regression and ANOVA." The American Statistician 32 , No 1: 17-22.
- HOLLAND, J.L. (1973). Making Individual Choices : A Theory of Careers. Englewood Cliffs. New Jersey : Prentice Hall.
- HOLMAN, E.W. (1972). " The Relation Between Hierarchical and Euclidean Models for Psychological Distances." Psychometrika 37: 417-423.
- HORAN, C.B. (1969). " Multidimensional Scaling: Combining Observations When Individuals Have Different Perceptual Structures." Psychometrika 34 : 139-165.
- HUBERT, L.J. (1982). " Analysis of Variance Procedures based on a Proximity measure Between Subjects." Psychological Bulletin 91, No 2: 424-430.
- INDOWS, T. (1975). " An Application of Multidimensional Scaling to Study of Binocular Visual Space." Paper presented to the U.S. Japan Seminar : Theory, Methods and Applications of Multidimensional Scaling and Related Techniques. San Diego USA.
- INDOWS, T. (1979). " Alleys in Visual Space." J. of Mathematical Psychology Vol 19, No. 3: 221-258.
- IWATSUBO, S. (1983). " The Analytical Solutions of Eigenvalue Problem in the Case of Applying Optimal Scoring Method to Some Types of Data." In DIDAY, E., Ed., INRIA, Rocquencourt: 24-33.
- JAFFRENOU, P.A. (1978). Sur l'Analyse des Familles Finies de Variables Vectorielles. Thèse. Université Claude Bernard. Lyon.
- JONES, L.E. et F.W. YOUNG. (1972). " Structure of the Social Environment : Longitudinal Individual Differences Scaling of an Intact Group." J. of Personality and Social Psychology 24 : 108-121.
- JORESKOG, K.G. (1970). "A General Method for Analysis Covariance Structures." Biometrika 57: 239-251.
- KAMENSKI, V.S. (1978). " Methods and Models of Nonmetric Multidimensional Scaling (Survey)." Avtomatika i Telemekhanika, No 8 (translated): 1212-1243.

- KAMENSKI, V.S. (1979). " A Matrix Approach of Nonmetric Multidimensional Scaling." Avtomatika i Telemekhanika, No 1 (translated): 98-106.
- KHATRI, C.G. (1966). " A Note on a MANOVA Model Applied to Problems in Growth Curve." Ann. Inst. Statist. Math. **18** : 75-86.
- KENDALL, M.G. (1962). Rank Correlation Methods (Third edition). Charles Griffin and Co. London.
- KENDALL, D.G. (1971 a). " Construction of Maps from 'Odds' Bits of Information." Nature **231** : 158-159.
- KENDALL, D.G. (1971 b). " Maps from Marriages : An Application of Nonmetric Multidimensional Scaling to Parish Register Data." Mathematics in Archeological and Historical Sciences. Edinburgh : University of Edingourgh Press.
- KRANE, W.R. (1978). " Least Squares Estimation of Individual Differences in Multidimensional Scaling." British J. of Math. and Stat. Psychology **31**, part 2: 193-208.
- KRANTZ, D.H. (1975). " Color Measurement and Color Theory : I. Representation Theorem for Grassman Structures." J. of Mathematical Psychology **12** : 283-303.
- KRUMHANSL, C.L. (1978). " Concerning the Applicability of Geometric Models to Similarity Data : the Interrelationship Between Similarity and Spatial Density." Psychological Review vol 85, No. 5 : 445-463.
- KRUSKAL, J.B. (1964 a). " Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." Psychometrika **29** : 1-27.
- KRUSKAL, J.B. (1964 b). " Nonmetric Multidimensional Scaling : a Numerical Method." Psychometrika **29** : 115-129.
- KRUSKAL, J.B. (1965). " Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data." J. Roy. Statist. Soc. **B27**: 251-263.
- KRUSKAL, J.B. (1977). " Multidimensional Scaling and Other Methods for Discovering Structures." Statistical Methods for Digital Computers. K. Enslein, A. Ralston et H. S. Wilf eds.. New York : Wiley.
- KRUSKAL, J.B., YOUNG, F.W, SEERY, J.B. (1973). " How to Use KYST, a Very Flexible Program to do Multidimensional Scaling and Unfolding ." Tec. Rep. Bell Labs., Murray Hill, N.J. .
- KRUSKAL, J.B. and WISH, M. (1978). " Multidimensional Scaling." SAGE publ. Beverly Hills.
- LALOUEL, J.M. (1977). " Linkage Mapping from Pairwise Recombination Data." Heredity **38**: 61-67.
- LAWSON, C.L. et R.J. HANSON. (1974). Solving Least Square Problems. Englewood Cliffs. New Jersey : Prentice Hall.
- LE FOLL, Y. (1979). Sur les Propriétés de l'Analyse des Correspondances pour Diverses Formes Complètes de Données. Thèse de 3-ième cycle. Université Pierre et Marie Curie. Paris.

- LEBART, L. (1969). "Analyse Statistique de la Contiguïté." Publication de l'Inst. Stat. Univ. Paris vol XVIII : 81-112.
- LEBART, L. (1978). "Programme d'Agrégation avec Contraintes (C.A.H. contiguïté)." Les Cahiers de l'Analyse des Données, 3 : 275-287.
- LEBART, L. (1984). "Correspondence Analysis of Graph Structures." Bulletin Technique du CESIA vol 2 : 5-19.
- LE CALVE, G. (1985). "Distances à Centre." Statistique et Analyse des Données, Vol 10, No 2: 29-44.
- LECHEVALLIER, Y. (1980). "Classification Sous Contraintes." In E. DIDAY et al., Eds. Optimisation en Classification Automatique. INRIA, PARIS: 677-696.
- LEE, J.C., GEISSER, S. (1972). "Growth Curve Prediction." Sankhya A 34 : 393-412.
- LEE, J.C., GEISSER, S. (1975). "Applications of Growth Curve Prediction." Sankhya A 37: 239-256.
- LEE, S.Y. (1984). "Multidimensional Scaling Models with Inequality and Equality Constraints." Comm. in Statistics (B) 13: 127-140.
- LEE, S.Y., BENTLER, P.M. (1980). "Functional Relations in Multidimensional Scaling." J. of Mathematical and Statist. Psychology, 33: 142-150.
- de LEEUW, J. (1977). "Applications of Convex Analysis to Multidimensional Scaling." Recent Developments in Statistics. J.R. BARRA et al. eds., Amsterdam : North Holland: 133-146.
- de LEEUW, J. (1980). "Majorization Algorithms for Individual Differences in Multidimensional Scaling." Papier présenté au symposium Multidimensional Scaling and Interindividual Differences. at the 22nd International Congress of Psychology. Leipzig.
- de LEEUW, J. et W. HEISER. (1977). "Convergence of Correction Matrix Algorithms for Multidimensional Scaling with Restrictions on the Configuration." Geometric Representation of Relational Data. J. LINGOES, ed., Ann Arbor, Mathesis Press: 735-752.
- de LEEUW, J. et W. HEISER. (1980). "Multidimensional Scaling with Restrictions on the Configuration." Multivariate Analysis V. P.R. Krishnaiah ed., Amsterdam : North Holland, pp 501-522.
- LEFKOVITCH, L.P. (1980). "Conditional Clustering." Biometrics, 36: 91-110.
- LINDMAN, H. et T. CAELLI. (1978). "Constant Curvature Riemannian Scaling." J. of Mathematical Psychology, 17: 89-109.
- LINGOES, J.C. (1971). "Some Boundary Conditions for the Monotone Analysis of Symmetric Matrices." Psychometrika 36 : 195-203.
- LUNEBURG, R.K. (1947). Mathematical Analysis of Binocular Vision. Princeton Univ. Press, Princeton.
- LUNEBURG, R.K. (1950). "The Metric of Binocular Visual Space." J. of the Optical Society of America, 40: 627-642.

- MacKAY, D.B. et J.L. ZINNES. (1981). " Probabilistic Scaling of Spatial Judgments." Geographical Analysis Vol 19, No. 1 : 21-37.
- MAGNUS, J.R. (1988). Linear Structures . Charles Griffin and Co. London.
- MAGNUS, J.R., NEUDECKER, H. (1988). Matrix Differential Calculus With Applications in Statistics and Econometrics . J. Wiley. New York.
- MAILLES, J.P. (1979). Analyse Factorielle des Tableaux de Dissimilarité. Thèse de 3ème cycle. Université Pierre et Marie Curie, Paris, France.
- MEULMAN, J. (1983). " Stability of Constrained Multidimensional Scaling Solutions." Paper presented at the European Meeting of the Psychometric Society. Jouy en Josas, France.
- MEULMAN, J. (1986). A Distance Approach to Nonlinear Multivariate Analysis. DSWO Press. Leiden.
- MEULMAN, J. et J. de LEEUW. (1985). " Relations Between Constrained Multidimensional Scaling and Principal Component Analysis." Paper presented at the Fourth European Meeting of the Psychometric Society. Cambridge, England.
- MEULMAN, J. et W. HEISER. (1984). " Constrained Multidimensional Scaling: More Directions than Dimensions." COMPSTAT. Vienna : Physica Verlag.
- MIZERE, D. (1981). Analyse d'un Cube de Données: Décomposition Tensorielle et Liens entre Procédures de Comparaison de Tableaux Rectangulaires de Données. Thèse de 3ème cycle. Université de GRENOBLE I. Grenoble. France.
- MONLEZUN, C.J., SPEED, F.M. (1980). " The Geometry of Estimation and Hypothesis Testing in the Constrained Linear Model. The Full Rank Case." Comm. in Statistics (A) 9, 2: 213-230.
- NOMA, E. et J. JOHNSON. (1979). " Constrained Nonmetric Multidimensional Scaling." Michigan Mathematical Psychology Program. Tech. Rep. No. 79/4. Ann Arbor, Michigan.
- OBENCHAIN, R.L. (1973). " Affine Multidimensional Scaling." Manuscript non publié. Bell Labs., Murray Hill.
- ORTH, B. (1980). " On the Foundations of Multidimensional Scaling : An Alternative to BEALS, KRANTZ and TVERSKY Approach." Similarity and Choices. E.D. LANTERMANN et H. FEJER, eds., Bern : Hans Huber Publ., pp 54-69.
- PIERRA, D. (1976). Méthodes de décomposition et Croisements d'Algorithmes pour des Problèmes d'Optimisation. Thèse de 3ème cycle. Université de GRENOBLE I. France.
- PIESZKO, H. (1975). " Multidimensional Scaling in Riemannian Space." J. of Mathematical Psychology 12 : 449-477.
- POLIT, M.O. (1986). Une  $\eta$ -ACP d'un Hypercube de données. Thèse de Troisième Cycle. Université de Grenoble II.
- POTGORNY, P. et W.R. GARNER. (1979). " Reaction Time as a Measure of Inter and Intra-Objet Visual Similarity : Letters of the Alphabet." Perception and Psychophysics 26: 37-52.

- POTTHOFF, R., ROY, S.N.. (1964). " A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems." Biometrika 51 : 313-326.
- PRENTICE, I.C. (1977). " Non Metric Ordination Methods in Ecology." J. of Ecology 65: 85-94.
- PRENTICE, I.C. (1980). " Vegetation Analysis and Order Invariant Gradient Models." Vegetatio 42: 27-34.
- QUINQUETON, J. (1981). " Le Concept de Dimension Intrinsèque en Reconnaissance des Formes." Thèse d'Etat, Université Pierre et Marie Curie, Paris VI.
- RAMSAY, J.O. (1969). " Some Statistical Considerations in Multidimensional Scaling." Psychometrika 34 : 167-182.
- RAMSAY, J.O. (1977). " Maximum Likelihood Estimation in Multidimensional Scaling." Psychometrika 42 : 241-266.
- RAMSAY, J.O. (1978). " Confidence Regions for Multidimensional Scaling Analysis" Psychometrika 43 : 145-160.
- RAMSAY, J.O. (1980). " Some Small Sample Results for Maximum Likelihood Estimation in Multidimensional Scaling." Psychometrika 45 : 139-144.
- RAMSAY, J.O. (1982). " Some Statistical Approaches to Multidimensional Scaling data." J. Roy. Statist. Soc. A 145, part 3 : 285-311.
- RAO, C.R. (1964). " The Use and Interpretation of Principal Component Analysis in Applied Research." Sankhya A 26 : 329-358.
- RAO, C.R. (1965). " The Theory of Least Squares When the Parameters are Stochastic and its Application to the Analysis of Growth Curves." Biometrika 52 : 447-458.
- RAO, C.R. (1966). " Covariance Adjustment and Related Problems in Multivariate Analysis." In P.R. KRISHNAIAH, ed. Multivariate Analysis . Academic Press, New York: 87-103.
- RAO, C.R. (1967). " Least Squares Theory Using an Estimated Dispersion Matrix and its Applications to Measurement of Signals. Proceedings of the 5th Berkeley Symposium 1 : 355-372.
- RAO, C.R. (1969). " A Decomposition Theorem for Vector Variables with a Linear Structure." Ann. Math. Statist. 40 : 1845-1849.
- ROBERT, F. (1967). " Calcul du Rapport Maximal de Deux Normes sur  $R^n$ ." RIRO 1 : 97-118.
- ROCKAFELLAR, T. (1970). Convex Analysis. Princeton : University Press.
- SAITO, T. (1983). " A Method of Multidimensional Scaling to Obtain a Sphere Configuration." Hokkaido Behavioral Science Report, Series M, No 4. Sapporo.
- SAITO, T. (1985). " Multidimensional Scaling to Represent Context Effect in Dissimilarity Judgment." Hokkaido Behavioral Science Report, Series M, No 7. Sapporo.
- SAITO, T. (1986). " A Method of Multidimensional Scaling to Explore Complex Aspects in Dissimilarity Judgment." Hokkaido Behavioral Science Report, Series M, No 8. Sapporo.

- de SARBO, W.S., MOHAJAN, V. (1984). " Constrained Classification: The Use of a priori Informations in Cluster Analysis." Psychometrika 49 : 187-215.
- SCHIFFMAN, S.S., REYNOLDS, M.L. et YOUNG, F.W. (1981). Introduction to Multidimensional Scaling: Theory, Methods and Applications. Academic Press, New York.
- SHEPARD, R.N. (1962). " The Analysis of Proximities. Multidimensional Scaling with Unknown Distance Function." Psychometrika 27 : 125-140, et 219-246.
- SHEPARD, R.N. (1974). " Representation of Structure in Similarity Data : Problems and Prospects." Psychometrika 39 : 373-421.
- SHEPARD, R.N., ROMNEY, A.K., NERLOVE, S.B. (1972). Multidimensional Scaling: Theory and Applications In The Behavioral Sciences. 2 Volumes. Seminar Press, New York.
- SHERMAN, C.R. (1972). " Nonmetric Multidimensional Scaling : A Monte Carlo study of the basic parameters." Psychometrika 37 : 323-355.
- SHONEMANN, P.H. (1972). " An Algebraic Solution for a Class of Subjective Metrics Models." Psychometrika 37 : 441-451.
- SIBSON, R. (1978). " Studies in the Robustness of Multidimensional Scaling: Procrustes Statistics." J. Roy. Statist. Soc. b, Vol . 40 : 234-238.
- SIK-YUM LEE et P.M. BENTLER. (1980). " Functional Relations in Multidimensional Scaling." British J. of Math. and Statist. Psychology 33 : 142-150.
- SKARABIS, H. (1978). " Multidimensional Scaling with Prescribed Structures." COMPSTAT Lectures : Lectures in Computational Statistics. H.SKARABIS and P.P.SINT eds. Wien : Physica Verlag.
- TAKANE, Y. (1977). " Statistical Procedures for Non Metric Multidimensional Scaling." Ph. D. Thesis. Univ. of North Carolina, Chapel Hill. Univ. Microfilm International 78-7169.
- TAKANE, Y. (1981). " Statistical Procedures for Non Metric Multidimensional Scaling." J. of Ecology 69: 135-152.
- TAKANE, Y. et J. SERGENT. (1983). " Multidimensional Scaling Models for Reaction Times and Some Different Models." Psychometrika 48 : 393-423.
- TAKANE, Y., YOUNG, F.W., de LEEUW, J. (1977). " Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features." Psychometrika 42 : 7-68.
- TEBBOUNE, F.O. (1981). Contribution aux Analyses en Composantes Principales Sous Contraintes Linéaires. Thèse de 3ème cycle. Université Paul Sabatier, Toulouse, France.
- TEN BERGE, J.M.F. (1984). " A Join Treatment of the Varimax Rotation and the Problem of Diagonalizing Symmetric Matrices Simultaneously in the Least Squares Sense." Psychometrika 49 : 347-358.
- TENENHAUS, M. (1977). " Analyse en Composantes Principales d'un Ensemble de Variables Nominales et Numériques." Rev. de Statist. Appliq., 25: 39-56.

- TENENHAUS, M. (1987). " Canonical Analysis of Two Convex Polyhedral Cones and Applications." Soumis à publication: 40p. .
- TER BRAAK, C.J.F. (1985). " Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model." Biometrics 49 : 347-358.
- TER BRAAK, C.J.F. (1986). " Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis." Ecology 67 (5): 1167-1179.
- TIMM, N.H. (1979). " Multivariate Analysis of Variance of Repeated Measurements." In P.R. KRISHNAIAH, ed., Handbook of Statistics , Vol. I: 41-88.
- TVERSKY, A. et D.H. KRANTZ. (1970). " The Dimensional Representation and Metric Structure of Similarity Data." J. of Mathematical Psychology 7 : 572-596.
- TORGERSON, W.S. (1958). Theory and Methods of Scaling. J. Wiley and Sons, New York.
- VAN DEN WOLLENBERG, A.L. (1977). " Redundancy Analysis. An Alternative for Canonical Analysis." Psychometrika 42 : 207-219.
- VAN DER BURG, E., de LEEUW, J. (1983). " Nonlinear Canonical Correlation." British J. of Math. and Statist. Psychology 36: 54-80.
- VELLEMAN, P.F., WELSCH, R.E. (1981). " Efficient Computing of Regression Diagnostics." The American Statistician 35 , No 4: 234-242.
- VERDEGALL, R. (1986). " Overals." Tech. Rep., Dept. Data Theory. University of Leiden.
- WALKER, J.J. (1976). " Some Statistical Procedures Based on Distances." Institute of Statistics Mimeo Series No. 1096. University of North Carolina, Chapel Hill, USA.
- WEEKS, D.G. et P.M. BENTLER. (1982). " Restricted Multidimensional Scaling Models for Asymmetric Proximities." Psychometrika 47 : 201-208.
- WHITTAKER, R.H. (1967). " Gradient Analysis of Vegetation." Biological Review 42: 207-264.
- WINSBERG, S. et J.D. CARROLL. (1984). " A Nonmetric Method for a Multidimensional Scaling Model Postulating Common and Specific Dimensions." Paper presented at the Psychometric Society Meeting.
- WINSBERG, S. et RAMSAY, J.O. (1983). " Monotone Spline Transformations for Dimension Reduction." Psychometrika 48 : 575-595.
- YOUNG, F.W., de LEEUW, J., TAKANE, Y. (1976). " The Principal Components of Mixed Measurement level Multivariate Data: An Alternating Least Squares Method with Optimal Scaling Features." Psychometrika 43 : 279-281.
- YOUNG, F.W., TAKANE, Y., de LEEUW, J. (1978). " Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features." Psychometrika 41 : 505-529.
- YOUNG, G. et A.S. HOUSEHOLDER. (1938). " Discussion of Set of Points in Terms of their Mutual Distances." Psychometrika 3 : 19-22.
- ZADEH, L.A. (1971). " Similarity Relations and Fuzzy Orderings." Information Sciences 3 : 177-200.

- ZINNES, J.L. et R.A. GRIGGS. (1974). Probabilistic Multidimensional Unfolding Analysis." Psychometrika 39 : 30-50.
- ZINNES, J.L. et D.B. McKAY. (1981). " Multidimensional Scaling Models: The Other Side." In I. BORG (ed.) Multidimensional Data Representations: When and Why. Mathesis Press, Ann Arbor .
- ZINNES, J.L. et R.D. WOLFF. (1977). " Single and Multidimensional Same Different Judgments." J. of Mathematical Psychology 16 : 30-50.



**PARTIE C**

**A TOOLKIT**

**FOR**

**MULTIDIMENSIONAL SCALING OF**

**DISSIMILARITY DATA**

- Chapitre I            INTRODUCTION**
- Chapitre II            EUCLIDEAN EMBEDDING OF DISSIMILARITY**  
**DATA**
- Chapitre III            EVALUATION OF THE DIFFERENTIAL QUALITY**  
**OF THE REPRESENTATION OF OBJETS**
- Chapitre IV            THE CASE OF ORDINAL DATA**



# A TOOLKIT FOR MULTIDIMENSIONAL SCALING OF DISSIMILARITY DATA

G. d'AUBIGNY

IUT 2, Dpt STATISTIQUE  
UNIVERSITE DES SCIENCES SOCIALES DE GRENOBLE  
47X - 38041 GRENOBLE CEDEX

## 1. INTRODUCTION

During the two preceding decades, the Multidimensional Scaling (MDS) methodology mainly focused on the elaboration of new methods, aimed at the data analytic treatment of ever more varied kinds of relational data, and the improvement of the numerical treatment of the difficult numerical problems encountered.

The applications to psychophysics and psychometry gave the lead in this active research area. Now the usefulness of those methods seems fully recognized in all the social and behavioral sciences. As shown in bibliographies by BICK and al. [1] and NISHISATO [2], that covers applications up to 1978 and the review paper by de LEEUW and HEISER [3] which gives more recent references. A helpful classification of data structure resorting to MDS methods and a complete presentation of the state of the art may be found in CARROLL and ARABIE [4].

The wide acceptance of the MDS methods as a routine part of the statistical knowledge in America is illustrated by their present integration to the leading professional statistical packages on the market (e.g. ALSCAL in SPSS<sup>X</sup>, SMACOF in SAS in next future,...). Their more cautious reception in France seems to result from the lack of elaborate tools needed for the evaluation of the quality of summaries resulting from analyses, as well as for assistance in the interpretation steps, a major contribution of the french statisticians community to the multivariate data analysis methodology. These tools exist in the domain of linear data analysis (see e.g. LEBART and al. [5]) and in the cluster analysis (c.f. JAMBU and LEBEAUX [6]), and somehow bridle the hasty and subjective readings of results. The situation is less satisfactory in MDS, since most existing programs give quite few statistical informations that could help a secure reading of the resulting mappings.

Our present contribution to the field mainly focuses on those topics. It is restrained to multidimensional scaling of one dissimilarity table, in both the numerical and the ordinal cases, so called metric and non metric MDS respectively in the psychometric litterature.

First a formal description of the necessary algebraic and geometrical foundations of the distance based methods is given. The next section provides a consequent guide for the prescription of interpretative rules and assistance. Here, without loss of generality, the discussion is limited to the numerical MDS setting. In fact the ordinal case follows directly when interpreted as the determination of an optimal scoring of the original dissimilarities, in such a way that the discrepancy between rescaled dissimilarities, or disparities, and Euclidean distances be minimal, from a numerical point of view. The next section focuses on the scoring step of the ordinal MDS methods. The savety devices against degeneracy and the determination of influential (pairs of) statistical units as well as stability problem are discussed.

## 2. EUCLIDEAN EMBEDDING OF DISSIMILARITY DATA

The MDS problem deals with a finite set  $I$  of statistical units, or objets. We consider the semi-metric  $d$  defined on  $I \times I$  and valued in  $\mathbb{R}^+$ , resulting from the observation of dissimilarities between pairs of elements of  $I$ . Now, let suppose that we are given a one-to-one application of  $I$  onto  $[n] = \{1, 2, \dots, n\}$ . The dissimilarities  $\delta_{ij}$  form a symmetrical matrix of order  $n$  with zero diagonal. The quantities  $\delta_{ij}$  are either directly observed or derived from more basic data, e.g. with help of one of the many dissimilarity coefficients in use in the taxonomic litterature (cf. CHANDON and PINSON [7], FICHET [8], GOWER [9]). In either case,  $(I, \delta_I)$  is not always embeddable in a Euclidean space, since one is scarcely able to exhibit a set of  $n$  points in a  $r$ -dimensional Euclidean space that generates all the  $n(n-1)/2$  values  $\delta_{ij}$ . When this Euclidean embedding in a vector space  $E_r$  is possible, we consider the  $Z$  matrix collecting the coordinates of the representative points in  $E_r$  ( $r < n-1$ ) and

$$\delta_{ij} = \delta_{ij}(Z) = \delta(z_i - z_j) \quad (1)$$

here  $\delta$  is a Euclidean distance function defined in  $E_r$  and  $z_i$  is the vector representing the unit  $i$  in  $E_r$ . Then a data reduction to a lower dimensional Euclidean spaces may be sought. One approximates the observed Euclidean distances  $\delta_{ij} = \delta_{ij}(Z)$  by fitting values  $d_{ij}(X)$  that minimize some badness of fit criterion such as one representative of the following class, proposed by OBENCHAIN [10]:

$$\sum_{i < j} q_{ij} |\delta_{ij}^a - d_{ij}^a(X)|^b \quad a, b \in \mathbb{R}^+ \quad (2)$$

In the ordinal case, this optimization problem is completed by the estimation of a monotone transformation of the data having some optimality property in the sens of (2). This point is treated later. Here, the presentation is constrained to the numerical case.

Two traditions need be differentiated. The first approach is based on the work of YOUNG and HOUSEHOLDER [11] and TORGERSON [12]. It received a comprehensive treatment in GOWER [13,14] who calls it the Principal Coordinates Analysis (PCA). It considers the triple  $(I, \delta_I, P_I)$  where  $P_I$  is a known weight function on  $I$ , satisfying

$$p_i > 0 \quad (\forall i, i \in I), \quad \sum_i p_i = 1$$

and makes use of the classical relation linking scalar products to Euclidean distances (CAILLET et PAGES [15]):

$$w_{il} = \frac{1}{2} (\delta_{i.}^2 + \delta_{.l}^2 - \delta_{il}^2 - \delta_{..}^2) \quad (3)$$

where

$$\delta_{i.} = \sum_{j=1}^n p_j \delta_{ij}^2 \quad \delta_{..} = \sum_{i=1}^n p_i \delta_{i.}^2$$

in a way to transform the problem in a scalar products approximation one. This model is often called the error free model, since (3) is an algebraic equation that does not take into account potential measurement errors. Moreover,  $w_{il}$  in (3) is a scalar product and the matrix  $W = (w_{il})$  is gramian only when  $\delta$  is Euclidean satisfying (1). In that case, we consider the  $n$ -dimensional free vector space  $F$  equipped with the arbitrary basis  $\{f_i, i \in [n]\}$ . The diagonal matrix  $N = D_p$  induced by the weights  $p_i$  defines a metric on  $F$  and the  $(i,l)$ -th element  $w_{il}$  of the matrix

$$W = Q_1 H {}^t Q_1 \quad (3bis)$$

is given by equation (3). Here the matrix  $H$  is defined in terms of Hadamard product  $\Delta * \Delta$  as:

$$H = -\frac{1}{2} \Delta * \Delta = \left( -\frac{1}{2} \delta_{il}^2 \right)$$

and  $Q_1$  is the  $D_p$ -orthogonal projector on the subspace  $F_1 \subset F$  containing the centered vectors in  $F$ .  $F_1$  is  $D_p$ -orthogonal to the constant vector  ${}^t 1 = (1, \dots, 1)$ , so:

$$Q_1 = I - {}^t 1 D_p$$

Consider  $F^*$  the algebraic dual space of  $F$ , equipped with the dual basis  $\{f_i^*, i \in [n]\}$ . One

observes (CAILLEZ and PAGES [15]) that  $w_{ii}$  given in (3) satisfies

$$w_{ii} = w(f^{i*}, f^{i*}) = \frac{1}{2} [w(f^{i*}) + w(f^{i*}) - w(f^{i*} - f^{i*})] \quad (3ter)$$

So  $W$  is the matrix of the quadratic form  $w(\cdot)$  on  $F_1^*$  in these dual bases. When  $\Delta$  is Euclidean, then

$$w(f^{i*} - f^{j*}) = \delta_{ii}^2(Z) = \delta_{ii}^2$$

A configuration representing  $I$  in  $F$ , in the sens that the inter-point distances describe the dissimilarities  $\delta_{ii}$  is obtained from the eigenvalue decomposition of the operator  $WD_p$ . Then the determination of summary subspaces mimics the Principal Components Analysis methodology. Both the nesting of principal subspaces and the contracting property of projectors constraint the solution to satisfy an internal approximation property (cf. d' AUBIGNY [20], MEULMAN [16]):

$$d_{ii}^2(X_1) \leq d_{ii}^2(X_2) \leq \dots \leq d_{ii}^2(X_r) \leq \dots \leq d_{ii}^2(X_p) = \delta_{ii}^2(Z) \quad (4)$$

where  $X_r$  is the configuration matrix in  $r$ -dimensions and  $p$  is the rank of  $Z$ . Moreover one criterion minimized by the configuration  $X$  can be deduced from (3bis) and the usual least squares property of principal components (cf. d'AUBIGNY [17]):

$$\text{STRAIN} = \text{Trace}((W_Z - W_X)N(W_Z - W_X)N) = \|w_Z - w_X\|_{N \otimes N}^2 \quad (5)$$

$$= \frac{1}{4} \|\delta^2(Z) - d^2(X)\|_{N_1 \otimes N_1}^2 \quad (5-bis)$$

where  $N$  is here  $D_p$  and  $N_1 = Q_1 N^t Q_1 = Q_1 D_p^t Q_1 = Q_1 D_p$ .  $w_X$  (resp.  $d^2(X)$ ) is the element of  $F_1 \otimes F_1$  associated with  $W_X$  (resp.  $d^2(X)$ ) by the Vec transformation that associates to a matrix its corresponding tensor in  $F_1 \otimes F_1$ ,  $w_X = \text{Vec}(W_X)$ . This component analysis is not the only possibility since one might extend the GUTTMAN Partial Image Analysis or any Common Factor Analysis method to the analysis of dissimilarity data (c.f. d'AUBIGNY [17]). The two first methods have in common their error free algebraic foundations given above.

The second approach directly approximates the dissimilarities  $\delta_{ii}$  and so it takes into account possible measurement errors. The approximation is understood in the sens of some special case of the criterion (2) characterized by one choice of  $a$  and  $b$ . They are based on the same explicit error theory and lead to a two-sided approximation property in contradiction to the inequations (4). Moreover, the extension to ordinal dissimilarity indices is straightforward then. In fact the work of SHEPARD [18], KRUSKAL [19] and GUTTMAN [47] initiated this approach in the ordinal setting. As shown in

d'AUBIGNY [20, 21], central to a common algebraic formulation of the various existing solutions is the geometric interpretation of the result of COXETER [22] :

**Lemma :**

Let  $q$  be a real function on  $I \times I$  satisfying  $q_{il} \geq 0$  and  $q_{il} = q_{li}$  for any  $(i,l)$  in  $[n] \times [n]$ . Then for any vector  $x$  in  $F_1$

$$\alpha = \sum_i \sum_l q_{il} (x_i - x_l)^2 = 2 \left\{ \sum_i q_{ii} x_i^2 - \sum_i \sum_{k \neq i} q_{ik} x_i x_k \right\} \quad (6)$$

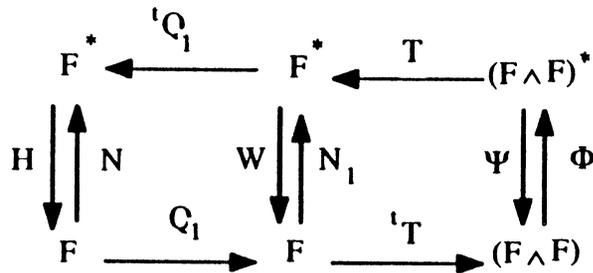
where

$$q_{ii} = \sum_{k \neq i} q_{ik}$$

So we can write  $\alpha = 2 \|x\|_{N_Q}^2$  for  $x \in F_1$ , where the matrix  $N_Q = D_Q - Q$  defines a (semi-) definite positive quadratic form on  $F_1$ , that is what is called with some abuse a (semi-) metric by :

$$n_{il} = \begin{cases} -q_{il} & \text{if } i \neq l \\ q_{ii} & \text{if } i = l \end{cases}$$

From a mathematical point of view, the Euclidean analysis of the triple  $(I, \delta_I, Q_{I \times I})$ , where  $Q$  is a weight function on  $I \times I$ , uses the fact that the distance associated with any gauge  $g$  defined on  $F$  by  $d_{il}(X) = g(x_i - x_l)$ , implies the skew vector  $x_i - x_l$  which is an element of the exterior product of  $F$  by itself, denoted by  $F \wedge F$ . The situation is summarized by the constituents of the following duality diagram :



The matrix  $\Phi = D_Q$  is diagonal with generic term  $q_{il}$  and  $T$  is the matrix of the restriction to  $F \wedge F$  of the simple contrasts operator in the bases discussed below

$${}^t T = S_n \otimes I_n \quad S_n = (I_n : -I_n)$$

$F \wedge F$  is a linear subspace of  $F \otimes F$  with dimensionality  $n^* = n(n-1) / 2$ . While not essential to the presentation, the restriction to  $F \wedge F$  takes into account the redundancies induced by the condition of symmetry  $\delta_{il} = \delta_{li}$ . Using classical duality arguments, one shows that  $T$  defines an isometry between  $(F \wedge F)^*$  and  $F_1^*$  characterized by

$$T (f^* \wedge f^*) = f^* - f^*$$

$\{(f^* \wedge f^*), i < l, (i, l) \in I \times I\}$  form a natural basis of  $F^* \wedge F^*$ , associated with the natural basis  $\{f^*, i \in [n]\}$  of  $F^*$ . This implies that for any  $x$  in  $F_1$  :

$${}^t T x = (x_i - x_l)_{i < l} \quad \text{if } x = (x_i)_{i \in I}$$

So,  ${}^t T x$  defines an skew-symmetric matrix of order  $n$ , That is an element of  $F \wedge F$  and of rank 1. Now, let us introduce the following notations:

$$Q = (q_{il}), \quad D_q = \text{diag}(q_i), \quad D_Q = \text{diag}(q_{il})$$

$$N_Q = D_q - Q = T D_Q {}^t T \quad (7)$$

$N_Q$  is shown to define a (semi-)metric of rank  $(n-1)$  on  $F_1$ , when  $Q$  satisfies the hypotheses of the Coxeter lemma. Moreover it may be interpreted as the covariance matrix of a multinomial distribution when  $q_{il} = p_i \times p_l$  for all  $(i, l)$ .  $F \wedge F$  is a natural space to consider when working on skew-symmetric scorings. It has a long history in statistics, since the work of DANIELS [23] and followers on association indices may be shown to receive a complete geometric treatment in  $F \wedge F$ . More is said on its interest in statistics in d'AUBIGNY [20]. Here, the well known fact that if  $d(X)$  is an Euclidean distance, then the gradient vector of  $d(X)$  belongs to  $F \wedge F$  is central to our formalization.

The analysis of the triple  $(I, \delta_I, Q_{I \times I})$ , proposed by KRUSKAL [19] minimizes the STRESS index, written in  $F \wedge F$  :

$$\text{STRESS} = \| \delta - d(X) \|_{D_Q}^2 \quad (8)$$

The associated MDS problem received a solution in terms of gradient methods. Since the differentiability of STRESS is not secured as soon as one  $d_{il}(X)$  is zero, a more satisfactory algorithm, using majorization arguments was proposed by de LEEUW and HEISER [3], de LEEUW [24]. A criterion more similar to the STRAIN is

$$\text{SSTRESS} = \| \delta^2 - d^2(X) \|_{D_Q}^2 \quad (9)$$

which can be written in case of an Euclidean dissimilarity

$$\text{SSTRESS} = \text{Trace}((W_Z - W_X)N_Q(W_Z - W_X)Q_1) = \| w_Z - w_X \|_{N_Q \otimes Q_1}^2 \quad (9\text{bis})$$

The change of Euclidean geometry in  $F \wedge F$ , observable by comparison to equation (5) has many consequences on the properties of the solution. The numerical resolution uses either an alternating least squares algorithm (TAKANE and al. [25]), or an iterated singular value decomposition of a diagonally perturbed  $WD_P$  operator (BROWNE [26]). The formulation of this second approach is closer to the PCA methodology and seems

promising on theoretical grounds. Comparing (5bis) to (9bis) shows the greater complexity of the optimization problem encountered when minimizing the SSTRESS function. This results from the asymmetrical structure of the metric  $N_Q \otimes Q_1$  that imposes an iteration of perturbed Principal Coordinates Analyses.

### 3. EVALUATION AND INTERPRETATION OF THE RESULTS

#### 3.1. The evaluation of the differential quality of the representation of objects:

The evaluation of the results of the MDS algorithms imposes some cautious steps. Interpreting distances informations is less easy and less intuitive than interpreting scalar products informations. We showed in the preceding section that one might express the problem in  $F \wedge F$ , in the form of a weighted least squares fit associated with the additive decomposition

$$\delta_{il} = d_{il}(X) + \epsilon_{il} \quad (i,l) \in I \times I \quad (10)$$

which yields bilateral approximations of the dissimilarities. In (10), the residuals  $\epsilon_{il}$  may be correlated or at least associated and a closer examination leads to a decomposition as the sum of two components

$$\epsilon_{il} = (\delta_{il}(Z) - d_{il}(X)) + \tau_{il} \quad (10bis)$$

The last term in (10bis) measures the strength of non Euclideanicity resulting from sources of variations in  $\Delta$ , while the first term quantifies the lack of fit resulting from a reduction of dimensionality. In Principal Coordinates Analysis, the common practice truly confuses both sources of missfit, since the underlying descriptive model considers that  $\tau_{il}$  is zero (cf. McDONALD [27]). This led to misconceptions about negative eigenvalues of  $WD_p$ , see d'AUBIGNY [20]. In a way to grasp on that point, let us consider a perfect embedding of elements of  $I$  in a  $r$  dimensional vector space  $E_r$ . Adding to this constellation of points a new point  $M_{n+1}$ , taken outside of  $I$ , generally imposes one more dimension. So consider  $E_{r+1} = E_r \oplus \Delta_u$  where  $u$  completes the basis of  $E_r$  in such a way that  $\Delta_u$  is the unicity axis of  $M_{n+1}$ , and write the decomposition of  $M_{n+1}$  on the completed basis as

$$x = \sum_{s=1}^r x_{n+1}^s e_s + x_{n+1}^{s+1} u$$

the dissimilarity of  $M_{n+1}$  to any element  $i$  of  $I$  is measured on Euclidean grounds by

$$\delta_{i(n+1)}^2(Z) = \sum_{s=1}^r (x_i^s - x_{n+1}^s)^2 + [x_{n+1}^{r+1}]^2 = d_{i(n+1)}^2(X) + [x_{n+1}^{r+1}]^2$$

where we note  $d_{i(n+1)}(X)$  the distance of the projection of  $M_{n+1}$  on  $E_r$  to points  $M_i$  ( $i \in I$ ) in  $E_r$ , and  $\delta_{i(n+1)}(Z)$  this distance evaluated in  $E_{r+1}$ , with  $Z = (X, x)$ . So the added point satisfies :

$$[x_{n+1}^{r+1}]^2 = \delta_{i(n+1)}^2(Z) - d_{i(n+1)}^2(X) \quad (11)$$

This coordinate is either real or complex depending on the sign of the right term in equation (11), whatever its modulus. The bias in the approximation of the  $\delta_{i(n+1)}(Z)$  by  $d_{i(n+1)}(X)$  is the same in both cases. One adaptation of the PCA methodology to this situation consists to solve the so called additive constant problem, and leads to a shift of the eigenvalues by a term greater than the absolute value of the smallest negative eigenvalue. This is not a satisfactory device on some respect since then, the reduction of dimensionality step of the analysis neglects the dimensions associated with the smallest modified eigenvalues. These are the imaginary ones, so that it results in inequation (4). A better way out (d'AUBIGNY [20]) would be to use the spectral decomposition of the operator  $WD_pWD_p$ , since the existence of imaginary coordinates has troublesome consequences on their statistical meaning as well as on the applicability of the interpretative tools usually used in Principal Components Analysis, which are based on analogies to mechanics and are no more valid in that situation. For example, the index measuring the global quality of the solution  $X$ , as

$$\frac{\|X\|_{N \otimes M}^2}{\|Z\|_{N \otimes M}^2} = 1 - \frac{\sum_i \sum_{i < l} q_{il} (\delta_{il}^2(Z) - d_{il}^2(X))}{\sum_i \sum_{i < l} q_{il} \delta_{il}^2(Z)}$$

is known to be derived from eigenvalues and so is clearly dependent on the sign problem in (11). In the same way, the evaluation of the contribution of an axis to the inertia of a point depends only on interpoints distances and weights. So it suffers from algebraic compensations between coordinates on real and complex axes, governed by the Sylvester law of inertia.

The additive decomposition of any vector  $x_i$  on coordinate axes induces via the matrix  ${}^tT$  a corresponding additive decomposition of the coordinates of vector  $y = (x_i - x_j)_{ij}$ , the norm of which yields the Euclidean distance  $d_{ij}(X)$ . Thus any interpretative tool aimed at the evaluation of the quality of the representation of  $x_i$  in  $F$  has a direct counterpart for the evaluation of the quality of the reconstitution of dissimilarities. Both suffer from the same weaknesses in the case of the application of a Principal Coordinate Analysis method to

non Euclidean dissimilarities.

Thus, while natural in MDS, the existence of imaginary axes implies the development of new interpretative assistance tools, when the diagonalization of  $WD_pWD_p$  is not used. In that case, the eigenvalues are squared and the associated eigenvectors are unchanged, leading either to direct or to inverse factors, in a way similar to the results of correspondence analysis when applied to a symmetric contingency table (Benzécri [29]). Thus the application of this kind of Principal Coordinate Analysis to  $(I, \delta_I, P_I)$  without data reduction step, results in a permutation of the axes, and the inner approximation constraint (4) is no more satisfied when some negative eigenvalues of  $WD_p$  exist. The MDS methods analysing the triple  $(I, \delta_I, Q_{I \times I})$ , with  $Q$  not diagonal, are free of such difficulties.

In this last case, the minimum of the objective function yields a simple way to compare the qualities of the representation of dissimilarities by the distances produced by the algorithm. Let us illustrate that point for the stress function. Then the minimum can be written as a sum of positive numbers

$$\text{STRESS}^* = \sum_{i < j}^n s_{ij}, \text{ with } s_{ij} = q_{ij} (\delta_{ij} - d_{ij}(X))^2$$

that yields the indices of quality of representation

$$r_{ij} = \frac{s_{ij}}{\text{STRESS}^*}$$

Moreover, the interpretation of the problem as the analysis of a weighted graph, with weights given by the dissimilarity function leads to an evaluation of the quality of the representation of any statistical unit by the half-outer degree  $s_i$  of the vertice  $i$ , given by

$$s_i = \sum_{j=1}^n \frac{q_{ij}}{q_i} (d_{ij} - d_{ij}(X))^2, \text{ with } q_i = \sum_{j=1}^n q_{ij}$$

which is better interpreted in the normalized form

$$r_i = \frac{s_i}{2 \times \text{STRESS}^*}$$

D'AUBIGNY [20] shows that the apparent quality of the MDS methods results from the fact that all of them use iterated transformations of  $Q$  matrices (see also TAKANE [28]). By the COXETER's lemma, the associated  $N_Q$  matrix has a dominant diagonal, and consequently only positive eigenvalues. The main consequence of this correcting property is the recovered availability of interpretative aids developed in Principal Components Analysis. A cautious reading of mappings, differentiating well represented points from others is possible, with help of classical indices quantifying the quality of the

representation of points in  $F$  on geometric grounds. Nevertheless, the original difficulty has been shifted to the equally complex one of determining the dimensionality of a summary configuration. The elaboration of a truly satisfactory methodology for that problem remains partly an open one. The following example of a simple graph associated with a regular pentagone with the similarity matrix  $S$ , shows the nature of difficulties met in both treatments. In the following example, the column named Torgerson gives the value of recovered eigenvalues of  $W$ , since  $D_p$  is taken equal to the identity here in PCA, while the one named Coxeter results from the spectral decomposition of the matrix  $N_Q$ .

|           | Eigenvalues | Torgerson | Coxeter |
|-----------|-------------|-----------|---------|
| $S =$     | $\lambda_1$ | .81       | 3.62    |
| 0 1 0 0 1 | $\lambda_2$ | .81       | 3.62    |
| 1 0 1 0 0 | $\lambda_3$ | .00       | 1.38    |
| 0 1 0 1 0 | $\lambda_4$ | -.31      | 1.38    |
| 0 0 1 0 1 | $\lambda_5$ | -.31      | .00     |
| 1 0 0 1 0 |             |           |         |

The incidence matrix  $S$  is associated with a dissimilarity matrix  $\Delta$  which evaluates the dissimilarity  $\delta_{ij}$  as the smallest number of edges of a path linking  $i$  and  $j$ . Since the diagonal terms of  $N_Q$  are all equal, one finds a perfect correspondence ( $T_1 = C_1$ ,  $T_2 = C_2$ ,  $T_4 = C_4$ ,  $T_5 = C_3$ ) between both sets of eigenfactors, aside from normalization problems, but their order is not identical. The planes  $(T_1, T_2)$  and  $(T_4, T_5)$  are in a one to one correspondence by reflexions and rotations. This results from the strong symmetry of the configuration. BENZECRI [29] gives an analytic expression of the eigen elements for such reference cases. Besides his arguments describes the functional relation linking the dimensions 4 and 5 with the dimensions 1 and 2 determined by the PCA, which adds complexity in the interpretative step.

This counterexample should appear paradoxical to some readers. In fact, the small number of points (5 here) has been chosen in a way to secure the observed shift of the true configuration on the plan generated by the fourth and the fifth eigenvectors, a well known weakness of the method. The treatment of an analogous cyclic graph with more vertices (e.g. 16) should produce a better situation, since the solution would be well determined and the original planar graph would then be correctly recovered in the plane (1,2).

When the diagonal terms of  $N_Q$  are not equal, the strong correspondence noticed here is lost. A canonical correlation analysis of the correlation structure of the factors derived from both approaches, usually shows very complex patterns. So both methods determine different summary representations, consequent to the data reduction step, when we impose  $r < \text{rank}(Z)$ .

### 3.2. Evaluation of the adequacy of the retained dimensionality

The evaluation of the dimensionality of the configuration space  $E_r$  is usually based on a scree test. One plots the optimal value of the criterion as a function of the number of estimated parameters, that is  $nr$  in the case of unconstrained MDS configurations in  $r$  dimensions. This evaluation is crossvalidated by the consultation of the so called Shepard diagram representing the pairs  $(\delta_{ij}, d_{ij}(X))$ . The descriptive model (10) assumes a linear fit of slope one. One often observes a bias corresponding to an overestimation of great dissimilarities and an underestimation of small ones. We argue here that this bias reflects in many situations an underestimation of dimensionality. The argument presented here for the STRESS (see d'AUBIGNY [17] for a direct equivalent in the case of SSTRESS). The minimum of the index STRESS in function of  $X$ , is achieved when the normal equations vanish. Then a simple calculation (cf. d'AUBIGNY [30], GOWER [14]) validates the two consequent equalities

$$\begin{aligned} \sum_i \sum_{i < j} q_{ij} (d_{ij}(X) - \delta_{ij}) \delta_{ij} &= 0 \\ \sum_i \sum_{i < j} q_{ij} (d_{ij}(X) - \delta_{ij})^2 &= \sum_i \sum_{i < j} q_{ij} d_{ij}(X)^2 - \sum_i \sum_{i < j} q_{ij} \delta_{ij}^2 \end{aligned} \quad (12)$$

so the optimum corresponds to the value of STRESS FORM 1:

$$S^* = \frac{\sum_i \sum_{i < j} q_{ij} (d_{ij}(X) - \delta_{ij})^2}{\sum_i \sum_{i < j} q_{ij} d_{ij}(X)^2} = 1 - \frac{\sum_i \sum_{i < j} q_{ij} \delta_{ij}^2}{\sum_i \sum_{i < j} q_{ij} d_{ij}(X)^2}$$

developping the left term shows by simple algebra, that  $S^*$  differs from the observed value of STRESS at the numerical optimum by the quantity

$$2 \left[ 1 - \frac{\sum_i \sum_{i < j} q_{ij} \delta_{ij} d_{ij}(X)}{\sum_i \sum_{i < j} q_{ij} d_{ij}(X)^2} \right] = 2[1 - b(X)]$$

This bias term depends on the estimated slope  $b(X)$  of the line adjusted to the SHEPARD diagram, in a weighted least squares sens. Since both statistics STRESS and  $b(X)$  are derived from the solution, the bias in the resulting configuration can be evaluated either visually or numerically. This slope corresponds to an increase of the lack of fit, measured e.g. in terms of STRESS. As a result of simulation studies, the bias may be attributed either to convergency problems of the algorithm, or to the impossibility to obtain a satisfactory fit in so few dimensions. Adapting the argument to STRESS FORM 2 or to SSTRESS is straightforward.

### 3.3. Interpretation of mappings

We discuss here neither the interpretative aids common in MDS nor the tools coming from multivariate linear data analysis commented upon in section 3.1 above. Among them, the projection of supplementary statistical units and of illustrative variates techniques are the most useful, see d'AUBIGNY [20]. Their applications to the MDS context present in fact few new technical problems. Besides we do agree with CARROLL and PRUZANSKI [31] on the advantages of additive trees over Euclidean representations as a descriptive model in presence of clustered units. Our interest here focuses on several techniques that seem promising in the elicitation of non linear structures describable in terms of non linear manifolds, such as simplexes (or horseshoes), circumplexes, etc... On that respect, the manipulation of the metric spaces is known to be only incompletely taken into account by algebraic tools, which are better suited to the normed linear spaces structures than to the description of metric properties. The elicitation and the description of non linear structures in some Euclidean space usually requires a linear dimensionality higher than their topological one. It is attainable with help of one appropriate constrained MDS method (cf. BENTLER et al. [32], [33], LINGOES and BORG [34], d'AUBIGNY [35] and references therein).

The use of interpoint distances in the quantification of curvature of the latent differential manifold is standard in differential geometry. So the detection, and later a first step to the recovery, of non linear structures, with help of elementary discrete analogues of the geometry of geodesics is promising for MDS methods. This point is partly exemplified by KRUSKAL and WISH [36]. These authors use a graphical display designed for the evaluation of the local recovery of dissimilarities by distances resulting from an ordinal MDS program. The accuracy of the recovery of the ordering of dissimilarities is appreciated by drawing the connectivity graphs associated with the following additive decomposition of  $\Delta$  induced by the ordinal information on  $I \times I$  contained in  $\Delta$ :

$$\Delta = \sum_{k=1}^K \alpha_k \Delta^k \quad (13)$$

where

$$\delta_{il}^k = \begin{cases} 1 & \text{if } \delta_{il} = \alpha_k \\ 0 & \text{elsewhere} \end{cases}$$

The thresholds  $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_K$  define a family of nested spanning subgraphs  $G_k = (I, U^k)$  of the complete graph  $G = (I, U)$ , the incidence matrix of which is  $\Delta^k$ : the edge  $u_{il}^k$  exists if  $\delta_{il}^k$  equals 1. Drawing the corresponding segments on the mapping issued from a MDS algorithm gives a visual idea of the respect of dissimilarities by the recovered distances in terms of concordance of the order relations induced by  $\Delta$  and

$D(X)$ . This idea is reminiscent to those underlying the definition of the KENDALL's  $\tau$  index of monotone association, and it can be quantified with help of generalized correlation coefficients, in a way similar to the approach of HUBERT and SUBKOVIK [37] and FRIEDMAN and RAFSKY [38]. Let us define

$$\Gamma_1(\Delta, D(X)) = \sum_i \sum_l a_{il} b_{il}$$

where  $a_{il}$  (resp.  $b_{il}$ ) is a scoring of  $\delta_{il}$  (resp.  $d_{il}(X)$ ).  $\Gamma_1(\Delta, D(X))$  offers an evaluation tool of the adequacy of the configuration, in terms of permutation tests of the association of  $\Delta$  and  $D(X)$ , whose statistical properties is studied in [38].

The interest of this tool in MDS, relies on the evaluation of such measures of association over the spanning subgraphs  $G_k$  of the valued graphs  $(G, \Delta)$  and  $(G, D(X))$ . FRIEDMAN and RAFSKY [38] decompose  $G$  in two ways. First they consider a family of orthogonal (that is without common edges) spanning subgraphs. This method is reminiscent to the projection pursuit one and at each stage, one derives the minimum spanning tree of the graph  $G$  orthogonal to the union of the minimum spanning trees resulting from former steps. Second, they consider the complementary family of  $k$ -nearest neighbours spanning subtrees of  $G$ , ordered by the integer  $k$ .

When applied both to  $(G, \Delta)$  and to  $(G, D(X))$ , this decomposition provides us with the corresponding evaluations of correlation indices reflecting the "local" association of the initial  $\Delta$  and final  $D(X)$  dissimilarities. In both cases, these indices give a higher weight to small distances than to big ones and so they share the interesting feature to produce a pessimistic evaluation of the solution of MDS algorithms, which fit more badly small distances than higher ones. Besides, both proposals order the multivariate space in a way usable for the elicitation of non linear manifolds. The second decomposition clearly corresponds to a refinement of the drawing proposed by KRUSKAL. It seems more prone to detect clustered configurations than the first one, which is more sensitive to topological structures unfolded along some geodesics.

FRIEDMAN and RAFSKY also elaborate a measure of prediction that can be used when evaluating the predictive power of a MDS configuration on an external numerical variable, or reciprocally of an illustrative ordinal variable on the MDS constellation of points. It is based too on the concept of spanning subgraphs. The rejection of the prediction hypothesis is associated with small values of the statistic

$$\Gamma_2 = \sum_{(i,l) \in G_X} R_i(l)$$

where  $G_X$  is e.g. a spanning subgraph associated with the predictive configuration  $X$  and

$R_i(l)$  is the rank of object  $l$  in the list of observations ordered by increasing  $Y$ -distance from object  $i$ . So  $Y$  is assumed to be either a numerical variable or a dissimilarity index, a situation met in the Unfolding Problem, where one is given an off-diagonal dissimilarity matrix. In both cases,  $Y$  is either observed or built as an indicator for some hypothetical property of the solution that one wants to validate. The power of the method may be improved by projecting this illustrative variable on the embedding space  $E_r$ . Besides, an improvement of the respect the potential non linearity of the manifold results from the categorization of  $Y$ . The illustrative variables method is then used for projecting the indicator function associated with each category on  $E_r$ , and so it yields one of the most efficient tools for the interpretation of the mappings resulting from a MDS algorithm.

Nevertheless it must be noted that the interest of these permutation tests in MDS has been rightly questioned by BORG and LINGOES [39], because of the uniformity assumption made over the set of permutations is to be questioned. In fact, a selection of order relations embeddable in an  $r$  dimensionnal Euclidean space would be necessary in that setting. Work on that problem is under development. While their denumeration has been done, no existing characterization seems to produce a feasible selection process of embeddable order relations, which would be necessary to a right application of the methodology of permutation tests in that setting.

### 3.4. Interpretation of the coordinate axes:

The description of linear structures presented in the preceding section uses coordinate free methods. A complementary approach using the normed linear space aspect of Euclidean space, is based on the determination of the optimal properties of coordinate axes in terms of inter point distances. The isometry  $T$  provides an interpretation in terms of description of the interaction term in the ANOVA like decomposition of the two way cross-tabulation  $H$ , with interaction and no replications. Then, the algebraic relation (3-bis) between  $W$  and  $H$  is interpreted in  $F$  as a double centering operation partialling out the main effects. Besides, d'AUBIGNY[20] shows that the equalities

$$I = \text{Trace}(WN_Q) = \text{Trace}(\Psi\Phi) = \sum_i \sum_{<l} q_{il} \delta_{il}^2(Z) \quad (14)$$

imply the matrix  $\Psi = TW^*T$  determined by

$$\Psi_{il,i'l'} = \begin{cases} \delta_{il}^2(Z) & \text{if } (i,l) = (i',l') \\ -\frac{1}{2} (\delta_{ii'}^2(Z) + \delta_{ll'}^2(Z) - \delta_{il}^2(Z) - \delta_{i'l'}^2(Z)) & \text{else} \end{cases} \quad (15)$$

$\Psi_{i_l, i'_l}$  is the 2x2 contrast  $( (i, l), (i, l') )$  associated with the matrix H. Nullity of  $\Psi_{i_l, i'_l}$  receives a geometric interpretation in terms of orthogonality since

$$\Psi_{i_l, i'_l} = (f^{i*} - f^{i'^*}) W (f^{i*} - f^{i'^*}) \quad (15bis)$$

So, the equation (12) shows that Principal Coordinates Analysis elicits the optimal (most dispersed) orthogonal contrasts of objects in I. When the  $h_{ij}$  may be considered independent, as is often true for applications in ecology, statistical tests apply for the determination of the number of non nul contrasts, and consequently for the number of dimensions to retain for a complete description of interactions, in direct application of the work by JOHNSON and GRAYBILL [40]. The only restriction to that methodology is the gaussian assumption prescribed to the variable  $\delta$ . As far as we know, the robustness of this assumption has not been evaluated. RAMSAY [41] reports that the skewness of Euclidean distances should better be taken into account by a log normal assumption. Besides, in applications where the  $\delta_{ij}^2$  are derived measurements of the observed variables, the validity of the probabilistic arguments founding the tests determining the number of non nul contrasts is not secured, since dissimilarities are then associated. So these tests should be cautiously used as a qualitative guideline. More conceptually, this correspondence shows that the dimensionality of the representation space is not directly linked to the numerical value of the  $\delta_{ij}$ 's. It depends on the 2x2 contrasts, that seem to bring a more satisfactory information when evaluating the dimensionality, the stability of MDS results in presence of missing data, or when designing incomplete collection of dissimilarity judgments.

#### 4. THE CASE OF ORDINAL DATA.

The extension of the MDS methodology to the analysis of  $(I, \delta_I, Q_{I \times I})$ , where  $\delta_I$  is measured on an ordinal scale, implies the resolution of an isotonic regression step at each iteration of the optimization of the criterion function. Two generalizations of the equation (10) may be set up :

$$m(\delta_{ij}) = d_{ij}(X) + e_{ij} \quad (16)$$

$$\delta_{ij} = m(d_{ij}(X)) + r_{ij} \quad (17)$$

where  $m$  is an unknown monotonic increasing function. The equation (16) is basic to the resolution of the MDS problem by KRUSKAL [19]. The ordinal MDS problem follows when alternating the construction of least squares estimations of  $m$  for a fixed  $d_{ij}(X^k)$  and the construction of configurations  $X^{(k+1)}$  in  $E_r$  ( $r$  fixed) in a way to minimize the criterion function. The evaluation of the quality of the regression function associated with the

optimal constellation of points is important in the analysis of the solution  $X$ . Two points are discussed here.

First, the interpretation of the weighted least squares isotonic problem in  $F \wedge F$  as a quadratic programming one with inequality constraints, defining a closed convex polyhedral cone in that space:  $K = \{ \underline{x} ; {}^t A \underline{x} \leq 0 \}$ , where the matrix  $A$  describes the ordinal constraints induced by the observed dissimilarity data  $\Delta$ . So any  $t$  in  $K$  is a monotonic function  $m(\Delta)$ , that is  $t_{ij} = m(\delta_{ij})$ , with  $m$  a monotonic increasing function.  $t$  is called a disparity. The resolution of this least squares problem provides us with an optimal disparity  $t^*$  defined as the orthogonal projection of  $d(X)$  on  $K$ . So  $t^*$  belongs to the lineality space supporting a face of  $K$ . So an a posteriori analysis of the problem can interpret  $t^*$  as the solution of a multiple linear regression problem, in which the predictors are the generators of this lineality space.

This is the translation in matrix form of the interpretation of the algorithm as a way to build the finest partition of  $I \times I$  which respect the order relation induces by  $\Delta$ . So, let  $A$  denote the indicator matrix of the blocs obtained by the isotonic regression problem. The columns of  ${}^t A$  form a basis of the lineality space of interest. They may be regarded as a subset of the possible predictors. Thus one may study the stability of the solution  $t^*$ , or equivalently of the lineality space, with help of Jack-knife or of Bootstrap methods. In the case of complete order relations induced by  $\Delta$ , the isotonic problem is not time consuming and does not involve high memory requirements, and these studies of stability are clearly feasible. MEULMAN [46] applies a Bootstrap method to the evaluation of the solution of the MDS solution in a distinct and complementary way, since she considers the stability of the whole process while here our interest is limited to the isotonic regression step of the iteration.

Besides, this evaluation is complementary to the global evaluation of the monotonic regression function advocated for by PLESZSZYNSKA et al. [42]. The interest of the graph of the monotone dependence curve lies in its skillful description of the strenght of the positive relationship between distances and dissimilarities, and it may be used as an indicator of influential pairs of statistical units in the elaboration of the solution. More specifically, since the STRESS function is closely related to the (weighted) cross product of the vectors  $\delta$  and  $d(X)$  in  $F \wedge F$  and large crossproducts, that is low values of the STRESS function, are often achieved by a good correspondence of very large values of  $\delta$  and  $d(X)$  while small and medium values are only weakly related. A thorough description of the relationship is more satisfactorily obtained by a function that evaluates the tendency of coappearance of values greater than their respective  $p$ -th quantile in both variables.

So consider the isotonic regression problem of  $D(X)$  on  $\Delta$ , where  $(D(X), \Delta)$  is interpreted as a bivariate random vector, defined on the probability space  $(\Omega, \mathcal{A}, P)$ . When one is given a one-to-one application of  $I$  onto  $[n]$ , let us consider  $\mathcal{C} = \{(i,j) \in I \times I, i < j\}$ . Since  $\text{card}(\mathcal{C}) = n^*$ , with  $n^* = n(n-1)/2$ , we consider a one-to-one application of  $\mathcal{C}$  on  $[n^*]$ . Now if  $\delta_{[s]}$  denotes the  $s$ -th quantile of the distribution of  $\Delta$ , let us define

$$L_{D(X),\Delta}(s) = E(D(X)) - E(D(X)/\Delta \leq \delta_{[s]})$$

$$L_{D(X),D(X)}(s) = E(D(X)) - E(D(X)/D(X) \leq d(X)_{[s]})$$

The monotonic regression function proposed by PLESZSZYNSKA et al. [42] may be written here:

$$\mu_{D(X),\Delta}^+(s) = \mu_{D(X),\Delta}^+(s) \text{ if } L_{D(X),\Delta}(s) \geq 0, \text{ and } \mu_{D(X),\Delta}^-(s) \text{ if } L_{D(X),\Delta}(s) \leq 0$$

where

$$\mu_{D(X),\Delta}^+(s) = \frac{L_{D(X),\Delta}(s)}{L_{D(X),D(X)}(s)}$$

and  $\mu_{D(X),\Delta}^-(s)$  is defined in an analogous way when substituting  $(\Delta \geq \delta_{[s]})$  for  $(\Delta \leq \delta_{[s]})$  in  $L_{D(X),\Delta}(s)$ . The sample counterpart of this function is obtained in  $F \wedge F$  when using the estimators

$$\hat{E}(D(X)) = \sum_{s=1}^{n^*} q_s d_s(X)$$

$$\hat{E}(D(X)/\Delta \leq \delta_s) = \sum_{t=1}^s q_t d_t(X)$$

This second expression is well known since it yields the ordinate of Cumulative Sum Diagram that is used to interpret the Lagrangian dual of the isotonic regression problem, see BARLOW et al.[48]. The abscissa are then

$$p_s = \sum_{t=1}^s q_t$$

d'AUBIGNY [20] shows that the Lagrangian multipliers of the isotonic regression problem are

$$\lambda_s = \sum_{t=1}^s q_t (d_t(X) - d_t^*)$$

where  $d^*$  is the disparity, solution of the problem, that is a monotonic transformation of  $\delta$ . As a consequence

$$\tilde{\lambda}_s = \frac{\lambda_s}{p_s} = \hat{E}(D(X)/\Delta \leq \delta_{[s]}) - \hat{E}(D^*/D^* \leq \delta^*_{[s]})$$

Thus

$$\mu_{D(X), D^*}^+(s) = 1 - \frac{\hat{\lambda}_s}{\hat{E}(D(X)) - \hat{E}(D^*/D^* \leq d_s^*)}$$

This gives a probabilistic interpretation of this dual problem, and one has

**Proposition:**

*If  $\lambda_s = 0$  for all  $s$  in  $\mathcal{T}$ , then the relation between  $\Delta$  and  $D(X)$  is isotonic and*

$$\mu_{D(X), D^*}^+(s) = 1 \text{ for all } s \text{ in } \mathcal{T}$$

*If  $\lambda_s > 0$  for some  $s$  in  $\mathcal{T}$ , then  $\mu_{D(X), D^*}^+(s) < 1$  for this  $s$ .*

*If  $\mu_{D(X), D^*}^+(s) = 0$  for all  $s$  in  $\mathcal{T}$  then  $\hat{E}(D(X)/\Delta) = \hat{E}(D(X))$ .*

In this last case the solution does not depend on the ordinal information induced on  $\mathcal{T}$  by  $\Delta$ . Moreover, PLESZSZYNSKA et al. [42] show that the monotone regression function is invariant by any isotonic transformation of  $\Delta$  and any positive affine transformation of  $D(X)$ . So this tool is clearly a valuable add-on to the Shepard diagram for the evaluation of the final transformation of the dissimilarity function, that realizes the minimum of the STRESS function. It gives complementary informations on the concentration of points around the monotone function drawn on the Shepard diagram, and permits us to detect potential non uniformities in the dependence of  $D(X)$  on  $\Delta$ , which is one of the causes of the appearance of horse-shoe shaped configurations.

A second consequence of this geometric approach refers to the related elicitation of influential points. The construction of the optimal lineality space of  $\mathfrak{t}^*$  furnishes the active constraints, corresponding to equalities plotted as constant steps on the scatterplot  $\{(\delta_{il}, t_{il}^*) ; (i, l) \in I \times I\}$ . It means that the indicator matrix of actives constraints is easily available and so is the corresponding Hat matrix  $H = WD_p$ . The consultation of its diagonal terms, using either graphical or statistical tools, or preferably both, provides a way to elicitate influential points if any.

The second point of interest specific to the Ordinal MDS problem is the recurring appearance of degenerate solutions, obtained when the recovered data points tend to cluster on well separated categories, so that the Shepard diagram shows a very small

number of steps. This is often the case when the statistical units are not numerous enough to secure a strong determinacy of the space  $E_r$ . This problem refers to the kind of monotonicity constraints prescribed to the solution, and a well known partial protection against their appearance may be obtained when substituting a rank image transformation of the dissimilarities on the distances (GUTTMAN [47]) in the ordinal regression step of the iteration. But this solution is not satisfactory from a theoretical point of view since the rank image transformation does not optimize the STRESS function. This situation indicates that a great majority of inequalities have been satisfied by transforming them in strict equalities. The dimensionality of the supporting hyperplane of the facet of  $K^*$  belongs to, is very low and then bootstrapping indicates that the solution is quite unstable.

We present here a two steps procedure, that seems promising, on the grounds of numericals experiments (cf. d'AUBIGNY and EL HADRI [43]). The first step of the method uses in the ordinal MDS context a smoothing monotone scatterplots method proposed by FRIEDMAN and TIBSHIRANI [44]. Let us denote by  $m$  the pair  $(i,l) \in [n] \times [n]$ ,  $i < l$ ,  $m \in [n^*]$ . The problem consists on the derivation of a smooth monotone function adjusting in a least squares sens the equation (16). The minimization of

$$ESE = \sum_m q_m (f(\delta_m) - m(\delta_m))^2 \quad (18)$$

where  $f(\delta_m) = d_m(X)$  up to errors  $e_m$ , has no solution as long as  $f$  is unknown. FRIEDMAN and TIBSHIRANI estimate  $f$  with help of an empirical smoother. For the sake of simplicity let us consider here a running means of span  $k$ :

$$s_k(\delta_m) = AVE(d_{m-k}(X), \dots, d_m(X), \dots, d_{m+k}(X)) \quad (19)$$

the windows are more over skrunken near the left and right endpoints. So the right set of indices in a window is  $\{\text{Max}\{1, m-k\}, \dots, m, \dots, \text{Min}\{M, m+k\}\}$ . Other more sophisticated smoothers are clearly possible, e.g. the use of filters interpreted as Kernel estimators of the conditional expectation  $E(D(X)/\Delta)$  (see d'AUBIGNY [20]) or of spline functions has been often advocated for in recent publications about MDS in the psychometric litterature. This empirical smoother has been retained because of its efficiency during numerical experimentations (see G. d'Aubigny and El Hadri [43]). The optimal span  $k$  is determined as the optimizer of the criterion function

$$\sum_m q_m (d_m(X) - s_k^m(\delta_m))^2 \quad (20)$$

in which  $s_k^m(\delta_m)$  estimates  $f(\delta_m)$  with help of a formula analogous to (19), where  $d_m(X)$  is excluded. The minimization of equation (18) results from the following algorithm that iterates two phases:

- (1) obtain  $s(\delta)$  in smoothing  $d(X)$  by (19), with  $k$  a solution of the minimization problem (20),
- (2) find the monotone transformation of  $\delta$  closest to  $s(\delta)$  and noted  $m(\cdot)$ .

The numerical experiment [43] produced an empirical obviousness of the efficiency of this smoothing techniques that revealed itself less expensive and often quite as good as the monotonic smoothing splines. Our present implementation of the smoothing step in MDS is basically founded on the method of Friedman. Nevertheless, when the optimum of the MDS problem is attained, we turn to a smoothing splines algorithm in a way to apply the results of EUBANK [45] in the evaluation of the optimalities of the solution. Then the method consist in the minimization of the penalized least squares Loss function

$$\sigma_{\rho}(\Delta, D(X)) = \sum_{t=1}^{n^*} q_t (d_t(X) - s(\delta_t))^2 - \rho \int_a^b \left[ \frac{\partial^m s(\delta)}{\partial \delta^m} \right]^2 d\delta$$

over all functions  $s$  having  $m-1$  absolutely continuous derivatives and a square intergrable  $m$ -th derivative on  $[a, b]$ , where the smoothing parameter belongs to  $(0, \infty)$  and  $n^* \geq m$ . The solution of the problem, say  $\hat{s}_{\rho}$ , is well known to be a polynomial spline of order  $2m$  and knots at  $\delta_t, t \in [n^*]$ , that is a smoothing spline, see e.g. SCHOENBERG [49]. the parameter  $\rho$  governs the amount of smoothing and it is null for the interpolation splines and  $\infty$  for the regression problem on polynomials of order  $m$ . Now, since the smoothing spline is a linear operator one may write the vector of fitted values as

$$\hat{t} = H(\rho)d(X) \quad \hat{t}_i = (\hat{s}_{\rho}(\delta_t))_{t \in [n^*]}$$

where  $H(\rho)$  is called the Hat matrix by analogy with the usual least squares multiple linear regression problem

$$d(X) = 'X\beta + \epsilon$$

which expresses the solution as a function of  $d(X)$  and of the matrix  $H$ , the Hat matrix, of the projector on the sub-space of  $F \wedge F$  spanned by the regressors, the rows of  $X$ :

$$\hat{t} = Hd(X) \quad H = XV_X^+ 'X\Phi$$

where  $\Phi$  defines the Euclidean geometry on  $F \wedge F$ , and  $V_X$  is the covariance matrix of the regressors. In the case of smoothing splines,  $H(\rho)$  is not the matrix of a projector. Nevertheless Eubank shows that it satisfies most of the properties of  $H$  of interest for the detection of influential observations. This author defines it as the generalized Hat matrix in the case of smoothing spline functions, which makes it possible to detect potential

influential points that could distort the shape of the regression function, and to evaluate their leverage, in a way analogous to the ordinary multiple linear regression situation. So it is clearly a valuable tool for the evaluation of the solution of a Smooth Ordinal Multidimensional Scaling method.

## 5. REFERENCES

- [1] BICK, W. , MUELLER, P. , BAUER, H., GIESEKE, O., Multidimensional Scaling and Clustering Techniques. (1978). Tech. Rep.. Institut für Angewandte Sozialforschung Universität zu Köln, West Germany.
- [2] NISHISATO,S. , Multidimensional Scaling : A Historical Sketch and Bibliography. (1978). Tech. Rep.. The Ontario Institut for Studies in Education, Toronto, Canada.
- [3] de LEEUW, J. , HEISER, W. , Multidimensional Scaling with Restrictions on the Configuration. (1980). In P. R. Krishnaiah (Ed.). *Multivariate Analysis V*. North Holland, Amsterdam. 501-522.
- [4] CARROLL, J. D., ARABIE, P., Multidimensional Scaling. In M.R. Rosenzweig, L.W. Porter (eds.). *Annual Review of Psychology* 31 (1980), 607-649.
- [5] LEBART, L. ,MORINEAU, A., WARWICK, K. M., *Multivariate Descriptive Statical Analysis*. (1986). Wiley, New York.
- [6] JAMBU, M. and LEBEAUX, M. O., *Cluster Analysis and Data Analysis*. (1983). North Holland, Amsterdam.
- [7] CHANDON, J. L. and PINSON, S. , *Analyse Typologique*. (1981). Masson, Paris.
- [8] FICHET, B., Distances and Euclidean Distances for Presence-Absence Characters and their Application to Factor Analysis. (1986). *Multidimensional Data Analysis*. J. de Leeuw, W. Heiser, J. Meulman (Eds.). DSWO Press, Leiden.
- [9] GOWER, J. C., Euclidean Distances Matrices. (1986). *Multidimensional Data Analysis*. J. de Leeuw, W. Heiser, J. Meulman (Eds.). DSWO Press, Leiden.
- [10] OBENCHAIN, R. L., Affine Multidimensional Scaling. Personal Communication. (1971).
- [11] YOUNG, G., HOUSEHOLDER, A. S., Discussion of a set of points in term of their mutual distances. *Psychometrika* 3 (1938), 19-22.
- [12] TORGERSON,W.S., *Theory and Methods of Scaling*. (1958), Wiley, New York.
- [13] GOWER, J. C., Some Distance Properties of Latent Root and Vector Methods used in Multivariate Analysis. *Biometrika* 53 (1966), 325-388.
- [14] GOWER, J. C., Euclidean Distance Geometry. *Mathematical Scientist* 7 (1982), 1-14.
- [15] CAILLEZ, F., PAGES, J. P., *Introduction à l'Analyse des Données*. (1976). SMASH 9 rue Duban, Paris.

- [16] MEULMAN, J., A Distance Approach to Nonlinear Multivariate Analysis. (1986). DSWO Press, Leiden.
- [17] d'AUBIGNY, G., Les méthodes de Codage Multidimensionnel. Tech. Rep.. (1987). IMSS, Grenoble.
- [18] SHEPARD, R. N., The Analysis of proximities: Multidimensional Scaling with an unknown distance function. *Psychometrika* 27 (1962), 125-140, 219-246.
- [19] KRUSKAL, J. B., Multidimensional Scaling by Optimizing Goodness-of-fit to a nonmetric Hypothesis. *Psychometrika* 29 (1964), 1-28, 115-129.
- [20] d'AUBIGNY, G., Un Modèle Algébrique pour l'Analyse des Données de dissimilarité. (1984). Troisièmes Journées Franco-Soviétiques de Statistique et analyse des Données. Paris.
- [21] d'AUBIGNY, G., Interpretating Mappings issued from Distances based Methods. Workshop on Multidimensional Data Analysis. Pembroke Colledge, Cambridge. (1985).
- [22] COXETER, H. S. M., Regular Polytopes. (1947). Methuen & Co., LONDON.
- [23] DANIELS, H. E., The Relation between Measures of Correlation in the Universe of Sample Permutations. *Biometrika* 33. (1944), 129-135.
- [24] de LEEUW, J., Applications of Convex analysis to Multidimensional Scaling. (1977). In J.R. Barra et al. (Eds.). Recent Developments in Statistics. North Holland, Amsterdam.
- [25] TAKANE, Y., YOUNG, F.W., De LEEUW, J., Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika* 42 (1977), 7-42.
- [26] BROWNE, M., W. , The Young-Householder Algorithm and the Least Squares Multidimensional Scaling of Squared Distances. (1987). Research Report. University of South Africa, Pretoria;
- [27] McDONALD, R. P., The Theoretical Foundations of Principal Factor Analysis, Canonical Factor Analysis and Alpha Factor Analysis. *The British Journal of Mathematical & Statistical Psychology* 23 (1970), 1-21.
- [28] TAKANE, Y., On the Relations among four Methods of Multidimensional Scaling. *Behaviormetrika* 4 (1977), 29-43.
- [29] BENZECRI, J. P. , l'Analyse des Données : II l'Analyse des Correspondances. (1973). Dunod, Paris.
- [30] d'AUBIGNY, G., Description Statistique des Données Ordinales: Analyse Multidimensionnelle. (1975). Thèse, Grenoble.
- [31] CARROLL, J.D. , PRUZANSKY, S., TVERSKI, A., Spatial versus Tree Representations of Proximity Data. *Psychometrika* 47. (1982), 3-24.
- [32] BENTLER, P. M., WEEKS, D. G., Restricted Multidimensional Scaling. *Journal of Mathematical Psychology* 17 (1978), 138-151.

- [33] LEE, S., BENTLER, P. M., Functional Relations in Multidimensional Scaling. *British Journal of Mathematical and Statistical Psychology* 33 (1980), 142-150.
- [34] LINGOES, J. C., BORG, I., A quasi-statistical Model for Choosing between Alternative Configuration derived from Ordinally Constrained Data. *The British Journal of Mathematical & Statistical Psychology* 36 (1983), 36-53.
- [35] d'AUBIGNY, G., Un Modèle Algébrique pour l'Analyse des Données de dissimilarité sous Contraintes. (1987). *Troisièmes Journées Franco-Soviétiques de Statistique et analyse des Données. Souzdal, URSS.*
- [36] KRUSKAL, J. B., WISH, M. *Multidimensional Scaling.* (1978), Sage P., Beverly hills, C.A.
- [37] HUBERT, L. J., SUBKOVIK, M. J., Confirmatory Inference and Geometric Models. *Psychometrik Bulletin* 86. (1979), 361-370.
- [38] FRIEDMAN, J., RAFSKY, L. C., Graph-Theoretic Measures of Multivariate association and Prediction. *The Annals of Statistics* 11. (1983), 377-391.
- [39] BORG, I., LINGOES, J. C., Multidimensional Scaling with Side Constraints on the Distances. (1977). In J.C. Lingoies, E.E. Roskam, I. Borg (Eds.) *Geometric Representation of Relational Data.* Mathesis Press, Ann Arbor, Michigan.
- [40] JONHSON, D. E., GRAYBILL, F. A. , An analysis of a Two-Way Model with Interaction and no Replication. *Journal of American Statistical Association* 67 (1972), 862-868.
- [41] RAMSAY, J. O., *MULTISCALE II Manual.* (1986). Tech. Rep.. MCGILL University, Montreal.
- [42] PLESZCYNKA, E., CWIK, J., PAPAGORASZ, T., PALINKAS, E., On seriation and Multidimensional Scaling. *COMPSTAT* (1978), Physica-Verlag, WIEN, 311-317.
- [43] d'AUBIGNY, G., EL HADRI, Regression Isotone Lisse : Evaluation Numérique comparative des Splines Monotones et des Lisseurs Empiriques. (1987). Tech. Rep.. IMSS, Grenoble.
- [44] FRIEDMAN, J., TIBSHIRANI, R., The Monotone Smoothing of Scatterplots. *Technometrics* 26. (1984), 243-250.
- [45] EUBANK, R. L., The Hat Matrix for Smoothing Splines. *Statistics & Probability Letters* 2. (1984). 9-14.
- [46] MEULMAN, J., Homogeneity Analysis of Incomplete Data. DSWO Press. Leiden. (1982).
- [47] GUTTMAN, L., A General Nonmetric Technique for Finding The Smallest Coordinate Space for a Configuration of Points. *Psychometrika* 33: 469-504.
- [48] BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., BRUNK, H.D., *A Statistical Inference under Order Restrictions.* J. WILEY, New-York.
- [49] SCHOENBERG, I.J., Spline Functions and the Problem of Graduation. *Proc. Nat.Acad.Sci.USA* 52: 947-950.



**PARTIE D****THE ADDITIVE DECOMPOSITION OF SOME  
ENTROPY FUNCTIONS****AND****(CONSTRAINED-) ORDINATION METHODS**

- Chapitre I**            **INTRODUCTION**
- Chapitre II**          **THE ANALYSIS OF DIVERSITY**
- Chapitre III**        **THE EUCLIDEAN EMBEDDING OF DISSIMILARITY  
DATA**
- Chapitre IV**        **ORDINATION METHODS WITH INSTRUMENTAL  
VARIABLES**



**THE ADDITIVE DECOMPOSITION OF SOME  
ENTROPY FUNCTIONS  
AND  
(CONSTRAINED-) ORDINATION METHODS**

**G. d' AUBIGNY**

**UNIVERSITE DES SCIENCES SOCIALES DE GRENOBLE  
47X - 38041 GRENOBLE CEDEX - FRANCE**

**S U M M A R Y**

The ordination methods, based on the multidimensional scaling of diversity and dissimilarity data, are interpreted as additive decompositions of a quadratic entropy function. Using the analogy with the analysis of variance methodology (Rao (1982)), the axes of the recovered mapping are interpreted in terms of 2x2 contrasts. Then the constraints induced either by a design of experiment, or by the conjoint measurement of external variables are taken into account for exploratory as well as confirmatory purposes. These results exhibit the complementary aspects of these ordination methods and of the usual analysis of experimental designs, which revealed themselves as fruitful for the sake of statistical analysis of dissimilarity data as are the well known and often used links that tie in Principal Component Analysis with ANOVA for the analysis of profile data.

**Key words :** Entropy function, diversity, dissimilarity, ordination method, 2x2 contrasts, ANOVA.

## 1. INTRODUCTION

This paper discusses a mixed methodology aimed at the analysis of measurements of inter-points dissimilarities with help of the Multidimensionnal Scaling (or MDS for short) approach. Ordination methods are now of common use in biometry (Gower (1984)), e.g. for the sake of analysis of incidence or abundance data and more broadly of measures of diversity within populations and of dissimilarity (or similarity) between populations.

The first kind of data is mostly encountered in ecology and analysed with help of gradient analysis methods (Whittaker (1967), Orloci (1979)) and of correspondence analysis, either detrended or not (Hill and Gauch (1980), ter Braak (1984)). Our interest focuses on the second kind of data sets which motivated numerous studies in anthropology (Rao (1948), Mahalanobis et al. (1949)), in genetics (Cavalli-Sforza (1969)) and in biology (Sokal and Sneath (1963), Pielou (1975), Patil and Taille (1979)). Dennis et al. (1979) give an extensive bibliography about measures of diversity and their applications.

Consider, say, some vegetation data. The information about species performance, community composition or ecological significance is often communicated in the form of a rectangular array of profile data. The  $i$ -th row describes the response of a given species, while columns are  $p$  dimensional quadrat descriptors (or relevés). In that situation, a great number of statistical methods exist and may be used in complementarily, in a way to analyse the data from either an exploratory or a confirmatory point of view, in accordance to their nature.

Our interest here is about a second type of data structure, where the data array  $\Delta = (\delta_{ij} ; i, j = 1, 2, \dots, n)$  carries informations about the strength of dissimilarity of the responses of given species. In that case a Multidimensionnal Scaling or MDS approach is generally used in a way to get rid of non linear data structures, from an exploratory point of view. The analyst consults the resulting (hopefully easily interpretable) mapping to grasp some latent information in the original array  $\Delta$ . The difficulties met in the determination of a satisfactory dimensionality for the representative space which is basic to this data reduction method, as well as the technical aids for reading and interpreting the resulting scatterplots will not be discussed here. The interested reader is referred to Kruskal and Wish (1978) and d'Aubigny (1987).

Although the MDS methodology relies on the assumption of (or merely the hope for) an efficient description of a latent, possibly non linear disjoint and/or curved, data structure linked with environmental influence gradients, it is often complemented with some clustering methods looking for discontinuities. The problems raised by this methodology are known for a long time (see Beals (1973) for criticisms in ecology) and are not to be discussed here.

Our aim here is to investigate methods tailored for handling experimental dissimilarity data in an inferential way, to show their coherence to the more well known ANOVA setting, in a way to motivate some progresses in the application of MDS techniques for the analysis of planned experiments. This approach is based on some new statistical interpretations of the components of the MDS approach, and on the demonstration of the complementarity of these exploratory tools to the testing methodology to be proposed.

Our approach recognizes the fact, first discussed by Rao (1982), that dissimilarity and entropy are just but two dual facets of the statistical treatment of diversity between individuals of the population of interest. On the one hand a dissimilarity function is chosen in practice for its biological meaning to reflect some intrinsic dissimilarity between individuals relevant to a particular investigation. This choice implicitly induces the definition to be emphasized here, of an entropy index which is most often neglected or ignored. The second way round directly considers a measure of entropy as a function defined on a space of distribution functions for the sake of comparison of their dissimilarity to the uniform distribution or of their differences in reference to it. This entropy must satisfy some classical postulates which do not refer explicitly to the individuals constituting the population of interest and ignores their dissimilarities. This way of thinking is exemplified in the generalization of linear models based on deviance measures and quasi likelihood estimation, see McCullagh and Nelder (1983). In that case it appears most often a difficult task to find a dissimilarity index that is associated to the deviance.

First the one way ANOVA is interpreted in terms of dissimilarities in section 2.1, for the sake of illustration. Its extension to the class of dissimilarity functions generating quadratic entropy functions (Rao (1982)) is advocated in section 2.2 and the resulting inferential problems are given a simple solution in terms of randomization tests, following Hubert et al. (1982). In section 3, the MDS methodology is interpreted as a tool tailored for the description of interactions evaluated on a symmetric two-way table. 2x2 contrasts play a prominent part in this interpretation and supply the latent variables associated to each axe of the configuration with an optimal contrasts interpretation. In section 4, an extension of this approach integrates some instrumental variables in the realm of the MDS

methodology. These external variables usually result either from some a priori knowledge about the problem at hand (e.g. the description of a design of experiment) or from the measurement of environmental variables (e.g. of pedological variables in vegetation).

## 2. THE ANALYSIS OF DIVERSITY

Let us consider the results (reported in d'Aubigny et al. (1988)) of the following experiment, aimed at the evaluation of the influence of everyday used industrial solvents on several morphometric parameters describing the myelinated fibre size distribution in the rat's trigeminal nerve. Trigeminal nerve impairment is in fact the main characteristic of the exposure to Trichloroethylene (or TRI). One of three treatments (no drug, exposure to trichloroethylene, and exposure to dichloroacetylene (or DCA, a degradation product from TRI) were administered in a unified protocol of intoxication, to a rat chosen at random in a group of animals selected in a way to assure as much homogeneity as possible. See Barret et al. (1988). This controlled experiment is characterized by one factor with three levels and four replications for each treatment. One wants to decide about the differential effect of intoxications. The response variable is multivariate, since several myelinated fibre size characteristics are observed (the inner and outer contours of the myelin sheaths of the myelinated fibres, and a circularity index based on the outer contour were measured, as well as the diameter of a circle of equal perimeter, the axonal contour of unmyelinated fibres, and the thickness of myelin lamellae). Each variate is measured here in the form of a repartition in predetermined contiguous categories that are identical for each rat. Since the consultation of the histogram associated to each rat during a pre-experiment showed that large inter-individual variabilities affect the whole shape of the repartitions, the results of this experiment could not be analysed in the usual terms of one-way ANOVA on a replicated design applied to central position summaries. Nevertheless, the next section discusses for the sake of comparisons the ANOVA methodology to be used, if it were sensical.

### 2.1. *The one way analysis of variance*

Let us consider the usual additive model  $y_{ik} = \mu + \alpha_i + \epsilon_{ik}$ , where the indices  $i$  and  $ik$  refer to the treatments and the replications respectively. It induces an additive decomposition of the sample variance which is basic to the derivation of F-statistics in the ANOVA setting

$$s_T^2(Y) = s_W^2(Y) + s_B^2(Y),$$

where  $s_T^2(Y)$ ,  $s_W^2(Y)$  and  $s_B^2(Y)$  denote the Total, Within and Between variances of the response variable  $Y$  respectively. In general, when the statistical units  $\omega_i$  (with  $i \in I = \{1, \dots, n\}$ ) are weighted by  $p_i$ ,  $p_i \geq 0$ ,  $\sum p_i = 1$ ,  $s_T^2(Y)$  may be written

$$s_T^2(Y) = \sum_{i=1}^n p_i (y_i - \bar{y})^2 = \frac{1}{2} \sum_i \sum_j p_i p_j (y_i - y_j)^2$$

where  $\delta_{ij}^2(Y) = (y_i - y_j)^2$  quantifies the strength of dissimilarity between the elements of each pair of statistical units  $(\omega_i, \omega_j)$ . Now  $\Delta(Y) = (\delta_{ij}(Y))$  summarizes these informations in a less usual form of a square symmetric dissimilarity matrix.

The usual decomposition of sums of squares corresponds here to a partition of  $I$  in  $K$  groups of statistical units submitted to the same (combination of) treatments, where  $K$  is the number of levels of the (combination of) factor(s) of interest. Under a proper indication of the elements of  $I$ , or equivalently after due permutation of rows and columns, there corresponds to this partition a decomposition of  $\Delta(Y)$  into blocs, such that one derives the Within weighted Sum of Squares by elementary summations of terms inside the diagonal blocs designed on  $\Delta(Y)$  by the factors. For example, in the trigeminal nerves experiment discussed above, there were one factor with  $K=3$  levels, designing three diagonal blocs. The analytical expression of  $s_W^2(Y)$  and  $s_B^2(Y)$  in terms of distances is derived by simple algebra (see Takeuchi et al. (1982)), since any partition of  $I$  in  $K$  classes  $k$  yields the Within class sample variance may be as a weighted sum of within diagonal blocs dissimilarities

$$s_W^2(Y) = \frac{1}{2} \sum_{k=1}^K \sum_{(i,j) \in k} p_i p_j \delta_{ij}^2$$

From a test of hypotheses point of view, the use of the  $F$  statistic rely on strong probabilistic assumptions about the distribution of  $Y$ , which are not satisfied by dissimilarity functions more general than the kernel  $\delta_{ij}^2(Y) = (y_i - y_j)^2$  of the variance operator used above. Hence we propose to restrain inferences to the set of observed statistical units, without references to a super population from which the might have been drawn from a controlled sampling process. We make use of design based inferences in that we follow Hubert et al. (1982) in using a randomization test methodology (Edgington (1980)). Since the Total Sum of Squares is invariant by permutations of the statistical units, the  $F$  statistics is a monotonic function of  $s_W^2(Y)$ . The numerical value of this statistic depends on the patterning of the dissimilarities and on the allocation of the statistical units to blocs. So, the results of the controlled experiment are associated to one

of the  $n! / \prod_{k=1}^K n_k!$  possible assignments of the  $n$  statistical units to the  $K$  groups (each group having  $n_k$  elements). The implementation of the randomization test evaluates the test statistic  $s_W^2(Y)$  for each possible allocation of the statistical units to the groups, and the significance level of the test is derived as the proportion of permutations leading to  $s_W^2(Y)$  values greater than or equal to the one obtained from the experiment. In practice Monte Carlo significance testing procedures must be carried out in these situations where the number of possible allocations is too large for complete enumeration (Edgington (1980)). Thus the significance level of each partition associated to an hypothesis is estimated from a sample of say, one thousand uniform random permutations of rows and columns of  $\Delta(Y)$ .

This methodology clearly extends to more complex designs, since it depends only on the lattice structure of the set of partitions of  $I$ . For instance, let  $A$  and  $B$  denote two factors on  $I$ , associated to the partitions  $\mathcal{P}_A$  and  $\mathcal{P}_B$  of  $I$  respectively. Since  $\mathcal{P}_A \cap \mathcal{P}_B$  is a partition finer than both  $\mathcal{P}_A$  and  $\mathcal{P}_B$ , the decompositions of  $\Delta(Y)$  in diagonal blocs associated to  $\mathcal{P}_A$ ,  $\mathcal{P}_B$  and  $\mathcal{P}_A \cap \mathcal{P}_B$  respectively make it possible to use a randomization strategy, based on summations of within diagonal blocs dissimilarities after due permutations of rows and columns, for testing the existence of main effects and/or interactions. The treatment of hierarchical designs of experiments in that setting is just but an example of such a technique presented in Hubert et al. (1978). The form of this extension of the ANOVA methodology is now clear. Given a sequence of nested models, we can use the Within classes Diversity derived from a given dissimilarity function  $\Delta$

$$H_W = \frac{1}{2} \sum_{k=1}^K \sum_{(i,j) \in k} p_i p_j \delta_{ij}$$

as a generalized measure of discrepancy analogous to the univariate additive decomposition of variance. The terms of such a methodology are discussed in the following. Nevertheless, the interpretation of the results must be cautious, since the orthogonality of the terms of the additive decomposition of the variance rely on the Euclideanicity of the dissimilarity function used in that case. In the general case two questions are in order. First some conditions must be satisfied by the dissimilarity function in a way to obtain an additive decomposition of the total diversity in nonnegative components. Second, in that case each test statistic measures the variation accounted for by that term having eliminated those terms above it, but it ignores any effect of the terms below it. Thus many models sequences may have to be considered in a complicated design of experiment, each producing its own decomposition of diversity, as in the more well known case of non orthogonal factors in ANOVA.

## 2.2. One extension of this methodology to some general dissimilarity functions

The apportionment of variance expressed as a function of  $\delta_{ij}^2(Y) = (y_i - y_j)^2$  is basic to the ANOVA methodology and sets the problem of what conditions are necessary for its extension to more general dissimilarity functions, as well as that of the appropriateness of a randomization testing strategy in that setting. Clearly the constraints of nonnegativeness discussed above are satisfied by most of the dissimilarity functions used in practice. We shall introduce now some complementary regularity conditions, that appear necessary for the apportionment of diversity measures leading to analogs of the ANOVA tables. Much work has been done in recent years in that direction. More specifically, Rao (1982, 1984) examines the convexity properties prescribed to entropy function in a way to analyse diversity data.

**Definition** (Rao (1984)) : Let  $(\Omega, \mathfrak{A})$  a measurable space and  $\mathfrak{P}$  the convex set of probability measures defined on it. A function  $H$  mapping  $\mathfrak{P}$  on the real line  $\mathbb{R}$  is said to be a diversity measure if it satisfies the condition

$$(C0) H(P) \geq 0 \quad \forall P \in \mathfrak{P} \text{ and } H(P) = 0 \text{ iff } P \text{ is degenerate.}$$

Thus the diversity is null if all the individuals of the population are identical for  $P$ . The following condition (C1) also seems natural to impose to a diversity function in the case of a mixture of populations with a priori probabilities  $\lambda_1$  and  $\lambda_2$  ( $\lambda_1, \lambda_2 \geq 0$   $\lambda_1 + \lambda_2 = 1$ ) :

$$(C1) H(\lambda_1 P_1 + \lambda_2 P_2) \geq \lambda_1 H(P_1) + \lambda_2 H(P_2)$$

with strict inequality when  $P_1 \neq P_2$ .

This is clearly equivalent to imposing that the diversity function  $J^{(0)} = -H$  is a strictly convex function. Thus, if  $P_1$  and  $P_2$  are different, (C1) implies that the diversity of a mixture is greater than the average diversity of its constituents. More generally, if  $\{P_k; k = 1, 2, \dots, K\}$  is a finite set of probability distributions defined on  $K$  sub-populations of the same probability space, and  $\lambda_k$  is the prior probability of the sub-population  $k$ , then the Total Diversity decomposes in

$$H\left(\sum_k \lambda_k P_k\right) = \sum_k \lambda_k H(P_k) + J^{(0)}(P_k, \lambda_k; k = 1, \dots, K)$$

which may be written  $H_T = H_W + H_B$  by analogy to the ANOVA setting, since by concavity both terms are nonnegative. One then associates to the decomposition of the dissimilarity table into blocs an additive decomposition of the total diversity, which is analogous to the additive decomposition of variance in the one way ANOVA table. This decomposition is complemented by the required permutation tests. The degree of

generality is clearly independent of the number of subpopulations, so let us look more closely to the case of two subpopulations. Then, the so called first Jensen difference between  $P_1$  and  $P_2$  is positive :

$$J^{(1)}((P_i, \lambda_i)_{i=1,2}) = \lambda_1 J^{(0)}(P_1) + \lambda_2 J^{(0)}(P_2) - J^{(0)}(\lambda_1 P_1 + \lambda_2 P_2) \geq 0$$

Hence  $J^{(1)}$  may be interpreted as a measure of dissimilarity between  $P_1$  and  $P_2$ . This mixing process may be iterated in a hierarchical way. For instance one should require that two pairs of probability  $(P_1^1, P_2^1)$  and  $(P_1^2, P_2^2)$  in  $\mathcal{P} \times \mathcal{P}$ , the dissimilarity between the two mixtures  $(\mu_1 P_1^1 + \mu_2 P_2^1)$  and  $(\mu_1 P_1^2 + \mu_2 P_2^2)$  is not larger than the average of dissimilarities between  $P_1^1$  and  $P_2^1$  and between  $P_1^2$  and  $P_2^2$ .

$$(C2) \quad J^{(2)}((P_i^j, \lambda_i^j, \mu_j)_{i,j=1,2}) = \mu_1 J^{(1)}((P_1^i, \lambda_i^j)_{i=1,2}) + \mu_2 J^{(1)}((P_2^i, \lambda_i^j)_{i=1,2}) - J^{(1)}(\mu_1 P_1^i + \mu_2 P_2^i, \lambda_i^j)_{i=1,2} \geq 0$$

that is, the first Jensen difference  $J^{(1)}$  is a convex function on  $\mathcal{P}^2$ . One may define recursively  $J^{(k)}$  on  $\mathcal{P}^{2^k}$  and lay down a series of positivity conditions

$$(Ck): J^{(k)} \geq 0 \quad k=1, 2, \dots$$

**Definition :** A diversity measure  $H$  satisfying (C0) is said to be completely convex if

$$(a) J^{(0)} = -H$$

(b)  $J^{(k)}$  defined appropriately satisfies (Ck) for all integer  $k$ .

It is said to be convex of order  $j$  if (Ck) is satisfied for  $k = 1, 2, \dots, j$ .

The practical interest of such properties is twofold. First, from a confirmatory point of view, ANOVA extends in a simple way to the treatment of a much larger class of diversity measures. Furthermore (Rao (1984)) the order of convexity necessary to ensure that the  $J^{(k)}$  statistic representing the main effects and interaction between factors is equal to the number of factors taken into account in the general ANalysis Of DIvERSity (or ANODIV for short) table, in which the decomposition of the overall diversity parallels the decomposition of the total sum of squares in the usual ANOVA table. Then the ANODIV table can be used for inferential purposes with help of randomization tests to determinate significant effects.

This methodology clearly extends to more complex designs of experiment, such as factorial, hierarchical (e.g. when testing the adequacy of an a priori given hierarchical clustering), repeated measures designs, ... . More over this formalization unifies the methodology since the technical intricacies that usually differentiate the mathematical treatment of the case of univariate response variables from the from multivariate one, is

treated here as a separate problem. The determination of the test statistic depends in effect on the raw data only through the dissimilarity matrix  $\Delta$ . So problems of dimensionality occur when choosing a suitable dissimilarity function.

For instance, when the response variable is distributed in  $\Omega_i$  with mean  $\mu_i$  and the covariance matrix  $\Sigma$  and is not necessarily gaussian, Rao (1982) defines  $J^{(2)}$  as the Mahalanobis distance between populations:

$$J^{(2)}(P_i, P_k) = \delta_{ik}^2 = (\mu_i - \mu_k) \Sigma^{-1} (\mu_i - \mu_k)$$

Furthermore, the myelinated fibres data set has been analysed with help of the Gini-Simpson dissimilarity function which can be derived as a  $J^{(2)}$  Jensen difference from the kernel function

$$\kappa_{ij} = 1 - \sum_{k=1}^K \kappa_{ij}^k \quad \text{where} \quad \kappa_{ij}^k = 1$$

if the populations of fibres  $\phi_i$  and  $\phi_j$  agree (that is they belong to the same predetermined class) and 0 otherwise. Thus, when a rat is identified to the population of its myelinated fibres and therefore to the associated observation of a multinomial vector, one obtains by simple algebra

$$J^{(2)}(P_i, P_l) = \delta_{il}^2 = \frac{1}{2} \sum_{k=1}^K (p_i^k - p_l^k)^2$$

This diversity measure, denoted by  $\delta_{il}$ , has been retained here because it is Euclidean. As shown above this quality induces many optimalities and yields several crossvalidations with help of better known methods either for the inferential treatments of the data or the exploratory one used during the analysis and to be discussed in the remainder of this paper.

Nevertheless, whatever the practical power of this new methodology, the reader should keep in mind that the usual limitations of the ANOVA methodology remain true. In particular, the complexity of the design of experiment is of importance for the choice of a regular enough dissimilarity function and for the implementation of the randomization test procedure. Moreover the usual limitations about testable hypotheses induced by the number of necessary replications cannot be overpassed.

The second point of interest refers to the exploratory point of view. Since the kernel  $\delta$  of the function

$$H(P) = \int_{\Omega^2} \delta(\omega_1, \omega_2) P(d\omega_1)P(d\omega_2)$$

determinates the elements of the Hessian matrix of  $H$  whose concavity properties depend on the regularity properties of  $\delta$ .

**Definition :** A non negative function  $\delta(\cdot, \cdot)$  on  $\Omega^2$  is conditionally negative definite (CND for short ) if

$$\sum_{i=1}^n \sum_{j=1}^n \delta(\omega_i, \omega_j) a_i a_j < 0$$

for any  $n$  and any  $a \in \mathbb{R}^n$  such that

$$\sum_{i=1}^n a_i = 0 .$$

**Theorem 1 :** if  $H(P) = \int_{\Omega^2} \delta(\omega_1, \omega_2) P(d\omega_1)P(d\omega_2)$ , then  $J^{(0)} = -H$  iff  $\delta$  is CND.

Hence using the result of Rao (1982) about the regularity of dissimilarity functions, the ANODIV can be applied to any data set and to any design of experiment provided that the dissimilarity function be a CND function. This assumption is closely related to the MDS methodology which is concerned with the Euclidean embedding of dissimilarity data since

**Theorem 2** (Schoenberg (1935)): Let  $\delta$  be a non negative symmetric function on  $\Omega$ .

Then

(1)  $\delta$  is CND iff  $d = \delta^{1/2}$  is a metric, and there exist an isometric embedding of the metric space  $(\Omega, d)$  into a Hilbert space.

(2) if  $\delta$  is CND, so is  $\delta^\beta$  is CND for  $0 \leq \beta \leq 1$ .

Hence the MDS methodology provides us with a tool complementary to the ANODIV approach, and aimed at the Euclidean representation of the elements of  $\Omega$  associated to a latent dimensional decomposition of a quadratic entropy, that is to a concave entropy function satisfying the assumptions of theorem 1. This interpretation of the aims of the MDS is reminiscent of the justification of Correspondence Analysis of a two way contingency table as a method producing a dimensional decomposition of the chi-2 statistic, which is then interpreted as a negentropy function .

It yields too a method for the construction of a quadratic diversity function, resulting from the approximation of the raw dissimilarities by an Euclidean one embeddable in a vector space of known dimensionality.

### 3. THE EUCLIDEAN EMBEDDING OF DISSIMILARITY DATA

Let us consider the data structure  $(I, \delta, P)$  resulting from the observation of some dissimilarity function between pairs of elements of a finite set  $I$  of  $n$  elements, which may be interpreted as a sample drawn from  $\Omega$ . The dissimilarities  $\delta_{ij}$  form a square symmetric matrix of order  $n$ , denoted by  $\Delta$ . Here  $\delta$  is assumed to satisfy the condition (C0).  $P$  is a weight function defined on  $I$  such that  $p_i > 0$  ( $\forall i, i \in I$ ) and  $\sum p_i = 1$ .

$(I, \delta, P)$  is embeddable in a  $p$ -dimensional Euclidean space  $E_p$  if there exists a set of  $n$  points in  $E_p$  that generates the  $n(n-1)/2$  values  $\delta_{ij}$ . In such a case, the matrix  $Z$  whose columns are formed by the coordinates of the representative points in an arbitrary fundamental basis of  $E_p$  is called a configuration matrix and since  $\delta$  is an Euclidean distance in  $E_p$ , then we have

$$\delta_{ij} = \delta_{ij}(Z) = \delta(z_i, z_j)$$

$p$  may be large in practice, so one usually approximates  $\Delta = \Delta(Z)$  by a matrix  $D(X)$  of Euclidean distances derived from a configuration matrix  $X$  of lower rank  $r < p$  (often  $r=2$  or 3). Two traditions consider that problem.

The first one is born with the works of Young and Householder (1938) and Torgerson (1958) and is known in the biometric literature under the heading of Principal Coordinates Analysis, since the comprehensive treatment by Gower (1966). Central to this approach is the derivation of scalar products from the squared interpoint distances in an  $n$ -dimensional Euclidean vector space  $F$ , referred to its natural basis :

$$w_{ij} = \langle z_i, z_j \rangle = \frac{1}{2} |\delta_{ij}^2(Z) - \delta_{i.}^2(Z) - \delta_{.j}^2(Z) + \delta_{..}^2(Z)|$$

with

$$\delta_{i.}^2(Z) = \sum_{l=1}^n p_l \delta_{il}^2(Z) \quad \delta_{.j}^2(Z) = \sum_{i=1}^n p_i \delta_{ij}^2(Z)$$

In matrix terms,  $W = Q_1 H^{-1} Q_1$  where  $H = (-\frac{1}{2} \delta_{ij}^2)$  and  $Q_1 = I - 1^{-1} 1 N$  is the orthogonal projector on  $F_1$ , the subspace of  $F$  spanned by vectors orthogonal to the constant vector.  $N$  is the diagonal matrix defined by  $n_{ii} = p_i$  ( $\forall i, i \in I$ ). Then  $W$  is

interpreted as the matrix of a linear application from  $F^*$ , the algebraic dual space of  $F$ , into  $F$  and the  $r$ -dimensional configuration matrix  $X_r$  derives from the  $r$  first eigen-vectors (i.e. associated to the  $r$  largest eigenvalues) of the matrix  $WN$ .  $W$  is gramian and secures non negative eigenvalues for  $WN$  only when  $\Delta$  results from an error free measurement of the Euclidean distances  $\delta_{ij}(Z)$ . The solutions  $X_r$ ,  $r=1, 2, \dots, p$  are nested and satisfy an inner projection property resulting from the contracting properties of orthogonal projectors (d'Aubigny (1984), Meulman (1986)), namely :

$$d_{ij}^2(X_1) \leq d_{ij}^2(X_2) \leq \dots \leq d_{ij}^2(X_r) \leq \dots \leq d_{ij}^2(X_p) = \delta_{ij}^2(Z)$$

Thus for any probability measure  $P$ , the apportionment of any quadratic entropy function  $H$  into non negative terms

$$H(P) = H_r(P) + H_{p-r}(P) = \text{explained diversity} + \text{residual diversity},$$

leads to the following inequalities :

$$H_1(P) \leq H_2(P) \leq \dots \leq H_r(P) \leq \dots \leq H_p(P) = H(P)$$

where  $H_r(P)$  denotes the diversity explained by the representation matrix  $X_r$ . Since any decomposition of  $\delta(\omega_i, \omega_j)$  as a sum of  $K$  terms  $\delta_k(\omega_i, \omega_j)$  provides a decomposition of the entropy  $H = \sum_{k=1}^K H_k$ , an explicit expression of both components may be obtained from the

$K$  non null eigenvectors of  $WN$ . d'Aubigny (1987) gives a discussion and references about the properties and weaknesses of the approximate configuration  $X_r$ , when  $\Delta$  is non Euclidean so that  $W$  and  $WN$  are no more gramian.

The second approach directly approximates in a least squares sense the dissimilarities  $\delta_{ij}$  and so it integrates possible errors in the measurement of dissimilarities. Besides, it is not interpretable as a projection method in  $F$ . Thus it produces two-sided approximations of the dissimilarities in contrast to the inner approximation of the dissimilarity function that characterizes Principal Coordinates Analysis. Moreover, the extension of this approach to the analysis of ordinal dissimilarity functions, which differentiates pairs of statistical units on an ordinal scale only, is straightforward in that case. In fact the work of Shepard(1962) and Kruskal (1964) initiated this approach in the ordinal setting. d'Aubigny (1984) proposes an unifying algebraic formulation of the existing solutions which is based on a geometric interpretation of a result of Coxeter (1947) :

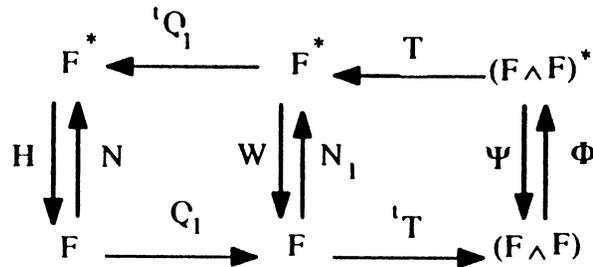
**Lemma :** *Let  $q$  be a real function on  $I \times I$  satisfying  $q_{ii} \geq 0$  and  $q_{il} = q_{li}$  for any  $(i, l)$  in  $I \times I$ . Then for any vector  $x$  in  $F_I$*

$$\alpha = \sum_i \sum_l q_{il} (x_i - x_l)^2 = 2 \left\{ \sum_i q_{i+} x_i^2 - \sum_i \sum_{k \neq i} q_{ik} x_i x_k \right\} \text{ where } q_{i+} = \sum_{k \neq i} q_{ik}$$

So we can write  $\alpha = 2 \|x\|_{N_Q}^2$  for  $x \in F_1$ , where  $N_Q = D_q - Q$  defines a semi-metric on  $F_1$  by :

$$n_{ij} = \begin{cases} -q_{ij} & \text{if } i \neq j \\ q_{i+} & \text{if } i = j \end{cases}$$

Since any distance  $d$  on  $F$ , derived from a gauge  $g$  as  $d_{ij}(X) = g(x_i - x_j)$ , is a function of the skew symmetric vectors  $x_i - x_j$ , the exterior product of  $F$  by itself denoted by  $F \wedge F$  is a natural space for describing the Euclidean embedding of the triple  $(I, \delta_I, Q_{|I|})$ , where  $Q$  is a weight function on  $|I|$ , as illustrated by the following duality diagram :



where the  $N$ -orthogonal projector  $Q_1$  on the subspace  $F_0$  of  $F$ ,  $N$ -orthogonal to the constant vector  $1_n$  verifies for any vector  $c$  in  $F_0$

$$\sum p_i c_i = 0 \Leftrightarrow Q_1 c = c .$$

$F \wedge F$  is a linear vector space with dimensionality equal to  $n(n-1) / 2$ , the metric  $\Phi = D_Q$  is diagonal with generic term  $q_{il}$  and  $T$  is the matrix of the restriction to  $F \wedge F$  of the simple contrasts operator  ${}^t T = S_n \otimes 1_n$   $S_n = (1_n : -1_n)$

$T$  is a linear application from  $(F \wedge F)^*$  onto  $F_1^*$  characterized by  $T(f^* \wedge f^*) = f^* - f^*$ , where  $\{(f^* \wedge f^*), i < l, (i, l) \in |I|\}$  forms a natural choice for a basis of  $F^* \wedge F^*$ , associated with the fundamental basis  $\{f^*, i \in I\}$  of  $F^*$ . Thus for any  $x$  in  $F_1$ ,  ${}^t T x = (x_i - x_l)_{i < l}$  if  $x = (x_i)_{i \in I}$  and using the notations  $Q = (q_{il})$ ,  $D_q = \text{diag}(q_{i+})$ ,  $D_Q = \text{diag}(q_{il})$ , we obtain

$$N_Q = D_q - Q = T D_Q {}^t T$$

So,  $N_Q$  defines a metric of rank  $(n-1)$  on  $F_1$ , when  $Q$  satisfies to the hypotheses of the Coxeter lemma. Using classical duality arguments, this choice of metrics makes  $T$  an partial isometry and describes a correspondence between Principal Coordinates Analysis (derivable in  $F$ ) and the more general analysis of the triple  $(I, \delta_I, Q_{|I|})$  (described in  $F \wedge F$ ) proposed by Kruskal (1964). This author minimizes the STRESS index, written in  $F \wedge F$  as:

$$\text{STRESS} = \| \delta - d(X) \|_{D_C}^2$$

This MDS problem is clearly not equivalent to the principal coordinates one. In fact, the former approximates dissimilarities, that is the topological properties of the space of representation of the statistical units, in an Euclidean affine space of a priori fixed dimensionality, while the latter provides biased estimations of these informations through a reconstitution of angles (d'Aubigny 1987). The numerical optimization of the STRESS function usually implements some gradient method, but since STRESS is not differentiable as soon as one  $d_{ij}(X)$  equals zero, a more satisfactory algorithm, using majorization arguments follows de Leeuw (1977) and de Leeuw and Heiser (1980).

The quadratic entropy  $H(P)$  associated to the STRESS function generalizes the definitions given in section 2, since now, the weights are not limited to the special case  $Q = p \otimes p$ . In this later case, we have  $q_{ij} = p_i p_j$ ,  $N$  the covariance matrix of the multinomial vector  $p$  satisfying  $n_{ii} = p_i(1-p_i)$  if  $i=j$  and  $n_{ij} = -p_i p_j$  otherwise, and then the constituents of the PCA and MDS solutions are related by the following theorem, in which  $\Delta * \Delta = (\delta_{ij}^2)$ .

**Theorem :** *If  $Q = p \otimes p$  and  $H = -\frac{1}{2} \Delta * \Delta$ , then any centered eigenvector of  $HN_Q$  is an eigenvector of  $WN$ .*

**Proof :** the eigenequation  $\lambda c = HN_Q c$  implies

$$\lambda c = \lambda Q_1 c = Q_1 HN_Q c = Q_1 H Q_1 N c = W N c$$

since the special form of  $N$  here implies  $Q_1 N = N$ .

By the classical duality arguments the partial isometry  $T$  associates to the matrix  $W$  the symmetric semi-definite positive matrix  $\Psi = {}^t T W T$  with generic term

$$\Psi_{il, i'l'} = \begin{cases} -\delta_{il}^2 & \text{if } (i,l) = (i',l') \\ -\frac{1}{2} [\delta_{ii'}^2 + \delta_{ll'}^2 - \delta_{il}^2 - \delta_{i'l'}^2] & \text{otherwise} \end{cases}$$

so that  $\text{Trace}(WN_Q) = \text{Trace}(\Phi\Psi) = \sum_{i,l} q_{il} \delta_{il}^2$

Hence, the dimensional decomposition of the entropy function associated to the kernel  $\delta$ , using Principal Coordinates Analysis yields too a factorization of the elements of  $\psi$ . One recognizes in  $\Psi_{il, i'l'}$  a 2x2 contrast associated with the following additive decomposition of the two-way table  $\Delta$

$$\delta_{il} = \mu_{il} + \epsilon_{il} \quad \mu_{il} = \mu + \alpha_i + \beta_l + \gamma_{il} \quad (M1)$$

in which the parameters are assumed to satisfy the identification constraints usually imposed to the parameters in ANOVA. Thus, (Johnson and Graybill (1972) theorem 4.1), the eigen-vector  $x$  associated to the largest eigenvalue of  $WNWN$  realizes the maximum of

$$\left[ \sum_i \sum_l u_i u_l \delta_{il} \right]^2 \quad \sum_i p_i u_i = 0 \quad \text{with } N_Q u = 1$$

is clearly associated to the largest eigenvalue of  $WN_Q$  in absolute value (d'Aubigny (1984)). So the Principal Coordinates Axes may be interpreted in that setting as a sequence of optimal contrasts ordered in terms of decreasing variance, and describing the interaction term of the model (M1). Moreover if the configuration matrix  $X$  is such that  $W = {}^t X M X$  for some metric  $M$  given a priori, then the  $2 \times 2$  contrasts are directly readable on the resulting mapping (with the usual cautions linked with quality of representation problems for objects in a summary representation space of low dimensionality), since then

$$\psi_{il, i'l'} = {}^t(x_i - x_l) M (x_{i'} - x_{l'}).$$

Most often  $M$  is taken equal to the identity and then, any  $2 \times 2$  contrast equal to zero is represented on mappings by two orthogonal segments. So that the less null  $2 \times 2$  contrasts we have, that is the stronger is the interaction potential derived from the dissimilarity matrix  $\Delta$ , and the less dimensions are necessary in a way to represent the statistical units in an Euclidean space.

Moreover, each combination of factors in a design of experiment determinates a partition of  $I$ . So if one considers the incidence matrix  $Q$  associated to the partition graph representing this combination, with  $q_{il} = 1$  if units  $i$  and  $l$  belong to the same class  $k$  ( $k = 1, \dots, K$ ) and 0 otherwise,  $\Phi$  is then a diagonal matrix  $\Phi = \text{diag}(q_{il})$ , so that

$$\text{Trace}(\Phi\Psi) = \sum_{i,l} q_{il} \delta_{il}^2 = \sum_{k=1}^K \sum_{i,l \in k} \delta_{il}^2$$

This expression receives a straightforward interpretation in the univariate ANOVA problem, since then  $\text{Trace}(\Phi\Psi)$  is the Within-class Sum of Squares, when  $\delta_{ij}^2(Y) = (y_i - y_j)^2$ . It shows too that the permutation of rows and columns of  $D$  is not necessary to evaluate the Within class Diversity. Besides the Between class diversity is obtained by complementarity with help of the matrix  $\Phi^c = \text{diag}(1 - q_{il})$ , and the case of weighted statistical units follows directly in considering the product matrix

$$\Phi = D_Q - D_{p \otimes p}$$

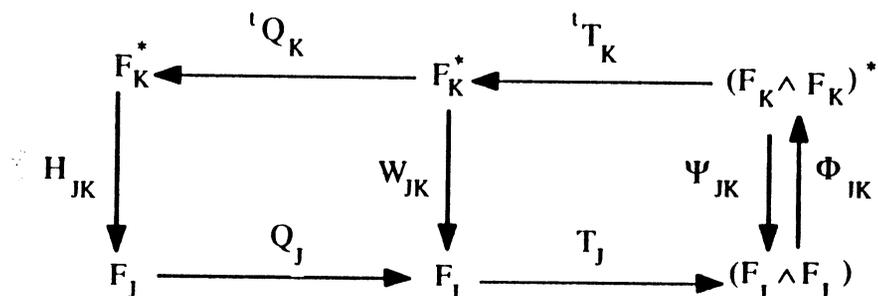
The interest of these results is twofold in practice. First they justify the application of Principal Coordinates Analysis to  $\Delta$  in the ANOVA realm, where  $H$  is derived from the kernel function  $\delta^2$ . In doing so one is given a powerful exploratory tool, which complements those discussed in section 2. On the other hand, a broader generality is obtained for the analysis of dissimilarity functions in common use in biology, especially with reference to incidence data (GOWER (1984)). As a matter of fact, many of them are semi-distances of  $L_1$ -type (Fichet (1987)). Le Calve (1987) shows that the square root of those semi-distances  $\sqrt{\delta}$  is an Euclidean dissimilarity. Then a direct application of Principal Coordinates Analysis to the kernel  $\sqrt{\delta}$  provides a dimensional decomposition of the associated quadratic entropy function and representations axes interpretable in terms of orthogonal optimal contrasts, in a way to describe the potential of interaction summarized by this diversity function.

This use of  $2 \times 2$  contrasts is rather special, since  $\Delta$  is identified to a two-way cross-tabulation. It is used for the additive decomposition of the dissimilarity function associated to a description of a peculiar term of interaction, since it defines the general term of a Gramian matrix in the Euclidean case. Now, this method extends nicely to an actual two factors setting. Then, we are given a dissimilarity matrix  $\Delta_{JK}$  the general term of which  $\delta_{jk}$  measures the strength of dissimilarity of the object  $k$  to some reference point subjectively fixed by a subject  $j$ ,  $j \in J$  and  $k \in K$ . The Euclidean representation of this kind of data matrix is known as the Unfolding problem in the psychometric literature, see e.g. KRUSKAL and WISH (1978) or SCHONEMANN (1972).

Let us consider the two simple contrasts matrices  $T_J$  and  $T_K$  defined respectively on  $J$  and  $K$  by analogy to the definition of  $T$  given above. Then the  $2 \times 2$  contrasts

$$\Psi_{jj',kk'} = -\frac{1}{2} \left[ (\delta_{jk}^2 - \delta_{jk'}^2) - (\delta_{j'k}^2 - \delta_{j'k'}^2) \right]$$

are the elements of the matrix  $\Psi_{JK}$  which is one element of the duality diagram



where  $Q_J$  and  $Q_K$  are the orthogonal projectors on the orthogonal supplementary subspace of the space of constant functions in  $F_J$  and  $F_K$  respectively and

$$H_{JK} = \left( -\frac{1}{2} \delta_{jk}^2 \right)_{j \in J, k \in K}$$

$$W_{JK} = Q_J H_{JK} Q_K = \left( h_{jk} - h_{j.} - h_{.k} + h_{..} \right)_{j \in J, k \in K}$$

So,  $W_{JK}$  may be interpreted in the ANOVA setting as the matrix of the estimates of the interactions between the factors  $J$  and  $K$ , associated to the model

$$h_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \epsilon_{ijk}, \quad \epsilon_{ijk} \text{ i.i.d. with } E(\epsilon_{ijk}) = 0 \text{ and } \text{Var}(\epsilon_{ijk}) = \sigma^2$$

and we suppose here that we have only one replication. Thus the preceding results interpret the spectral analysis of  $W_{JK}$  as a dimensional decomposition of the dissimilarity matrix  $\Delta_{JK}$  in terms of optimal contrasts aimed at the description of the  $2 \times 2$  contrasts contained in  $\Psi_{JK}$ . But this spectral analysis is clearly equivalent to the algebraic resolution of the metric Unfolding problem proposed by SCHONEMANN (1972), which receives here a statistical interpretation. The  $2 \times 2$  contrast  $\Psi_{jj',kk'}$  has a straightforward interpretation here, since it measures the difference in the judgments of the subjects  $j$  and  $j'$  for the paired comparison of the objects  $k$  and  $k'$ .

Moreover, this formalization of the Unfolding problem may be obtained on other grounds, when considering  $J$  and  $K$  as the constituents of a bipartition of the set  $I$ . Then one may write

$$\Delta_{II} = \begin{bmatrix} \Delta_{JJ} & \Delta_{JK} \\ \Delta_{KJ} & \Delta_{KK} \end{bmatrix}$$

This means that  $\Delta_{JK}$  is an off-diagonal dissimilarity matrix, as usually said in the Multidimensional Scaling literature, with the informations  $\Delta_{JJ}$  and  $\Delta_{KK}$  missing. In the ANOVA setting this point is interpreted in considering either the one way incomplete design of experiment on  $I$  or the complete two-way factorial design that crosses the two factors  $J$  and  $K$ .

#### 4. ORDINATION METHODS WITH INSTRUMENTAL VARIABLES

The above methodology is strongly complemented by the introduction of constraints for the treatment of dissimilarity data. This problem justified numerous publications in the realm of the MDS methodology, since the work of Bloxom (1978), and the above algebraic tools offer an unifying formalism (d'Aubigny (1987)), for the integration of instrumental variables, or with more generality of instrumental dissimilarity

relations. These constraints can be measured on a nominal scale, for instance in a planned experiment, when defining the incidence matrix (or the derived dissimilarity matrix) associated to some design of experiment in the case of a dissimilarity response variable. In biology, these instrumental variables are often quantitative measurements as is the case for vegetation dissimilarity data collected jointly with pedological characteristics of the soil.

This methodology receives a comprehensive treatment in d'Aubigny (1987) and is illustrated here in the simplest case of graph constraints, because the treatment of partition graphs specializes to an exploratory complement of the inferential approach discussed above, that mimics for dissimilarity data, the dual use of linear discriminant analysis and multivariate ANOVA approaches.

#### 4.1. *The Discriminant Coordinates Analysis*

Let us consider the partition  $I = \bigcup_{k=1}^K I_k$  of the set of statistical units, with  $\text{card}(I_k) = n_k$ .

When the data are measurements of  $p$  variables on  $I$ , one usually applies a Linear Discriminant Analysis (or Canonical Coordinates Analysis) methodology in a way to get a representation of the data in an Euclidean space of low dimensionality. This classical add on to a multivariate ANOVA considers two arrays. The  $p$  quantitative measurements define a matrix  $X$ , which is  $p \times n$ , and the design is described by an array of dummy (or indicator) variables  $Y$ , of order  $K \times n$ . Then Linear Discriminant Analysis is interpreted either in terms of Canonical Correlation Analysis applied to  $(X, Y)$  or equivalently as an ACPVI of  $X$  on the Instrumental variables  $Y$ . Both interpretations lead to the same solution, which is derived (see for example Cailleux and Pages (1976)) from the spectral decomposition of the triple  $(V_{22}, D_K^{-1}, V_{11}^{-1})$ , where

$$V_{21} = YN^t X$$

$$V_{11} = XN^t X$$

$V_{22} = YN^t Y \Rightarrow V_{22} = D_K = \text{diag}(p_k)$ .  $p_k$  is the relative weight of the class  $k$ , and  $N = D_p$  is a matrix of weights a priori given to the statistical units.

$$p_k = \sum_{i \in I_k} p_i$$

Moreover, one simply shows that the  $k$ -th row of the matrix  ${}^t B = V_{22}^{-1} V_{21} = D_K^{-1} V_{21}$  yields the coordinates of the centroid of the  $k$ -th category. So it is a classical result that the Principal Component Analysis of the triple  $(B, D_K, V_{11}^{-1})$  yields an equivalent way to

derive the solution of the Linear Discriminant Analysis problem . See Caillez and Pages (1976) for more details about these correspondences.

Here, we adopt an alternative and less well known formulation of the Linear Discriminant Analysis problem in terms of Generalized Canonical Analysis, applied on vectors representing the statistical units and not the variables. So, one speaks of the Q version of this method. While this formulation of the problem is often ignored, it is of interest here because it leads to an extension of the methodology to dissimilarity data.

Let us consider the following decomposition of the data matrix in K sub-tables :

$$Z = (Z^1 : \dots : Z^K) \quad Z^k \in \mathcal{M}_{p, n_k}(\mathbb{R})$$

based on the array Z derived from X by a permutation of columns that regroups the elements of each class  $I_k$  . From an algebraic point of view, Z describes a linear mapping  $F^* \rightarrow E$ . So when the vector space F is decomposed as the direct sum

$$F = \bigoplus_{k=1}^K F_k, \quad \dim(F_k) = n_k, \quad \sum_{k=1}^K n_k = n = \dim(F), \quad \dim(E) = p,$$

For any fundamental basis of F which respects this decomposition:  $\mathfrak{B} = \bigcup_{k=1}^K \mathfrak{B}_k$  , with

$\text{card}(\mathfrak{B}_k) = n_k$ , one is able to write any metric N, defining an Euclidean geometry on F as  $N = (N_{kk'})_{k=1, \dots, K, k'=1, \dots, K}$  .

Then one writes the usual equations of the  $\Sigma$ -strategy in Canonical Correlation Analysis (cf. CAILLEZ et PAGES (1976)) specified here by a direct sum decomposition of F, and not of E as usual. As a consequence, one gets the following decomposition of the sample covariance matrix of Z:

$$V = ZN^tZ = \sum_{k=1}^K Z_k N_k^t Z_k$$

when  $N_k$  is a block diagonal matrix  $N_k = (N_{kk})$ , and the sample connexion matrix

$$W = {}^tZ M Z \text{ is}$$

$$W = (W_{kk'})_{kk'} \text{ where } W_{kk'} = {}^tZ_k M Z_{k'}$$

As before it is associated with the array of squared distances  $H = (-\frac{1}{2} d_{ij}^2(Z))$  by the equation

$$W = Q_1 H {}^tQ_1$$

Besides, let us denote by  $\Delta_{1_k}$  the subspace of  $F_k$  containing the vectors with constant coordinates in  $F_k$ . When  $N_k = p_k Q_K$ , with  $Q_K$  the  $D_K$ -orthogonal projector on the subspace of  $F_k$ , which is the  $D_K$ -orthogonal supplementary of  $\Delta_{1_k}$ , we have:

$$\begin{aligned} D_K &= \text{diag}\left(\frac{p_i}{p_K}\right) \\ Q_K &= I_k - \frac{1_k}{n_k} \frac{1_k}{n_k} D_K \\ Z_k N_k {}^t Z_k &= p_k Z_k Q_K {}^t Z_k = p_k V_k \end{aligned}$$

where  $V_k$  is the sample covariance matrix derived from  $Z$  in the class  $k$ . Consequently,  $V$  is the sample Intra-class covariance matrix of  $Z$ . Thus using the metric  $M = T^{-1}$ , that is the inverse of the sample Total covariance matrix of  $Z$ :

$$T = Z Q_1 {}^t Z \quad Q_1 = I_n - \frac{1_n}{n} \frac{1_n}{n} D_P$$

leads to the usual equations of Linear Discriminant Analysis, when interpreted as a Principal Component Analysis of the triple  $(Z, T^{-1}, W)$ :

$$\begin{aligned} T^{-1} W \underline{u} &= \lambda \underline{u} \\ \|\underline{u}\|_W^2 &= 1 \quad \text{and} \quad \langle \underline{u}^s, \underline{u}^t \rangle = 1 \text{ if } s = t, \text{ and } 0 \text{ otherwise.} \end{aligned}$$

Now, let us consider an Euclidean dissimilarity function  $\delta(Z)$  observed on the partitioned set  $I = \bigcup_{k=1}^K I_k$ , with  $\text{card}(I_k) = n_k$ , and note  $\underline{b}_k$  the vector of coordinates of the centroid of the class  $k$ .

**Proposition :** *The derivation of the squared Euclidean distances  $d_{kk'}^2(Z)$  separating two centroids  $k$  and  $k'$  depends only on the observed dissimilarities  $\Delta(Z)$ , on the partition graph  $G$ , and on the weighting of the statistical units  $P_1$ .*

**Proof :** By the law of cosines, one has

$$d_{kk'}^2(Z) = \|\underline{b}_k - \underline{b}_{k'}\|_M^2 = \|\underline{b}_k\|_M^2 + \|\underline{b}_{k'}\|_M^2 - 2\langle \underline{b}_k, \underline{b}_{k'} \rangle_M \quad (1)$$

and since  $\underline{b}_k$  is the centroid of the weighted points  $\{(z_i, p_i), i \in I_k\}$ , it satisfies

$$\underline{b}_k = \frac{Z D_P \underline{y}_k}{{}^t \underline{y}_k D_P \underline{y}_k} = \frac{Z_k D_k \underline{1}_k}{{}^t \underline{1}_k D_k \underline{1}_k}$$

and

$${}^t \underline{y}_k D_P \underline{y}_k = {}^t \underline{1}_k D_k \underline{1}_k = p_k$$

where  $\underline{y}_k$  is the  $k$ -th column of the indicator matrix  ${}^t Y$  derived from the partition. Moreover

$$y_k = (0, \dots, \underline{1}_k, 0, \dots, 0) \Leftrightarrow \underline{1}_k \otimes e_k = y_k \quad e_k \in \mathbb{R}^K, \quad e_k = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \leftarrow k\text{-th row}$$

In an analogous way the overall centroid  $\underline{b}$  of the weighted points  $\{(z_i, p_i), i \in I\}$  may be written, with the normalization chosen for  $P_i$ :

$$\underline{b} = \frac{ZD_P \underline{1}_n}{\sum \underline{1}_n D_P \underline{1}_n} = ZD_P \underline{1}_n$$

So

$$\langle \underline{b}_k, \underline{b}_{k'} \rangle_M = \frac{1}{p_k p_{k'}} \sum y_k D_P \sum Z M Z D_P y_{k'} = \frac{1}{p_k p_{k'}} \sum y_k D_P W_Z D_P y_{k'}$$

Now, since  $H(Z) = -\frac{1}{2} \Delta(Z) * \Delta(Z)$  and  $W_Z = Q_1 H_Z Q_1$ :

$$\begin{aligned} Q_1 D_P y_k &= D_P Q_1 y_k = D_P (I - \underline{1}_n \underline{1}_n^T D_P) y_k \\ &= D_P y_k - p_k \underline{1}_n \end{aligned}$$

as a consequence,

$$\begin{aligned} \langle \underline{b}_k, \underline{b}_{k'} \rangle_M &= \frac{1}{p_k p_{k'}} \sum (D_P y_k - p_k \underline{1}_n) H_Z (D_P y_{k'} - p_{k'} \underline{1}_n) \\ &= \frac{1}{p_k p_{k'}} \sum D_P y_k H_Z D_P y_{k'} - \frac{1}{p_k} \sum \underline{1}_n H_Z D_P y_{k'} - \frac{1}{p_{k'}} \sum y_k D_P H_Z \underline{1}_n + \sum \underline{1}_n H_Z \underline{1}_n \end{aligned}$$

let us note too  $p_k = D_P y_k$ . Then

$$\langle \underline{b}_k, \underline{b}_{k'} \rangle_M = \frac{1}{p_k p_{k'}} \sum p_k H_Z p_{k'} - \frac{1}{p_k} \sum \underline{1}_n H_Z p_{k'} - \frac{1}{p_{k'}} \sum p_k H_Z \underline{1}_n + \sum \underline{1}_n H_Z \underline{1}_n \quad (2)$$

and after some simplifications, substituting the terms (2) in (1) gives

$$d_{kk'}^2(B) = \frac{\sum p_k H_Z p_k}{p_k^2} + \frac{\sum p_{k'} H_Z p_{k'}}{p_{k'}^2} - 2 \frac{\sum p_k H_Z p_{k'}}{p_k p_{k'}} \quad (3)$$

which yields the announced result. It may be done more explicit in writing  $\underline{a}_k = \frac{1}{p_k} p_k$  and

$$\begin{aligned} d_{kk'}^2(B) &= \|\underline{a}_k\|_{H_Z}^2 + \|\underline{a}_{k'}\|_{H_Z}^2 - 2 \langle \underline{a}_k - \underline{a}_{k'} \rangle_{H_Z}^2 \\ &= \|\underline{a}_k - \underline{a}_{k'}\|_{H_Z}^2 \end{aligned} \quad (4)$$

(3) may be given an analytic expression

$$d_{kk'}^2(B) = -\frac{1}{2} \left[ \frac{1}{2} \sum_{i \in I_k} \sum_{j \in I_k} p_i p_j d_{ij}^2(Z) + \frac{1}{2} \sum_{i \in I_{k'}} \sum_{j \in I_{k'}} p_i p_j d_{ij}^2(Z) - \frac{2}{p_k p_{k'}} \sum_{i \in I_k} \sum_{j \in I_{k'}} p_i p_j d_{ij}^2(Z) \right] \tag{5}$$

that clearly demonstrates that only the knowledge of the  $p_i$ 's and  $d_{ij}^2(Z)$ 's is necessary to evaluate these distances. Besides writing the last equation as

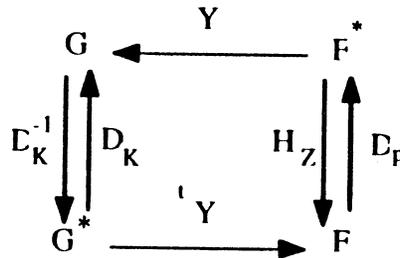
$$d_{kk'}^2(B) = w_{kk}(B) + w_{k'k}(B) - 2 w_{kk'}(B)$$

where

$$w_{kk'}(B) = \frac{{}^t p_k H_Z p_{k'}}{p_k p_{k'}}$$

is the generic term of the matrix  $W(B) = D_K^{-1} Y D_P H_Z D_P {}^t Y D_K^{-1}$  associated to the linear application

$w(B) : G \rightarrow G^*$ . One may get an insight into the statistical interpretation of  $W(B)$  by studying the components of the following duality diagram



where  $V_Y = Y D_P {}^t Y = D_K$ , since  $W(B) = V_Y^{-1} Y D_P H_Z D_P {}^t Y V_Y^{-1}$ . So one could write  $W(B) = \widehat{H}_{Z,Y}$ , in a way to indicate the relation of  $W(B)$  to the projection of  $H_Z$  on the linear space span by the partition graph. In fact, with  $P_2 = {}^t Y D_K^{-1} Y D_P$  the projector on the subspace of  $F$  spanned by  $Y$ :

$${}^t Y W(B) Y = P_2 H_Z {}^t P_2$$

This equation does not yield the matrix used in the classical resolution of the ACPVI problem defined by  $X$  and  $Y$ . This one uses in fact (see d'Aubigny (1984)):

$$\widehat{W} = P_2 W_Z {}^t P_2 = P_2 Q_1 H_Z {}^t Q_1 {}^t P_2$$

Nevertheless, this is an equivalent problem if  $P_2$  is a projection matrix on a subspace of  $F_1$ .

**Proposition:** *The Solution of the Discriminant Coordinates Analysis problem is given by the one of the two equivalent spectral decompositions:*

$$(P1) \quad W(B) D_K \underline{u} = \lambda \underline{u}, \quad \underline{u}_s D_K \underline{u}_t = 1 \text{ if } s = t, \text{ and } 0 \text{ otherwise}$$

$$(P2) \quad W_{KK} D_K \underline{u} = \lambda \underline{u}, \quad \underline{u}_s D_K \underline{u}_t = 1 \text{ if } s = t, \text{ and } 0 \text{ otherwise}$$

**Proof:** since

$$\langle \underline{b}_k, \underline{b}_{k'} \rangle_M = \frac{{}^t \underline{p}_k D_P W_Z D_P \underline{p}_{k'}}{p_k p_{k'}}$$

let us define the matrix of scalar products of centroids by  $W_{KK} = D_K^{-1} Y D_P W_Z D_P {}^t Y D_K^{-1}$ . Then the application of the law of cosines to  $D(Z)$  gives

$$d_{ij}^2(Z) = w_{ii}(Z) + w_{jj}(Z) - 2w_{ij}(Z)$$

and  $d_{kk}^2(B)$  is expressed by (5) where one substitutes  $w_{ij}(Z)$  for  $d_{ij}^2(Z)$ . It means that the Principal Coordinates Analysis of the matrix of Euclidean distances between centroids may be treated in two equivalent ways with help of a spectral analysis either of the matrix  $W(B)D_K$ , or of the matrix  $W_{KK}D_K$ .

The relation to the ACPVI problem is then expressed by the identities:

$$\text{Trace}(\widehat{W}D_P) = \text{Trace}({}^t Y W_{KK} Y D_P) = \text{Trace}(W_{KK} Y D_P {}^t Y) = \text{Trace}(W_{KK} D_K)$$

Now, let us consider  $G=(I,U)$  a non oriented simple graph (see Berge (1973) for definitions), with  $u_{ik} = 1$  if and only if the statistical units  $\omega_i$  and  $\omega_k$  are related by a constraint, and let  $N_Q$  be the symmetric non negative definite matrix derived as above from the connexion matrix  $Q$  of this graph, with generic term  $q_{ik} = 1$  if  $(i,k)$  belongs to  $U$  (that is  $u_{ik} = 1$ ) and 0 otherwise. Then d'Aubigny (1987) shows that the spectral decomposition of  $W N_Q$  results in a local ordination method similar to the Local Principal Components Analysis approach advocated by Lebart (1969). Due to the identity of equations involved in the case of an Euclidean dissimilarity function  $\Delta(Z)$  with those of the Local Principal Coordinates Analysis under the constraints defined by a partition graph is equivalent to the Discriminant Coordinates Analysis problem defined by  $\Delta(Z)$  and  $G$ .

#### 4.2. The Representation of the statistical units in Discriminant Coordinates Analysis

In Linear Discriminant Analysis, the visual evaluation of the dispersion of the statistical units of each class around its centroid is straightforward, since the interpretation of the method in terms of ACPVI indicates that their representative vector should be projected on the summary representation space obtained in the above section. This technique is often called the supplementary (or illustrative) points method (see Lebart et

al. (1984)). Its algebraic formulation is adapted to the Principal Coordinates Analysis problem in Mailles (1979), while in that case too, a solution has been derived by analytic arguments in Gower (1968) and Wilkinson (1970) among others. d'Aubigny (1984) shows that both solutions are equivalent. Here the formulation of Mailles is used, and we derive its specific application to the representation of the statistical units in Discriminant Coordinates Analysis.

As usual in Principal Component Analysis, let us interpret the supplementary points as statistical units with weight zero, and assume that only  $K$  statistical units among  $p = n+K$ , play an active part in the construction of the solution, the weight function is represented in  $F$  by the diagonal matrix

$$D_p = \begin{bmatrix} D_K & 0 \\ 0 & 0 \end{bmatrix}$$

the order of  $D_p$  (resp.  $D_K$ ) is  $p$  (resp.  $K$ ) and the  $D_p$ -orthogonal projector on  $F_1$  is now

$$Q_1 = I_p - \mathbf{1}_p \mathbf{1}_p' D_p = \begin{bmatrix} Q_K & 0 \\ -A & I_n \end{bmatrix}$$

where

$$Q_K = I_K - \mathbf{1}_K \mathbf{1}_K' D_K$$

$$A = \mathbf{1}_p \mathbf{1}_K' D_K$$

So the following partition of  $H$  and  $W$  :

$$H = \begin{bmatrix} H_{11} & H_{12} \\ \mathbf{1}' H_{12} & H_{22} \end{bmatrix} \quad W = Q_1 H \mathbf{1}' Q_1 = \begin{bmatrix} W_{11} & W_{12} \\ \mathbf{1}' W_{12} & W_{22} \end{bmatrix}$$

gives, with elementary algebra:

$$W_{11} = Q_K H_{11} \mathbf{1}' Q_K$$

$$W_{12} = Q_K H_{12} - Q_K H_{11} \mathbf{1}' A$$

$$W_{22} = A H_{11} \mathbf{1}' A + H_{22} - A H_{12} - H_{21} \mathbf{1}' A$$

The coordinates of the  $K$  active units and of the  $n$  supplementary ones are solutions of the Principal Coordinates Analysis equations, with  $\Lambda$  the matrix of eigenvalues (supposed distinct here for simplicity), and  $C$  is the orthogonal matrix whose columns are the corresponding eigenvectors of  $W D_p$ .

**Proposition:** *the coordinates of the supplementary units are given by the columns of the matrix*

$$\mathbf{1}' C_2 = \Lambda^{-1} \mathbf{1}' C_1 D_K W_{12}$$

**Proof:**

$$\begin{aligned}
 WD_p C = CA &\Leftrightarrow WD_p \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \Lambda \\
 &\Leftrightarrow \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix} \begin{bmatrix} D_K & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \Lambda \\
 &\Leftrightarrow \begin{cases} W_{11} D_K C_1 = C_1 \Lambda \\ W_{21} D_K C_1 = C_2 \Lambda \end{cases} \\
 &\Leftrightarrow \begin{cases} W_{11} D_K C_1 = C_1 \Lambda \\ W_{21} D_K C_1 \Lambda^{-1} = C_2 \end{cases} \quad (6)
 \end{aligned}$$

So we get the solution by equation (6).

In the case of the Discriminant Coordinates Analysis, we take the  $K$  centroids as the active statistical units, while the  $n$  original statistical units are treated as supplementary units in the analysis of the triple  $(K, \delta_K, P_K)$ . As a consequence,  $W_{12} = Q_K H_{12} - Q_K H_{11} {}^t A$  is derivable from the sub-blocks  $H_{11}$  and  $H_{12}$  of  $H$ , the expression of which is specific to the problem at hand. Now, we have by definition

$$\begin{aligned}
 H_{12} &= (d_{ki}^2(Z)) = (\|z_i - \underline{b}_k\|_{D_p}^2) \\
 d_{ki}^2(Z) &= \|z_i - \underline{b}\|_{D_p}^2 + \|\underline{b}_k - \underline{b}\|_{D_p}^2 - 2 \langle z_i, \underline{b}_k \rangle_{D_p} \\
 &= (W_{II})_{ii} + (W_{KK})_{kk} - 2(W_{KI})_{ik} \\
 \Leftrightarrow H_{12} &= \text{diag}(W_{KK}) \underline{1}_K {}^t \underline{1}_n + \underline{1}_K {}^t \underline{1}_n \text{diag}(W_{KK}) - 2W_{KI}
 \end{aligned}$$

In the same way, one has

$$H_{11} = \text{diag}(W_{KK}) \underline{1}_K {}^t \underline{1}_K + \underline{1}_K {}^t \underline{1}_K \text{diag}(W_{KK}) - 2W_{KK}$$

where

$$W_{KK} = {}^t B M B$$

and

$$\begin{aligned}
 B &= Z D_p {}^t Y D_K^{-1} \\
 W_{KK} &= D_K^{-1} Y D_p {}^t Z M Z D_p {}^t Y D_K^{-1} \\
 &= D_K^{-1} Y D_p W_Z D_p {}^t Y D_K^{-1} \\
 &= D_K^{-1} Y D_p Q_1 H_Z {}^t Q_1 D_p {}^t Y D_K^{-1}
 \end{aligned}$$

$$W_{KI} = {}^t B M Z = D_K^{-1} Y D_p {}^t Z M Z = D_K^{-1} Y D_p W_Z = D_K^{-1} Y D_p Q_1 H_Z {}^t Q_1$$

So,

$$W_{12} = Q_K [ \text{diag}(W_{KK}) \underline{1}_K ({}^t \underline{1}_n - {}^t \underline{1}_K {}^t A) - 2W_{KI} + 2W_{KK} {}^t A ]$$

But,

$$\underline{1}_n - A\underline{1}_K = \underline{1}_n - \underline{1}_K {}^t\underline{1}_n D_K \underline{1}_K = \underline{1}_n - \underline{1}_n = 0$$

and  $W_{12} = -2 [W_{KI} - W_{KK}{}^tA]$ . Now,

$$\begin{aligned} W_{KK}{}^tA &= D_K^{-1} Y D_p W_Z D_p {}^t Y D_K^{-1} D_K \underline{1}_K {}^t\underline{1}_n \\ &= D_K^{-1} Y D_p W_Z D_p {}^t Y \underline{1}_K {}^t\underline{1}_n = D_K^{-1} Y D_p W_Z D_p \underline{1}_n {}^t\underline{1}_n \end{aligned}$$

so that

$$\begin{aligned} {}^tC_2 &= \Lambda^{-1} {}^tC_1 D_K [W_{KI} - D_K^{-1} Y D_p W_Z D_p \underline{1}_n {}^t\underline{1}_n] \\ &= \Lambda^{-1} {}^tC_1 D_K [D_K^{-1} Y D_p W_Z - D_K^{-1} Y D_p W_Z D_p \underline{1}_n {}^t\underline{1}_n] \\ &= \Lambda^{-1} {}^tC_1 Y D_p W_Z (I - D_p \underline{1}_n {}^t\underline{1}_n) \\ &= \Lambda^{-1} {}^tC_1 Y D_p W_Z {}^tQ_1 = \Lambda^{-1} {}^tC_1 Y D_p Q_1 H_Z {}^tQ_1 \end{aligned} \quad (7)$$

Thus we have the

**Proposition:** *The coordinates of the projections of the statistical units in the linear vector space spanned by the centroids are the solutions of*

$${}^tC_2 = \Lambda^{-1} {}^tC_1 Y D_p Q_1 H_Z {}^tQ_1.$$

*Their derivation depends only on the matrices  $H_Z$ ,  $D_p$  and  $Y$ , which are known informations in the analysis of dissimilarity data.*

In a closely related paper, Digby and Gower (1981) give an analytic solution of the problem of representation of the statistical units in the representation space of the centroids of classes. In fact the equation (5) is common to both derivations and the solution of Digby and Gower may be shown to be equivalent to our equation (7). By comparison, our derivation assumes the existence of a weight function on  $I$ , and makes easier the comparison of Discriminant Coordinates Analysis (or DCA) to the usual treatment of the Linear Discriminant Analysis (or LDA) problem. DCA parallels LDA as closely as Principal Coordinates Analysis parallels Principal Component Analysis.

Moreover, Discriminant Coordinates Analysis is explicitly characterized here as a specific ACPVI. In fact, this relationship to ACPVI extends nicely to a wider class of instrumental variables problems (see d'Aubigny (1987)). Let us give some examples of related interest.

First, when considering the complementary graph, characterized by  $q_{ij}^c = 1 - q_{ij}$   $\forall (i,j) \in I \times I$ , one obtains an additive decomposition  $N_Q + N_{Q^c} = Q_I$  that yields a simple way to partial out a set of instrumental variables. In that case too, the matrix  $T$  (see section 3 above for a definition) is interpreted as the incidence matrix of a graph and governs the duality between the vertices, described in  $F_I$ , and its edges, described in  $F \wedge F$ , since  $t_{iu} = 1$  and  $t_{lu} = -1$  iff the edge  $u$  originates in  $i$  and attains  $l$  with  $i < l$ . besides,  $q_{i+} = \sum_{l \neq i} q_{il}$  is interpreted as the half outer degree of the vertice  $i$  in the graph  $G$  (cf. Berge (1973)), so that the  $N_Q$  metric on  $F$  weights more strongly the statistical units which are more often referred to in the design of relational constraints.

More generally, the constraints defined by one or several dissimilarity matrices are interpreted in the valued graphs setting, where the valuation is given by the constraining dissimilarities. The method may be extended as well as to some non linear constraints as thoses imposing an order on the projected coordinates of the statistical units on the axes, which revealed themselves a valuable add on to this methodology originally taylorred for applications in psychometry (see Noma and Johnson (1977), d'Aubigny (1987)).

The MDS approach to the problem is especially valuable when DCA is no more justified from a mathematical or from a statistical point of view. This is the case either when the response  $\delta$  is not an Euclidean dissimilarity function, or when it must be interpreted as the measurement of an Euclidean dissimilarity function contaminated by noise (measurement errors, sampling fluctuations, etc...). This is the case too, when our primary interest lies in the search for the best approximation, in the sense of a definite criterion, of the dissimilarity function by an Euclidean distance one.

Besides, the graph constraints are easily implemented in the MDS methodology, leading to a Discriminant MDS method. In the case of a metric interpretation of the dissimilarities, the treatment of the problem is straightforward with help of some elementary adaptations of the majorization algorithm of de Leeuw and Heiser (1980), aimed at the minimization of what should be called here a Local STRESS function, associated to the determination of a specific weight function  $Q_{I \times I}$ . When an ordinal assumption is made on the measurement scale of the response dissimilarity, the solution is more intricate ( see also Meulman 1986).

An alternative to this constrained MDS approach results from Section 4.1. above, by the direct application of a MDS algorithm to the triple  $(K, \delta_K, P_K)$ . But in that situation the embedding of the statistical units on the representation subspace obtained from the embedding of the centroids is not a simple problem, as indicated by Greenacre (1978). Since the MDS solution cannot be characterized in terms of projection of representative

points of the centroids on a subspace of  $E$ , the derivation of the coordinates of each statistical unit in the linear space spanned by the centroids is a non linear regression problem in its own.

Whatever these technical problems may be, the point of interest here is the preserved complementarity between the Principal Coordinates Analysis and the MDS approach for the analysis of a dissimilarity response variable in the setting of controlled experiments. The former method looks for optimal contrasts, while the latter yields a closer description of the topological properties of the response surface described by the observed dissimilarities.

Each of these approaches will gain in efficiency in the interpretative step from the use of a nonlinear Biplot methodology, recently developed by Gower and Harding (1988). This interpretative tool yields an extension of the original proposal of Gabriel (1971) in Linear Multivariate Analysis to the case of dissimilarity data. As one could expect, both methods are identical in the case of the application of Principal Coordinates Analysis to an Euclidean dissimilarity response.

Digby and Gower (1981) recommend to draw circles or of ellipses containing a fixed percentage of the illustrative points ( say 90% ) and centred in the corresponding centroids, in a way to compare with some care the dispersions of classes, in projection on a representation subspace. This recommendation mimics an usual practice in Linear Discriminant Analysis. It expresses once again the evidence that the choice of a dissimilarity response variable does introduce much less limitations to the potential exploratory data analysis tools available than is generally thought.

**Aknowledgments:** A first draft of this paper, has been presented in the XIV-th IBC conference, held in Namur, July 1988. We are thankfull to J.C. Gower for having indicated then to us his closely related work with P.G.N. Digby. His interest motivated some improvements and completions, especially for the sake of comparison of the two independent derivations of what is called Here the Discriminant Coordinates Analysis.

## REFERENCES

- Barret L., Stoklov M., Torch S., Stoebner P., and Saxod R. 1988. A Morphometric Evaluation of the Effects of Trichloroethylene and Dichloroacetylene on the Rat Mental Nerve. Preliminary Results. Submitted for Publication.
- Beals E.W. 1973. Ordination: Mathematical elegance and ecological naivete. *Journal of Ecology* **61**, 23-35.
- Berge C. 1973. *Graphes et Hypergraphes*. (2nd Ed.). Dunod, Paris.
- Caillez F. and J.-P. Pages 1976. *Introduction à l'Analyse des Données*. SMASH. 9 rue Duban, Paris.
- Bloxom B. 1978. Constrained multidimensional scaling in N spaces. *Psychometrika* **43**, 397-408.
- Cavalli-Sforza L.L. 1969. Human diversity. In *The Proceedings of the XII international congress of genetics, TOKYO Vol 3*, 405-416.
- Coxeter H. S. M. 1947. *Regular Polytopes*. Methuen & Co., LONDON.
- d'Aubigny G. 1984. Un Modèle Algébrique pour l'Analyse des Données de dissimilarité. *Proceedings of the Troisièmes Journées Franco-Soviétiques de Statistique et Analyse des Données*. Paris.
- d'Aubigny G. 1987. Un Modèle Algébrique pour l'Analyse des Données de dissimilarité sous Contraintes. *Proceedings of the Troisièmes Journées Franco-Soviétiques de Statistique et Analyse des Données*. Souzdal, URSS.
- d'Aubigny G. and C. d'Aubigny, 1988. The influence of the drug on the myelinated fibre size distribution: a generalization of the ANOVA to general diversity functions. *Technical Report*. IMSS, Grenoble. (Submitted for publication).
- de Leeuw J. 1977. Applications of Convex analysis to Multidimensional Scaling. In J.R. Barra et al. (Eds.). *Recent Developments in Statistics*. North Holland, Amsterdam.
- de Leeuw J. and W. Heiser 1980. Multidimensional Scaling with Restrictions on the Configuration. In P. R. Krishnaiah (Ed.). *Multivariate Analysis V*. North Holland, Amsterdam. 501-522.
- Dennis B., G.P. Patil, O. Rossi, S. Stehman and C. Taille 1979. A bibliography of literature on ecological diversity and related methodology. In *Ecological diversity in theory and practice Vol 1* CPH, 319-354.
- Digby P.G.N. and J.C. Gower 1981. Ordination Between- and Within-Groups Applied to Soil Classification. In D.F. Merriam (ed.) *Down to Earth Statistics: Solutions Looking for Geological Problems*. Syracuse University, Geological Contributions, No 8.

- Edgington E.S. 1980. *Randomization tests Statistics* : textbooks and monographs. D.B. Owen (Ed.). Marcel Dekker, Inc. New York and Basel.
- Fichet B., 1987. The role played by  $L_1$  in data analysis. In *Statistical Data Analysis based on the  $L_1$ -norm and related methods*. Y.DODGE (ed.). North-Holland. 185-193.
- Gabriel K.R. 1971. The Biplot Graphic Display of Matrices with Applications to Principal Component Analysis. *Biometrika* **58**, 453-467.
- Gower J. C. 1966. Some Distance Properties of Latent Root and Vector Methods used in Multivariate Analysis. *Biometrika* **53**, 325-388.
- Gower J. C. 1968. Adding a Point to Vector Diagrams in Multivariate Analysis. *Biometrika* **55**, 582-585.
- Gower J.C. 1984. Multivariate analysis : ordination, multidimensional scaling and allied topics. In E.H. Lloyd (ed.) *Handbook of applicable mathematics VI*, 727-781. Wiley.
- Gower J. C. 1986. Euclidean Distances Matrices. *Multidimensional Data Analysis*. J. de Leeuw, W. Heiser, J. Meulman (Eds.). DSWO Press, Leiden.
- Gower J. C. and S.A. HARDING 1988. Non Linear Biplots. *Biometrika* **75**, 3, 12 p. (to appear).
- Greenacre M.J. 1978. Some objective Methods of Graphical Display of a Data Matrix. Special Report, Univ. of South Africa, Pretoria.
- Hill M.O. and H.G. Gauch, 1980. Detrended correspondence analysis : an improved ordination technique. *Vegetatio* **42**, 47-58.
- Hubert L.J. 1978. Generalized proximity function comparisons. *British Journal of Mathematical and Statistical Psychology* **31**, 179-192.
- Hubert L.J. 1982. Analysis of variance procedures based on a proximity measure between subjects. *Psychological Bulletin* Vol **91** No 2, 424-430.
- Johnson D. E. and F.A. Graybill 1972. An analysis of a Two-Way Model with interaction and no replication. *Journal of American Statistical Association* **67**, 862-868.
- Kruskal J. B. 1964. Multidimensional Scaling by Optimizing Goodness-of-fit to a nonmetric Hypothesis. *Psychometrika* **29**, 1-28, 115-129.
- Kruskal J. B. and M. Wish, 1978. *Multidimensional Scaling*. Sage P. Beverly hills C.A.
- Lebart L. 1969. Analyse statistique de la contiguité. *Publications de l'ISUP* **18**, 81-112.
- Lebart I., A. Morineau and K. Warwick, 1984. *Multivariate Descriptive Statistical Analysis*. Wiley. New York.
- Le Calve G. 1987.  $L_1$  embeddings of a data structure (I,D). In *Statistical Data Analysis based on the  $L_1$ -norm and related methods..* Y.DODGE (ed.). North-Holland. 195-202.
- McCullagh P. and Nelder J.A., 1983. *Generalized Linear Models*: Chapman and Hall, London.
- Mahalanobis P.C., D.N. Majumdar and C.R. Rao, 1949. Anthropometric survey of the united provinces 1941 : a statistical study. *Sankhya* **9**, 90-324.

- Mailles J.P. 1979. Analyse Factorielle des Tableaux de Dissimilarités. Thèse, Univ. Pierre et Marie Curie, Paris.
- Meulman J. 1986. *A Distance Approach to Nonlinear Multivariate Analysis*. DSWO Press, Leiden.
- Noma E. and Johnson J. 1977. Confirmatory Multidimensional Scaling. Unpublished manuscript. The University of Michigan.
- Orloci L. 1979. Non linear data structures and their description . In *Multivariate methods in ecological work* 191-202. L. Orloci, C.R. RAO, W.M. STITELER eds. International cooperative Publishing House, Fairland, Maryland.
- Patil G.P. and C. Taille, 1979. An overview of diversity. In *Ecological diversity in theory and practice Vol 1* CPH, 3-28.
- Pielou E.C. 1975. *Ecological Diversity*. Wiley, New York.
- Rao C.R., 1948. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society B* 10, 159-193.
- Rao C.R., 1982. Diversity and dissimilarity coefficients : a unified approach. *Theoretical Population Biology Vol 21* Nr 1, 24-43.
- Rao C.R. 1984. Convexity properties of entropy functions and analysis of diversity. In *Inequalities in Statistics and Probability*. IMS Lecture. Monograph Series Vol 5, 68-77.
- Schoenberg I. 1935. Remarks to Maurice Frechet's article "Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicables sur l'espace de Hilbert." *Annals of Mathematics (2)* 36, 724-732.
- SCHONEMANN P. H. 1972. An Algebraic Solution for a Class of Subjective Metrics Models. *Psychometrika* 37, 441-451.
- Shepard R. N., 1962. The Analysis of proximities: Multidimensional Scaling with an unknown distance function. *Psychometrika* 27, 125-140, 219-246.
- Sokal R.R. and P.H.A. Sneath, 1963. *Principles of Numerical Taxonomy*. Freeman, San Francisco.
- ter Braak J.F. 1984. Correspondence analysis of incidence and abundance data : properties in terms of a unimodal response model. *Technical Report A84ST11*. IWIS-TNO, Wageningen Section, Wageningen, the Netherland.
- Takeuchi K., Yanai H., Mukherjee B.N., 1982. *The Foundations of Multivariate Analysis*. Wiley, New York.
- Torgerson W.S. 1958. *Theory and Methods of Scaling*. Wiley, New York.
- Wilkinson C. 1970. Adding a Point to a Principal Coordinates Analysis. *Syst. Zool.* 19, 258-263.
- Whittaker R.H. 1967. Gradient analysis of vegetation. *Biological Review* 42, 207-264
- Young G. and A.S. Householder, 1938. Discussion of a set of points in term of their mutual distances. *Psychometrika* 3, 19-22.



AUTORISATION DE SOUTENANCE

DOCTORAT D'ETAT

Vu les dispositions de l'Article 5 de l'Arrêté du 16 avril 1974,

Vu les rapports de M...J...de LEEUW.....

M...G...LECALVÉ.....

M...G...ROMIER.....

M...Gerald...DROUET...d'AVIGNY..... est autorisé à  
présenter une thèse en vue de l'obtention du grade de DOCTEUR D'ETAT  
ES SCIENCES.

Fait à Grenoble, le 17 JAN. 1989

Le Président de l'U.J.F.



  
J.J. PAYAN



## **Résumé**

Notre étude des méthodes d'Analyse des Tableaux de dissimilarité poursuit quatre buts principaux.

Tout d'abord, un cadre algébrique unificateur est proposé pour décrire les deux classes de méthodes existantes, dites de réduction de dimension et de Codage Multidimensionnel. Dans le premier cas, l'Analyse en Coordonnées Principales est complétée par la proposition de méthodes nouvelles d'Analyse en Coordonnées d'Image d'une part et d'Analyse factorielle du triple d'autre part. Le statut méthodologique de ces approches est établi et les problèmes de constante additive sont réinterprétés en termes de Regression Ridge. L'étude des relations algébriques entre ces méthodes et celle de Codage Multidimensionnel permet de dégager le statut statistique des solutions de l'Analyse en Coordonnées Principales.

En second lieu, nous proposons un ensemble d'outils de diagnostic et d'aide à l'interprétation des résultats d'analyse.

L'étude des méthodes d'Analyse sous Contraintes est le troisième objectif. Nous montrons à l'aide d'une reformulation tensorielle du problème les relations liant ces méthodes à l'Analyse Multivariée des courbes de croissance, nous dégageons le rôle des métriques dans l'élaboration et les propriétés de la solution et nous proposons des algorithmes résolutifs pour une classe très générale de contraintes, non nécessairement linéaires et incluant le problème des rotations procrustes.

La nouvelle interprétation des méthodes par analogie à la physique permet en dernier lieu de proposer une méthodologie d'analyse de données de dissimilarité structurées analogue à l'ANOVA et complétée, du point de vue exploratoire par une méthode nouvelle dite d'Analyse en Coordonnées Discriminantes.

## **Mots clés**

Coordonnées Principales - Constante Additive - Codage Multidimensionnel - Dissimilarité et Quantité d'Information - Contraintes - Contrastes 2x2 - Analyse de la Diversité - Coordonnées Discriminantes.