

Analyse des Proximités et programmes de Codage Multidimensionnel

Gérard d'Aubigny

LABSAD,
Université Pierre Mendès-France,
Grenoble, France

1 Introduction

L'objectif poursuivi par ce texte est de fournir des points de repère méthodologiques et logiciels aux utilisateurs désireux d'analyser des tableaux de données de proximités. On dispose alors d'un ou de plusieurs tableaux de données brutes (\mathbf{R}^i , $i = 1, 2, \dots, n$) dont le terme général r_{jk}^i (r comme raw data) fournit une mesure de proximité entre les objets j et k ($j, k = 1, 2, \dots, p$), relevée dans la situation i . Les objets sont des stimuli en psychologie, des biens en économie, des espèces ou des repères en biologie, des atomes en chimie, des sites, des zones, des lignes en géographie, des sujets en sociologie, etc. Les situations correspondent aux unités statistiques d'un protocole expérimental ou d'une enquête, à des temps ou des lieux d'observation, etc. Nous parlons dans la suite de sujets pour les éléments i et d'objets pour les éléments j .

Parmi les méthodes d'analyse des structures de proximité, on s'intéresse à la classe de modèles qui se proposent de représenter les coefficients de proximité entre objets sous la forme de distances entre points d'un espace affine. On obtient ainsi un graphique dont la consultation, plus attrayante que la lecture d'un tableau de chiffres, permet souvent de dégager des constatations qui seraient passées inaperçues sans représentation, des modes de structuration de l'information et donc des stratégies de lecture du (ou des) tableau(x) de données étudié(s).

Historiquement, les concepts d'axe et de dimension ont joué un grand rôle dans la recherche de structures. Ils apparaissent même à beaucoup d'utilisateurs néophytes comme l'unique finalité de l'emploi de méthodes d'analyse multivariée, l'unique type de régularité à rechercher dans la lecture des graphiques. Ceci explique l'adoption du terme de *Codage Multidimensionnel* pour désigner la classe de méthodes dont l'ambition est de construire une représentation des données dans un cadre géométrique caractérisé par la construction de nuages de points dans des espaces affines de dimension aussi faible que possible et dépendant des données, afin de visualiser les positions relatives d'un ensemble fini d'objets. On parle aussi parfois de *Positionnement Multidimensionnel*, cf. Escoufier (1980), ou encore d'*Echelonnement Multidimensionnel*, cf. Tournois et Dickes (1992)), comme traduction des termes anglais de *Multidimensional Scaling* utilisé en psychométrie et d'*Ordination Method* préféré en biométrie.

Le recours à la perception visuelle suppose de plus un modèle de représentation qui fixe en particulier le choix d'une géométrie métrique, c'est à dire un mode de calcul des distances interpoints sensées représenter au mieux (au sens de critères discutés ci-dessous) les relations observées de proximité entre objets.

Les méthodes d'analyse envisagées dans ce texte, concernent uniquement la recherche d'une représentation affine. Elles diffèrent des approches combinatoires étudiées par exemple par Chandon et Pinson (1981) et Barthélemy et Guénoche (1988) et fondées sur l'emploi de représentations arborées. Le choix de l'un de ces deux types de modèles de représentation dépend en général de la nature des phénomènes étudiés, des objectifs poursuivis, et parfois même des a priori culturels ou esthétiques de l'utilisateur. Il convient donc d'insister sur la complémentarité de ces deux approches parfois perçues à tort comme antinomiques.

De plus, l'adoption d'une géométrie pose un problème pratique d'apprentissage des règles de lecture des graphiques obtenus. Par exemple, lorsqu'on munit \mathbb{R}^2 d'une métrique de Minkowski distincte de la métrique Euclidienne, on produit une géométrie dans laquelle la boule unité n'a plus la forme d'un disque conforme à la perception visuelle usuelle. La lecture des graphiques perd alors son support intuitif. Afin d'éviter une discussion de ces problèmes, la présentation adoptée restreint son ambition au cadre Euclidien.

Enfin, nous n'aborderons pas les méthodes de collecte de l'information conduisant à la construction des tableaux de proximités, par souci d'économie de place. Nous renvoyons le lecteur intéressé aux ouvrages classiques tels que Chandon et Pinson (1981) ou à Tournois et Dickes ((1992) chap. 2) pour une discussion récente des problèmes posés par le passage des observations aux données. La difficulté du problème naît de la nécessité de prendre en compte les fluctuations d'échantillonnage (dans le cas $n > 1$) et les erreurs de mesure inhérentes à la collecte de l'information sous la forme de comparaisons par paires (d'objets), par triples, etc, lors de la construction de mesures de proximité. Cette phase délicate et très importante pour la pertinence de l'étude, a de nombreuses conséquences sur les propriétés des tableaux de dissimilarités obtenus et sur le choix de stratégies d'analyse. Elle est cependant exclue de cet exposé, dont la portée est limitée à une présentation technique des méthodes.

Donnons tout d'abord trois exemples classiques d'analyse de tableaux de proximités:

Exemple 1: Proximités géographiques.

Mardia et al. (1979) fournissent un tableau récapitulatif des distances routières séparant 12 villes britanniques et se proposent de reconstruire de façon approchée une carte géographique de ces villes. Le propos des auteurs est moins élémentaire qu'il n'y paraît, dans la mesure où le modèle de représentation proposé reconstruit des distances Euclidiennes (ou à vol d'oiseau) dans un espace à deux dimensions, à partir de la géométrie définie par des géodésiques complexes puisque définies par un réseau de routes serpentant par monts et par vaux dans la campagne britannique, c'est à dire selon une géométrie à priori tri-dimensionnelle dont le support (la surface de la terre) peut être assimilée à une variété Riemannienne particulière, la sphère.

Towns	1	2	3	4	5	6	7	8	9	10	11	12	v propres
Aberystw	116 16	-4 42	-5 59	-14 92	38 74	-21 05	-17 63	-40 59	3 08	-5 26	-25 42	-23 09	394473
Brighton	244	118 99	-70 17	89 99	55 16	-85 63	-19 22	-96 17	-33 51	75 66	-60 01	29 33	63634 4
Carlisle	218	350	90 66	-70 17	-49 01	75 20	2 12	55 66	15 33	-61 01	59 83	-42 84	13544 1
Dover	284	77	369	137 99	27 16	-85 63	-17 22	-94 17	-31 51	75 66	-58 01	40 82	10245 3
Exeter	197	167	347	242	158 33	-64 47	-28 55	-84 01	-27 34	36 83	-57 34	-5 51	2464 68
Glasgow	312	444	94	463	441	153 74	-13 84	128 20	-0 13	-76 47	48 37	-58 30	1450 27
Hull	215	221	150	236	279	245	63 58	-22 38	32 28	-9 55	16 28	14 12	500 895
Invernes	469	583	251	598	598	169	380	271 66	-10 67	-87 01	46 83	-67 34	-4 65e-10
Leeds	166	242	116	257	269	210	55	349	55 99	-24 34	25 49	-4 67	-16 9392
London	212	53	298	72	170	392	168	531	190	85 33	-50 84	40 99	-214 076
Newcastl	253	325	57	340	359	143	117	264	91	273	85 99	-31 17	-1141 22
Norwich	270	168	284	164	277	378	143	514	173	111	256	107 66	-7063 42

Table 1: Distances routières entre 12 villes britanniques. (Source: Mardia et al. 1979).

La partie triangulaire inférieure (resp. supérieure) fournit les distances routières (resp. les pseudo produits scalaires)

La dernière colonne produit les valeurs propres de la matrice soumise à Codage en Composantes Principales.

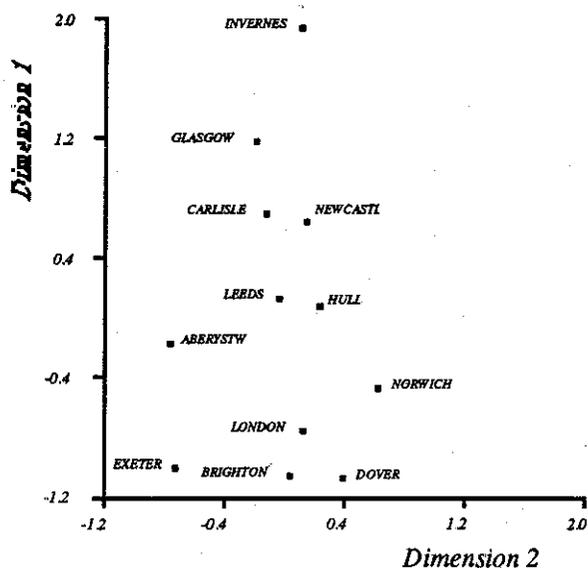


Figure 1 Codage en Composantes Principales de 12 villes Britanniques.

Exemple 2: Proximités subjectives.

L'un des jeux de données les plus célèbres de la littérature consacrée au Codage Multidimensionnel est fourni par les données d'Ekman (1954). Celui-ci mesura le degré perçu de similarité (ou de ressemblance) entre 14 couleurs variant en teinte (de 434 nm à 674 nm en longueur d'ondes). Les sujets exprimaient leurs jugements de similarité à l'aide d'une note fournie pour chaque combinaison de deux couleurs. Ekman construisit ensuite une note synthétique associée à chaque paire et obtenue

par moyennage puis normalisation afin de varier de 0 (totalement dissemblable) à 1 (identique).

434	445	465	472	490	504	537	555	584	600	610	628	651	674
	0.86	0.42	0.42	0.18	0.06	0.07	0.04	0.02	0.07	0.09	0.12	0.13	0.16
0.86		0.50	0.44	0.22	0.09	0.07	0.07	0.02	0.04	0.07	0.11	0.13	0.14
0.42	0.50		0.81	0.47	0.17	0.10	0.08	0.02	0.01	0.02	0.01	0.05	0.03
0.42	0.44	0.81		0.54	0.25	0.10	0.09	0.02	0.01	0.00	0.01	0.02	0.00
0.18	0.22	0.47	0.54		0.61	0.31	0.26	0.07	0.02	0.02	0.01	0.02	0.00
0.06	0.09	0.17	0.25	0.61		0.62	0.45	0.14	0.08	0.02	0.02	0.02	0.01
0.07	0.07	0.10	0.10	0.31	0.62		0.73	0.22	0.14	0.05	0.02	0.02	0.00
0.04	0.07	0.08	0.09	0.26	0.45	0.73		0.33	0.19	0.04	0.03	0.02	0.02
0.02	0.02	0.02	0.02	0.07	0.14	0.22	0.33		0.58	0.37	0.27	0.20	0.23
0.07	0.04	0.01	0.01	0.02	0.08	0.14	0.19	0.58		0.74	0.50	0.41	0.28
0.09	0.07	0.02	0.00	0.02	0.02	0.05	0.04	0.37	0.74		0.76	0.62	0.55
0.12	0.11	0.01	0.01	0.01	0.02	0.02	0.03	0.27	0.50	0.76		0.85	0.68
0.13	0.13	0.05	0.02	0.02	0.02	0.02	0.02	0.20	0.41	0.62	0.85		0.76
0.16	0.14	0.03	0.00	0.00	0.01	0.00	0.02	0.23	0.28	0.55	0.68	0.76	

Table 2: Similarité perçue des couleurs correspondant à des longueurs d'ondes variant entre 434 nm et 674 nm

Données de Ekman (1954).

Ekman soumit ces données à une Analyse Factorielle qui conduisit à retenir 5 facteurs, identifiés comme les facteurs violet, bleu, vert, jaune et rouge. L'article fondateur de Shepard (1962) reprit les données d'Ekman pour étalonner une proposition d'algorithme. Cette heuristique conduisit à une représentation jugée par les spécialistes du domaine plus économe en dimensions (deux seulement), plus pertinente et plus aisément interprétable.

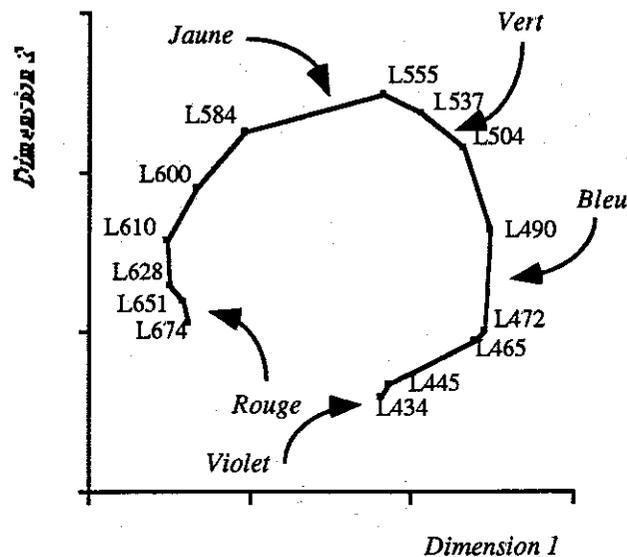


Figure 2 Données de Ekman, représentation en dimension 2, Méthode MINRES.

Exemple 3: Proximités déduites d'un protocole expérimental.

Tout aussi célèbres que les données d'Ekman parce qu'elles aussi traitées par Shepard (1962), les données de ROTHKOPF (1957) concernent une expérience de reconnaissance de caractères du code MORSE: les sujets non formés au code MORSE écoutaient une suite de points et de traits, tels que - - (la lettre K) ou . . . - (le chiffre 4), qui leur étaient présentés par paires, à un rythme rapide géré par une machine et séparés par un intervalle de 1,4 secondes. Ils devaient dire si les deux éléments de la paire étaient identiques ou différents. Lorsque les réponses sont cumulées sur l'ensemble des sujets étudiés, on dispose d'une matrice de données de confusion, dont la case (j, k) fournit le nombre de sujets qui ont confondu les signaux j et k . Cette démarche interprète le nombre de confusions r_{jk} comme un indice de ressemblance entre les codes j et k .

La structure de la matrice de ROTHKOPF originelle (cf. Kruskal et Wish (1978) page 11) n'est qu'approximativement symétrique. Sa diagonale est forte, mais non constante. Ces caractéristiques montrent une limitation additionnelle de notre présentation pour l'essentiel limitée au cadre fixé par la définition suivante:

Définition 1 On appelle matrice de dissimilarités propre d'ordre p toute matrice $p \times p$ symétrique \mathbf{R} de terme général r_{jk} telle que pour tous j et k variant entre 1 et p :

$$r_{jj} = 0, \quad r_{jk} \geq 0 \quad j \neq k \quad (1)$$

Cette définition repose sur une interprétation de r_{jk} comme une mesure de dissemblance et non de ressemblance. Le passage de l'un à l'autre concept n'est pas toujours simple. Nous ne le traitons cependant pas ici. De même renvoyons nous, pour des impératifs de gain de place, à Fichet (1988) ou d'Aubigny (1989) pour une étude des problèmes mathématiques liés à l'obtention d'une dissimilarité propre à partir d'une dissimilarité observée.

Dans le cas des données de Rothkopf, citons à titre d'exemple deux possibilités offertes pour prendre en compte l'absence de symétrie. La solution la plus couramment adoptée consiste à symétriser le tableau de données :

$$r_{jk} \longrightarrow s_{jk} = \frac{1}{2}(r_{jk} + r_{kj})$$

Une solution alternative consiste à ajuster un modèle qui intègre un terme d'antisymétrie, tel que par exemple :

$$r_{jk} = (\alpha_j - \alpha_k) + \mu_{jk} + \epsilon_{kj} \text{ avec } \mu_{jk} = \mu_{kj}$$

2 Analyse algébrique du triple $(J, R_{J \times J}, P_J)$

2.1 Introduction

La construction d'une représentation Euclidienne d'un ensemble fini J muni de masses p_j et de distances (cf. Benzécri et al. 1973) est la méthode de codage multidimensionnel la plus étudiée en France, où son succès tient sans doute au surplus de faveur généralement accordé aux méthodes algébriques. Par suite, on en trouve maintes présentations sous les dénominations diverses (1) d'Analyse du Triple (Benzécri 1973),

(2) d'Analyse Factorielle sur Tableaux de Distance (cf. Cailliez et Pages (1975) par exemple), (3) d'Analyse des Proximités (cf. Benzécri (1965a, b), Davy de Virville (1972)), (4) de Positionnement Multidimensionnel (Escoufier 1975), etc...

Nous utilisons ici le vocable de *Codage en Composantes Principales* (ou CCP) qui marque de façon claire la filiation de cette approche à l'Analyse en Composantes Principales (ou ACP), en tant que problème inverse. La donnée du problème d'ACP est la matrice \mathbf{X} à partir de laquelle on construit un résumé $\widehat{\mathbf{X}}$ par projection, dont on déduit une matrice de distances Euclidiennes approchées $\mathbf{D}(\widehat{\mathbf{X}})$ par calcul algébrique. En CCP la donnée est un tableau de pseudo-distances \mathbf{R} à partir duquel on construit une matrice $\mathbf{X} = \mathbf{X}(\mathbf{R}) = \widehat{\mathbf{A}}$ appelée *configuration* dans la suite. Il s'agit d'une approximation non unique d'une configuration supposée \mathbf{A} , qui est obtenue par calcul algébrique puis par projection.

Cette analyse algébrique du triple $(J, R_{J \times J}, P_J)$ appartient donc à la famille des méthodes dont l'objectif est de construire des composantes, c'est à dire des combinaisons linéaires *exactes* de variables initiales observées (en ACP) ou latentes (en CCP). Elle doit donc être distinguée des méthodes factorielles qui construisent des combinaisons linéaires *approximations* des variables initiales observées ou latentes.

Le théorème suivant montre que dans les deux cas, traiter le problème inverse conduit à une solution non unique. Notons E et G deux espaces vectoriels de dimensions respectives p et n , E^* l'espace vectoriel dual algébrique de E et $\mathbf{1}$ le vecteur constant de E de composantes toutes égales à l'unité. Notons de plus \mathbf{X} et \mathbf{Y} deux matrices réelles d'ordre $p \times n$ associées, pour un choix de base donné, aux applications linéaires x et y définies sur E^* et à valeurs dans G . Soit enfin $\boldsymbol{\mu}$ un vecteur de G . Alors:

Théorème 1 Soient \mathbf{O} une matrice orthogonale d'ordre n , $\boldsymbol{\mu}$ un vecteur de dimension n et α un scalaire positif décrivant respectivement une rotation ou une symétrie, une translation et une homothétie dans G . Soient de plus \mathbf{X} et \mathbf{Y} deux matrices réelles d'ordre $p \times n$ telles que

$$\mathbf{Y} = \alpha \mathbf{X} \mathbf{O} + \mathbf{1}^t \boldsymbol{\mu}$$

Alors, les matrices de distance Euclidienne associées $\mathbf{D}(\mathbf{X})$ et $\mathbf{D}(\mathbf{Y})$ vérifient pour tout (j, k) dans $J \times J$:

$$d_{jk}(\mathbf{X}) = \alpha^2 d_{jk}(\mathbf{Y}).$$

La configuration de points obtenue par les méthodes d'analyse Euclidienne des données de proximité et donc en particulier par Codage en Composantes Principales n'est pas identifiée. La solution est invariante par toute transformation appartenant au groupe des similitudes de G .

2.2 La solution classique

L'analyse algébrique du triple $(J, R_{J \times J}, P_J)$ est justifiée par l'affirmation du caractère Euclidien des dissimilarités observées. La relation entre produits scalaires et distances est alors fournie par le théorème de Pythagore appliqué au triangle $\triangle(O, S_j, S_k)$, avec les notations portées sur la figure ci-dessous:

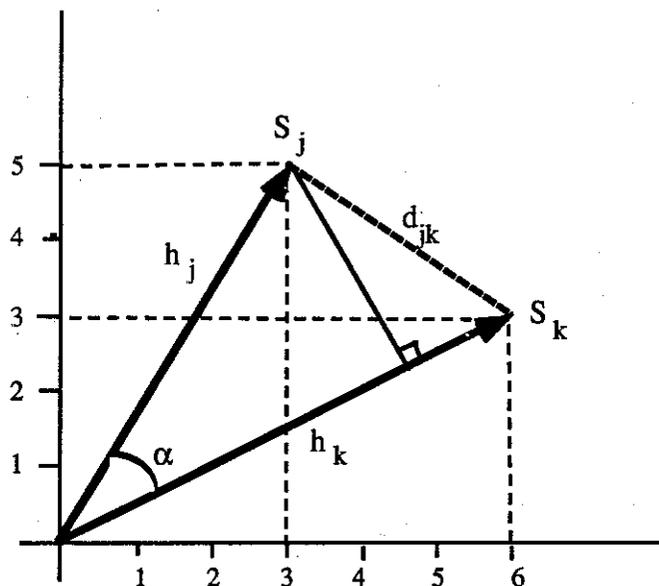


Figure 3 Illustration de la relation entre produit scalaire et distance Euclidienne

$$d_{jk}^2 = h_j^2 + h_k^2 + 2h_j h_k \cos(\alpha) \quad (2)$$

où de façon classique, le choix du barycentre \bar{x} comme origine conduit à écrire:

$$h_j^2 = b_{jj} = d_j^2 - \frac{1}{2}d^2 \quad (3)$$

$$= \langle \mathbf{x}_j - \bar{\mathbf{x}}, \mathbf{x}_j - \bar{\mathbf{x}} \rangle = \|\mathbf{x}_j - \bar{\mathbf{x}}\|^2 \quad (4)$$

$$h_j h_k \cos(\alpha) = b_{jk} = \langle \mathbf{x}_j - \bar{\mathbf{x}}, \mathbf{x}_k - \bar{\mathbf{x}} \rangle$$

$$d_j^2 = \frac{1}{p} \sum_{k=1}^p d_{jk}^2$$

$$d^2 = \frac{1}{p} \sum_{j=1}^p d_j^2$$

$$\bar{\mathbf{x}} = \frac{1}{p} \sum_{j=1}^p \mathbf{x}_j$$

lorsque p_j est égal à $\frac{1}{p}$ pour tout j . De plus, le cas d'une pondération quelconque p_j ; $j = 1, 2, \dots, p$ s'écrit sous forme matricielle:

$$B = -\frac{1}{2} Q_1 H Q_1 \quad (5)$$

où $Q_1 = I_p - \mathbf{1}^t \mathbf{1} D_p$ est le projecteur D_p -orthogonal sur le sous espace de G , D_p -orthogonal au vecteur constant de G . En pratique, les difficultés proviennent du caractère non nécessairement Euclidien de la matrice de dissimilarités observée R . Le résultat suivant, démontré initialement par Schoenberg (1937) pour une pondération uniforme, permet d'évaluer la difficulté du problème rencontré :

Théorème 2 Soit \mathbf{R} une matrice de dissimilarités propre d'ordre p et \mathbf{B} la matrice déduite de \mathbf{R} par la relation (5) où \mathbf{H} est la matrice de terme général r_{jk}^2 . Alors \mathbf{R} est Euclidienne si et seulement si \mathbf{B} est positive semi-définie.

En particulier, la méthode classique s'appuie sur les résultats corollaires suivants:

(1) Si \mathbf{R} est une matrice de distances Euclidiennes entre points de la configuration \mathbf{A} , id est ${}^t\mathbf{A} = (\lambda_1, \lambda_2, \dots, \lambda_p)$, alors

$$b_{jk} = {}^t(\lambda_j - \bar{\lambda})(\lambda_k - \bar{\lambda}) \quad (6)$$

(2) Inversement, si \mathbf{B} est semi-définie positive et de rang n , alors une configuration peut être construite par le procédé suivant :

Soient $l_1 > l_2 > \dots > l_n$ les valeurs propres positives de \mathbf{B} associées aux vecteurs propres $\mathbf{A} = (\lambda^1, \dots, \lambda^n)$ pour $\|\lambda^s\|^2 = l_s$, $s = 1, 2, \dots, n$. Alors les points P_j de \mathbb{R}^n de coordonnées $\lambda_j = {}^t(\lambda_j^1, \dots, \lambda_j^n)$ ont pour matrice de distances inter-points \mathbf{R} , pour centre de gravité $\bar{\lambda} = \mathbf{0}$ et pour matrice de produits scalaires \mathbf{B} .

Remarques: (1) Sous forme matricielle (6) devient

$$\mathbf{B} = (\mathbf{Q}_1\mathbf{A}) {}^t(\mathbf{Q}_1\mathbf{A})$$

et donc \mathbf{B} est semi-définie positive. \mathbf{B} s'interprète alors comme la matrice des produits scalaires associés à la matrice centrée \mathbf{A} : $\mathbf{B} = \mathbf{B}(\mathbf{A})$.

(2) La j -ème ligne de \mathbf{A} contient les coordonnées du j -ième point alors que la s -ième colonne de \mathbf{A} contient le s -ème vecteur propre correspondant à la valeur propre l_s .

(3) Le théorème ne s'applique pas au cas où \mathbf{B} admet des valeurs propres négatives puisqu'il serait alors possible de contruire un vecteur dont le carré de la norme Euclidienne serait négatif.

2.3 Un algorithme pratique

2.3.1 Présentation générale

Supposons que la matrice \mathbf{R} représente approximativement des distances interpoints dans un espace Euclidien de dimension n faible (usuellement 1, 2 ou 3). Si $\text{rang}(\mathbf{B}) \geq n$, on choisit pour configuration dans \mathbb{R}^n la configuration $\mathbf{X} \equiv \hat{\mathbf{A}}$ de points dont les coordonnées sont fournies par les n premiers vecteurs propres de \mathbf{B} , si les valeurs propres correspondantes sont positives. Si les valeurs propres d'ordre supérieur à n sont proches de zéro, on peut alors espérer que les distances reconstruites $\mathbf{D}(\mathbf{X})$ fournissent une bonne approximation de \mathbf{R} . Cette solution, appelée la solution classique de Codage en Coordonnées Principales depuis les travaux de Gower (1966) procède en quatre étapes:

- (1) déduire de \mathbf{R} la matrice $\mathbf{H} = (-1/2r_{jk}^2)$
- (2) déduire de \mathbf{H} la matrice $\mathbf{B} = (h_{jk} - h_j - h_k + h)$
- (3) Calculer les n plus grandes valeurs propres positives de \mathbf{B}

$$l_1(\mathbf{B}) > l_2(\mathbf{B}) > \dots > l_n(\mathbf{B}) \quad (7)$$

et les vecteurs propres correspondants \mathbf{x}^s normés à la racine carrée de la valeur propre associée. Ils forment les colonnes de la matrice \mathbf{X}

(4) Les lignes de \mathbf{X} donnent les coordonnées cherchées des points représentatifs des objets. De plus, $\mathbf{X}^t \mathbf{X}$ est une approximation de \mathbf{B} au sens des moindres carrés.

Le tableau 1 a été obtenu par application de ce procédé aux données de Mardia et al. (1979). La partie triangulaire supérieure du tableau fournit alors les valeurs (arrondies à la 1-ère décimale) des distances Euclidiennes reconstruites en $n = 2$ dimensions. Celles-ci illustrent une propriété essentielle de la méthode: à l'instar de l'ACP, celle-ci reconstruit un sous-espace de G de dimension q par approximation interne de l'espace support du nuage de points étudié. Par suite, si $d_{jk}(\hat{\Lambda}^{(m)})$ désigne la distance entre j et k reconstruite dans un sous-espace affine de dimension m , on la chaîne d'inégalités

$$0 \leq d_{jk}^2(\hat{\Lambda}^{(1)}) \leq d_{jk}^2(\hat{\Lambda}^{(2)}) \leq \dots \leq d_{jk}^2(\hat{\Lambda}^{(m)}) \leq \dots \leq d_{jk}^2(\Lambda) = r_{jk}^2$$

où l'égalité $\mathbf{R} = \mathbf{D}(\Lambda)$ est vérifiée si et seulement si \mathbf{R} est une matrice de distance Euclidienne.

Définition 2 On appelle *diagramme de Shepard*, le graphique bivarié croisant dissimilarités observées r_{jk} et distances Euclidiennes reconstruites $d_{jk}(\mathbf{X})$.

Cette propriété d'approximation interne se traduit sur le diagramme de Shepard par une structure particulière: toutes les paires représentées sont situées au-dessous de la première bissectrice des axes. Ceci n'est pas vrai pour la solution fournie par les méthodes de codage multidimensionnel qui sont étudiées dans la section suivante, comme le montre la consultation de la table 4 fournie ci-dessous.

0 000	118 087	126 850	168 568	29 847	186 342	129 421	276 470	92 318	121 021	153 796	187 164
118 087	0 000	264 214	7 534	83 473	379 858	100 795	496 957	118 472	3 116	234 684	43 865
126 850	264 214	0 000	275 417	264 986	10 573	62 619	36 888	32 947	227 492	7 865	188 921
168 568	7 534	275 417	0 000	139 714	393 749	89 705	505 298	119 363	3 958	231 598	21 969
29 847	83 473	264 986	139 714	0 000	357 921	203 249	480 586	176 235	102 655	286 197	205 502
186 342	379 858	10 573	393 749	357 921	0 000	120 834	9 068	80 232	336 122	27 700	284 811
129 421	100 795	62 619	89 705	203 249	120 834	0 000	177 469	6 634	71 512	35 161	34 789
276 470	496 957	36 888	505 298	480 586	9 068	177 469	0 000	133 519	443 513	54 676	369 008
92 318	118 472	32 947	119 363	176 235	80 232	6 634	133 519	0 000	91 018	19 707	65 226
121 021	3 116	227 492	3 958	102 655	336 122	71 512	443 513	91 018	0 000	194 281	23 979
153 796	234 684	7 865	231 598	286 197	27 700	35 161	54 676	19 707	194 281	0 000	139 605
187.164	43.865	188.921	21.969	205.502	284.811	34.789	369.008	65.226	23.979	139.605	0.000

Table 4 Distances routières entre 12 villes Britanniques reconstituées
par Codage Multidimensionnel Ordinal en dimension 2.

Remarque: Lorsque l'on construit une représentation de \mathbf{R} dans le champ complexe $G = \mathbb{C}^n$ et non plus \mathbb{R}^n . Les valeurs propres négatives de \mathbf{B} correspondent alors à des composantes principales à coordonnées complexes. Par suite si $n = r + i$, où

r (resp. i) est le nombre de composantes réelles (resp. imaginaires pures), on peut décomposer les carrés des distances reconstruites en:

$$\begin{aligned} d_{jk}^2(\mathbf{A}) &= \sum_{s=1}^r (\lambda_j^s - \lambda_k^s)^2 + \sum_{s=r+1}^{r+i} (i\lambda_j^s - i\lambda_k^s)^2 \\ &= \sum_{s=1}^r (\lambda_j^s - \lambda_k^s)^2 - \sum_{s=r+1}^{r+i} (i\lambda_j^s - i\lambda_k^s)^2 \\ &= d_{jk,R}^2(\mathbf{A}) - d_{jk,I}^2(\mathbf{A}) \end{aligned} \quad (8)$$

Ce carré s'écrit donc sous forme d'une différence de deux termes positifs, correspondant respectivement aux coordonnées réelles et imaginaires pures de la configuration observée, de sorte qu'il est possible d'écrire pour $q \leq r$:

$$d_{jk}^2(\mathbf{A}) = [d_{jk}^2(\mathbf{X}^{(q)}\mathbf{O}) + d_{jk}^2(\mathbf{X}^{(r-q)}\mathbf{O})] - d_{jk,I}^2(\mathbf{A}) \quad (9)$$

où \mathbf{X} est la solution obtenue en analysant la partie réelle $\mathbf{D}_R(\mathbf{A})$ et \mathbf{O} est la rotation qui permet de faire coïncider \mathbf{X} et la partie réelle de \mathbf{A} . Retenir un résumé de dimension q ($q \leq r$), revient à écrire la décomposition suivante des carrés des distances

$$d_{jk}^2(\mathbf{A}) = d_{jk}^2(\mathbf{X}^{(q)}\mathbf{O}) + [d_{jk}^2(\mathbf{X}^{(r-q)}\mathbf{O}) - d_{jk}^2(\mathbf{A})] \quad (10)$$

Le terme entre crochets peut être positif, nul ou négatif. $\mathbf{X} = \mathbf{X}^{(q)}$ fournit donc une approximation bilatérale $\mathbf{D}_R(\mathbf{X})$ de $\mathbf{D}(\mathbf{A})$.

Maintenant, la distance entre j et k reconstruite dans un sous-espace affine réel de dimension q , notée $\hat{d}_{jk}^2 = d_{jk}(\hat{\Lambda}^{(q)})$, obtenue par extraction des éléments propres de la matrice \mathbf{B} , donc par approximation interne des dissimilarités observées, vérifie la propriété d'optimalité suivante:

Théorème 3 Soient $\mathbf{R} \equiv \mathbf{D}(\mathbf{A})$ une matrice de distances Euclidienne correspondant à une configuration \mathbf{A} dans $G \sim \mathbb{R}^n$ et q un entier donné compris entre 1 et n . Alors parmi toutes les projections $\mathbf{A}\mathbf{P}_q$ de \mathbf{A} sur un sous-espace affine de G , la quantité

$$\Phi = \sum_{j,k=1}^p (d_{jk}^2(\mathbf{A}) - \hat{d}_{jk}^2) = \sum_{j,k=1}^p d_{jk}^2(\mathbf{A}) \left[1 - \frac{\hat{d}_{jk}^2}{d_{jk}^2(\mathbf{A})} \right]$$

est minimisée quand \mathbf{A} est projetée sur le sous-espace de dimension q engendré par ses coordonnées principales.

La figure 4 ci-dessous illustre sur le tableau des distances entre villes britanniques, la différence de comportement entre les méthodes d'approximation interne (ici la solution pour $q = 2$ de la méthode CCP) et les méthodes d'approximation bilatérale (ici la solution de la méthode MINRES (cf. infra) de Codage Multidimensionnel Ordinal pour $q = 2$). La position du nuage de points de coordonnées $(d_{jk}(X^{(MINRES)}), d_{jk}(X^{(CCP)}))$ montre un biais systématique par défaut de la méthode CCP. Le théorème (3) fournit l'une des raisons de ce biais. La méthode CCP n'est pas associée à un critère naturel d'approximation de \mathbf{R} par une matrice de distances Euclidiennes.

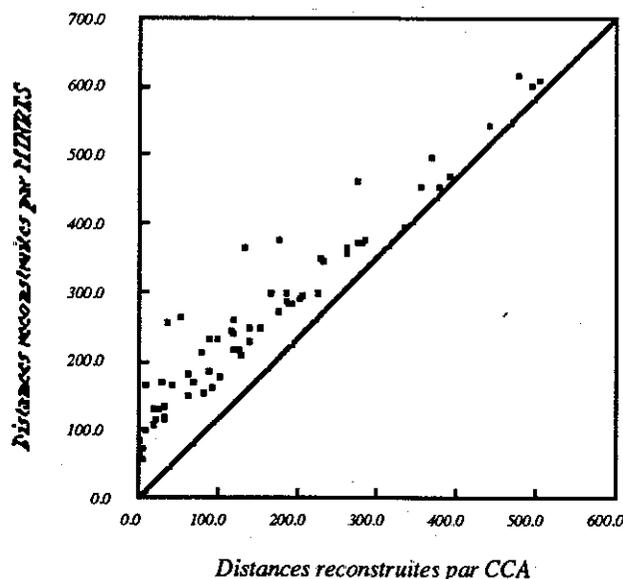


Figure 4 Comparaison des distances construites par CCP (abscisse) et par Codage Multi-dimensionnel MINRES (ordonnée)

La méthode de Codage en Composantes Principales n'est certes pas la seule solution pour construire une approximation interne de R . d'Aubigny (1989) fournit une classe de méthodes concurrentes possibles. La pratique actuelle semble cependant se limiter à ce cas, sans doute pour des raisons de disponibilité de logiciels.

Remarque: Les théorèmes 2 et 3 suggèrent de mesurer l'adéquation d'une configuration aux données en termes analogues à l'ACP. On calcule alors la proportion d'inertie de la matrice de distance $D(A)$ expliquée par une solution de dimension q à l'aide de l'un des rapports suivants:

$$\begin{aligned}\alpha_{1,q} &= \frac{\sum_{s=1}^q l_s}{\sum_{s=1}^p |l_s|} \times 100, \\ \alpha_{2,q} &= \frac{\sum_{s=1}^q l_s^2}{\sum_{s=1}^p l_s^2} \times 100\end{aligned}\quad (11)$$

L'exemple des villes anglaises donne pour $q = 2$, $\alpha_{1,q} = 92,6\%$ et $\alpha_{2,q} = 99,8\%$

Un calcul simple montre que $\Phi = 2p(l_{q+1} + \dots + l_p)$. Le caractère un peu artificiel du critère Φ montre la nécessité de caractériser la méthode dans le cadre de critères plus usuels, par exemple en termes de moindres carrés. On observe alors que le critère cherché porte naturellement sur l'approximation des produits scalaires.

En effet, étant donnée R une matrice Euclidienne ou non, on se propose d'approcher la matrice $B = Q_1 H Q_1$ au sens des moindres carrés par une matrice de produits scalaires $\widehat{B} = B(X)$ de façon à minimiser la fonction critère Ψ donnée sous la forme

$$\begin{aligned}\Psi &= \sum_{j,k=1}^p (b_{jk} - \hat{b}_{jk})^2 p_j p_k, \\ &= \text{Trace}((B - \widehat{B})D_P(B - \widehat{B})D_P)\end{aligned}\quad (12)$$

pour laquelle la solution des CCP est optimale, (cf. d'Aubigny (1989), Fichet (1983), Gower (1983), Mardia (1978) et Mathar (1985)):

Théorème 4 *Si R est une matrice de dissimilarités non nécessairement Euclidienne, pour un q fixé, la fonction critère Ψ est minimisée par rapport à toutes les configurations X de dimension q quand \widehat{X} est la solution classique du problème de CCP.*

2.4 les difficultés rencontrées

2.4.1 Le cas des données de similarité

Les développements présentés ci-dessus posent un problème pratique spécifique lorsque les données sont collectées sous forme de mesures de ressemblance entre objets. Nous adoptons alors la définition suivante:

Définition 3 *Soit S une matrice carrée d'ordre p . On dit que S est une matrice de similarité si elle vérifie*

$$\begin{aligned} (1) \quad & s_{jk} = s_{kj} \\ (2) \quad & s_{jk} \leq s_{jj} \quad \text{pour tout } (j, k) \end{aligned}$$

De nombreuses transformations sont proposées dans la littérature pour déduire une matrice de dissimilarités d'une matrice de similarités donnée (et réciproquement). Nous observons ici qu'il est par exemple possible de transformer une dissimilarité par la transformation standard:

Définition 4 *On appelle transformation standard $S \rightarrow R$ la fonction réelle définie par*

$$r_{jk}^2 = s_{jj} - 2s_{jk} + s_{kk} \quad (13)$$

Le lien entre transformation standard et méthode CCP est fournie dans ce cas par le résultat suivant:

Théorème 5 (Mardia et al. 1979) *Si S est une matrice de similarités à termes positifs ou nuls et si R est déduite de S par la transformation standard, alors R est Euclidienne et la matrice de produits scalaires associée est $B = Q_1 S Q_1$*

Preuve: C'est une conséquence directe de la relation suivante

$$r_{jk}^2 = s_{jj} - 2s_{jk} + s_{kk} = {}^t k_{jk} S k_{jk} \geq 0$$

où k_{jk} est le vecteur de i -ème coordonnée

$$k_{jk}(i) = \begin{cases} +1 & \text{si } i = j \\ -1 & \text{si } i = k \\ 0 & \text{sinon} \end{cases}$$

Remarque: Lorsque la décomposition spectrale de B fournit des valeurs propres

négatives, il n'est pas possible d'interpréter R comme une métrique sur E . On est donc amené à analyser B comme la matrice associée à une application linéaire définie sur E^* et à valeurs dans E , dont la décomposition singulière donne des facteurs directs et des facteurs inverses (cf. Benzécri 1973) :

$$B = UD_l {}^tV = UD_l D_\epsilon {}^tU \quad (14)$$

où

$$D_\epsilon = \text{diag}(\epsilon_s)$$

En effet, B est une matrice symétrique dont les vecteurs singuliers associés à la valeur singulière l_s vérifient

$$v^s = \epsilon_s u^s \quad \epsilon_s = \begin{cases} +1 & \text{facteur direct} \\ -1 & \text{facteur inverse} \end{cases}$$

Par suite, dans le cadre d'une analyse linéaire, regarder B comme la matrice d'une forme bilinéaire sur E^* , sans considérer pour autant qu'elle définit un produit scalaire, conduit à réaliser l'analyse spectrale de B^2 . Or :

$$B^2 = UD_l D_\epsilon {}^tUUD_\epsilon D_l {}^tU = UD_l^2 {}^tU$$

Par suite, analyser B^2 conduit à ne pas retenir les mêmes dimensions pour reconstruire R que la solution habituelle et mathématiquement peu fondée qui néglige les dimensions associées à des valeurs propres négatives de B . Celles ci apparaissent dans la matrice diagonale $D = D_l D_\epsilon$ obtenue dans de nombreux programmes standards de décomposition spectrale de B et l'ordre des termes diagonaux de D et D_l^2 diffèrent. // L'élimination des valeurs propres négatives par adjonction d'une *constante additive* résoud certes le problème numérique rencontré, mais cet échappatoire est dépourvu de justification méthodologique.

3 Codage Multidimensionnel du triple $(J, R_{J \times J}, W_{J \times J})$

Une faiblesse commune aux approches algébriques présentées dans la section précédente (cf. aussi d'Aubigny (1989)) réside dans l'affirmation implicite de l'absence d'erreurs de mesure. R est regardée comme une donnée que l'on analyse de façon indirecte au travers de B . Pourtant, faute de vérifier l'inégalité triangulaire, R est en général une matrice de pseudo-distances que l'on ne sait pas relier simplement à une matrice de produits scalaires. Ce cas conduit donc à rechercher les moyens d'analyser directement les dissimilarités observées.

Une deuxième faiblesse de l'approche précédente est liée aux difficultés inhérentes à l'existence de données manquantes. Aucune solution naturelle n'apparaît à cause du caractère indirect de toute méthode d'analyse fondée sur la factorisation de B .

Une troisième faiblesse est liée à la difficulté à construire un modèle d'erreurs naturel à la fois pour le tableau de dissimilarités et pour celui des produits scalaires.

$$\begin{aligned} R_{jk}^2 &= \delta_{jk}^2 + \epsilon_{jk}^2 \Rightarrow b_{jk}(R) = b_{jk}(\Delta) + b_{jk}(E) \\ b_{jk} &= \beta_{jk} + \tau_{jk} \Rightarrow R_{jk}(B) = ? \end{aligned} \quad (15)$$

On observe de plus que la relation (15) s'exprime entre carrés des distances Euclidiennes et produits scalaires et non entre distances et produits scalaires. Cette propriété est une conséquence du théorème de Pythagore. On retrouve ici une difficulté courante en statistique lorsqu'on se propose de travailler sur les écarts-types et non sur les variances. La classe de méthodes présentées dans cette section traite de façon directe le problème d'approximation des tableaux de dissimilarités. Le tableau \mathbf{R} est alors interprété comme définissant une mesure positive symétrique, définie sur les arêtes d'un graphe complet (en l'absence de données manquantes) défini sur J . On suppose de plus donnée une mesure a priori $W_{J \times J}$ sur ce graphe. La donnée du problème est alors résumée par le triple $(J, R_{J \times J}, W_{J \times J})$ dont l'analyse pose un problème d'approximation, dit de Codage Multidimensionnel. Nous retenons ici l'acronyme MDS, abréviation de l'intitulé anglais de Multidimensional Scaling.

d'Aubigny (1989) montre que le cas où $W_{J \times J}$ factorise en $P_J \otimes P_J$, conduit à des simplifications importantes permettant de se ramener au cadre d'analyse de la section précédente. Ce cas est important en pratique où souvent on choisit implicitement $p_j = \frac{1}{p}$, $\{\forall j = 1, 2, \dots, p\}$. Nous traitons ici le cas général.

On pose dans la suite $\Delta = D(\mathbf{A})$, id est $\delta_{jk} = d_{jk}(\mathbf{A})$, de sorte que δ_{jk} figure la distance Euclidienne supposée par le modèle. On définit de plus un modèle d'erreurs en supposant par exemple que les distances vérifient l'un des modèles suivants. Le modèle le plus simple est:

Modèle 1: Erreur additive

$$r_{jk} = \delta_{jk} + \epsilon_{jk}$$

où ϵ_{jk} représente les erreurs de mesures et les erreurs de spécification qui proviennent du fait que les r_{jk} ne correspondent pas exactement à une configuration dans \mathbb{R}^n . Ce modèle est caractérisé par le fait que ϵ dépend implicitement de la configuration théorique \mathbf{A} :

$$r_{jk} = d_{jk}(\mathbf{A}) + e_{jk}(\mathbf{A}) \quad \epsilon_{jk} = e_{jk}(\mathbf{A})$$

Cependant, de nombreuses situations rendent plus réaliste l'hypothèse d'une relation moins rigide liant les $d_{jk}(\mathbf{A})$ aux r_{jk} . On considère alors le modèle plus général :

Modèle 2: Erreur décrivant une structure statistique de type exponentiel

$$r_{jk} = f(\tau_{jk})$$

où $\tau_{jk} = E(R_{jk})$ est l'espérance mathématique de la variable aléatoire R_{jk} dont la loi définit une structure exponentielle et f est une fonction monotone croissante et différentiable. Dans la littérature psychométrique, τ_{jk} est appelée la disparité de j et k de sorte que si l'on pose ici la

Définition 5 On appelle fonction de disparité, toute transformation monotone croissante d'une fonction de dissimilarité donnée.

Le modèle probabiliste adopté ici est une extension du modèle (1), pour lequel chaque sujet observé ω dans une population Ω fournit un tableau de proximités $\mathbf{R} = (r_{jk})$ où $r_{jk} = R_{jk}(\omega)$, de sorte que le modèle d'erreur est décrit par le choix de la fonction

f retenue, c'est à dire par le choix d'une famille paramétrée de lois de probabilités appartenant à la classe des structures exponentielles. La seule information empirique reconnue comme pertinente pour estimer les δ_{jk} , est alors fournie par les relations de préordre entre paires du type

$$r_{jk} \leq r_{j'k'} \Rightarrow d_{jk}(\mathbf{X}) \leq d_{j'k'}(\mathbf{X}) \text{ pour tout } i, j, i', j'$$

Il conduit donc à des jugements comparatifs du type: Londres est plus proche de Brighton que ne l'est Hull de Leeds. On parle alors de méthodes de *Codage Multidimensionnel Ordinal* (ou CMO). Une heuristique fut proposée par Shepard (1962 a,b) et rationalisée par Kruskal, dans le cadre du formalisme statistique plus traditionnel de la régression isotone au sens des moindres carrés. L'algorithme de Kruskal consiste en fait à itérer deux phases répondant à deux objectifs distincts :

- (1) Connaître la configuration \mathbf{S}^ν pour une transformation fixée $f^{\nu-1}$,
- (2) Connaître un codage optimal f^ν des dissimilarités pour une configuration fixée \mathbf{S}^ν , c'est à dire la disparité \mathbf{T}^ν associé à \mathbf{R} par f^ν .

Si \mathbf{X} est une configuration de points de \mathbb{R}^q et si $\widehat{\mathbf{D}}^q = \mathbf{D}(\mathbf{X}) = \mathbf{D}(\widehat{\mathbf{A}})$ est la matrice de distances associée, Kruskal propose d'optimiser la fonction d'écart appelée le STRESS et définie en $\mathbf{X} = \widehat{\mathbf{A}}$ par la formule :

$$STRESS(\mathbf{X}) = \min_t \left\{ \frac{\sum_{j < k} (t_{jk} - d_{jk}(\mathbf{X}))^2}{\sum_{j < k} d_{jk}^2(\mathbf{X})} \right\}$$

où les disparités t décrivent un cône polyédral \mathcal{C} dans \mathbb{R}^{p*} ($p* = \frac{p(p-1)}{2}$) formé par les fonctions monotones croissantes de $r : t = m(r)$. Ici, la fonction t définit la matrice symétrique de disparités \mathbf{T} de terme général t_{jk} et t est de plus un codage isotone de r , lorsque m est de plus bijective.

On construit ainsi la transformation monotone de r qui est la plus proche (au sens des moindres carrés) d'une distance Euclidienne entre p points plongés dans un espace Euclidien de dimension q . L'optimum est donc atteint lorsque le STRESS est nul. Cela se produit très rarement en pratique et une valeur quasi nulle du STRESS traduit en général l'obtention d'une solution dégénérée.

Remarques: (a) Le STRESS est en particulier invariant par toute transformation du groupe des déplacements:

$$\mathbf{Y} = \mathbf{X}\mathbf{O} + \mathbf{T}$$

où \mathbf{O} est une matrice orthogonale et $\mathbf{T} = \mathbf{1}^t$ représente une translation.

(b) Le présence du dénominateur permet de standardiser le STRESS et rend cette fonction objectif invariante par transformation de la forme $y_s = cx_s$, $c > 0$, c'est à dire par homothétie.

(c) Pour chaque dimension q fixée, les configurations qui réalisent la valeur minimum du STRESS ne sont pas nécessairement uniques à un déplacement près. Elles sont dites meilleures configurations en dimension q et l'on note $\widehat{\mathbf{S}}(q)$ la valeur $\min_{\mathbf{X}(q,p)} \mathbf{X}(\widehat{\mathbf{A}})$ de ce STRESS minimum.

De plus, d'Aubigny (1989) montre que les solutions du problème CMO associées à des choix de valeurs consécutives de q ne sont pas emboîtées pour des raisons mathématiques fondamentales.

(d) Pour déterminer la dimension la plus satisfaisante, on trace le graphe de la fonction $q \rightarrow \hat{S}(q)$ et l'on choisit la dimension optimale q_0 de façon que $\hat{S}(q_0)$ soit jugé assez petit. La configuration retenue comme solution du problème réalise alors $\hat{S}(q_0)$.

Kruskal (1964) donna des règles empiriques de choix de $\hat{S}(q_0)$ qui apparaissent maintenant très peu adaptées et il est préférable d'avoir recours aux tables établies par simulation et multiplication des configurations de départ.

Compte tenu de leur importance, un grand nombre de travaux leur furent consacrés, parmi lesquels nous citons Arabie (1973), Clark (1976), Isaac et Poor (1974), Klar (1969), Levine (1978), Spence (1972,1974), Spence et Graef (1973, 1974), Spence et Ogilvie (1973), Spence et Young (1978), Stenson et Knoll (1969), Wagenaar et Padmos (1971). Le problème posé par la recherche du minimum de la fonction STRESS

est complexe du point de vue de l'analyse numérique et il n'existe actuellement pas de méthode uniformément meilleure. Les travaux de de Leeuw (1977a,b,1980, 1984, 1988), de Leeuw et Stoop (1984), de Leeuw et Heiser (1977) et Chine (1982) permettent d'évaluer la complexité de ce problème de minimisation d'une fonction $D - C$ (Différence de deux fonctions convexes) non différentiable, pour lequel on ne peut garantir que la convergence (lente) vers un optimum local. Les travaux récents portent sur la recherche de protections contre la captation de l'itération dans le bassin d'attraction d'un des optima locaux. On recourt alors soit à des méthodes de recuit simulé (Machmouchi (1992)), soit à des méthodes de types SEM. (Machmouchi(1992) adapte à ce problème les approches de Celeux (1988) et Celeux et Diebolt (1986)). Ces travaux portent aussi sur la recherche d'accélérateurs de procédures numériques de résolutions de problèmes de type $D - C$ (Chine (1991)).

Nous abordons ici le problème de façon élémentaire. tout d'abord nous présentons une approche heuristique du problème, proche dans son esprit de l'approche de Shepard. Puis, la section suivante décrit l'approche adoptée par de Leeuw et Heiser (1977) et implémentée dans SMACOF, le logiciel commercial le plus satisfaisant sur le plan numérique, actuellement sur le marché. On note ainsi la forte parenté qui existe entre les deux approches.

3.1 Présentation heuristique des méthodes de MDS

Supposons que l'écart entre dissimilarité initiale et distance reconstruite à l'aide de la configuration \mathbf{X} est évalué à l'aide des disparités $t_{jk} = m(r_{jk})$, par la quantité

$$e_{jk} = t_{jk} - d_{jk}(\mathbf{X})$$

L'heuristique proposée par Shepard (1962) repose sur le principe suivant :

Etant donnée une configuration \mathbf{X} , chaque objet j est représenté par le point de coordonnée \mathbf{x}_j et l'on réduit l'écart e_{jk} lorsqu'on déplace le point j dans la direction du point k . Le nombre e_{jk} définit le sens et la valeur du vecteur du déplacement à

partir de j et à destination de k , de sorte que l'étude des positions de j et k conduit au facteur de correction :

$$f_{jk} = \frac{e_{jk}}{d_{jk}(\mathbf{X})} = 1 - \frac{t_{jk}}{d_{jk}(\mathbf{X})}$$

Ce facteur de correction permet d'écrire que le mouvement du point j relativement au point k résulte des mouvements v^s le long de l'axe s .

$$v_{jk}^s = f_{jk}(x_j^s - x_k^s) \quad s = 1, 2, \dots, n$$

ceci pour toute paire (j, k) . On associe donc à tout point j le déplacement résultant

$$v_j^s = \frac{1}{p} \sum_{k=1}^n v_{jk}^s = \frac{1}{p} \sum_{k=1}^n \left(1 - \frac{t_{jk}}{d_{jk}(\mathbf{X})}\right) (x_j^s - x_k^s)$$

où par convention, on a posé pour tout j

$$\frac{t_{jj}}{d_{jj}(\mathbf{X})} = 0$$

La nouvelle coordonnée du point représentant l'objet j sur l'axe s est donnée à l'itération $(\nu + 1)$ par moyennage

$$x_j^{s(\nu+1)} = x_j^{s(\nu)} + v_j^{s(\nu)}$$

L'interprétation de cette heuristique dans un cadre d'analyse numérique plus formel, repose en fait sur la constatation suivante

Lemme 1 Si

$$\sigma(\mathbf{X}) = \sum_j \sum_k (t_{jk} - d_{jk}(\mathbf{X}))^2$$

alors le gradient de $\sigma(\mathbf{X})$ a pour s -ème coordonnée

$$\frac{\partial}{\partial x_j^s} \sigma(\mathbf{X}) = 2pv_j^s$$

Cela signifie que l'on peut écrire l'heuristique proposée originellement par Shepard sous la forme d'une méthode de gradient à pas constant:

$$x_j^{s(\nu+1)} = x_j^{s(\nu)} - \frac{1}{2p} \left[\frac{\partial}{\partial x_j^s} \sigma(\mathbf{X}) \right]^{(\nu)}$$

qui fait peu de cas de la non différentiabilité de $\sigma(\mathbf{X})$ aux points $\mathbf{X}^{(\nu)}$ pour lesquels $d_{jk}(\mathbf{X}^{(\nu)}) = 0$ et $j \neq k$.

On retrouve ainsi, sous l'angle numérique, l'approche de Kruskal (1964) qui se caractérise par une gestion particulièrement efficace du pas d'itération $\alpha(\nu)$. Les recherches menées depuis, en particulier par de Leeuw et Heiser (1977, 1980) à la suite des travaux

de Guttman (1968), sont toutes formulables comme des élaborations et des raffinements de cette méthode itérative de base.

$$\mathbf{x}^{s(\nu+1)} = \mathbf{x}^{s(\nu)} - \alpha(\nu) \left[\frac{\partial}{\partial x_j^s} \sigma(\mathbf{X}) \right]^{(\nu)}$$

destinée à calculer la configuration \mathbf{X} solution du problème

$$\text{Min}\{\sigma(\mathbf{X}), \mathbf{X} \in \mathcal{M}_{p \times q}(\mathbf{R}), \mathbf{X} \text{ centrée}\}$$

$\sigma(\mathbf{X})$ est appelé RAW STRESS dans la littérature. Le cas général est donné par

Définition 6 Soient \mathbf{W} une pondération définie sur $J \times J$ et $(J, R_{J \times J}, W_{J \times J})$ le triplet analysé, on appelle STRESS brut la fonction

$$\mathbf{X} \rightarrow \sigma(\mathbf{X}) = \sum_j \sum_k w_{jk} (t_{jk} - d_{jk}(\mathbf{X}))^2$$

où la disparité $t_{jk} = m(r_{jk})$ est une transformation des données brutes r_{jk} par une fonction monotone m .

3.2 Un algorithme de résolution de problème CMM

Nous nous plaçons maintenant dans le cas simple où la transformation des données adoptée est l'identité. On parle alors de cas *métrique* et on utilise l'acronyme CMM. Cela ne remet pas en cause la généralité de cette approche proposée initialement par Guttman (1968) puis formalisée et affinée par de Leeuw et Heiser (1977a). Le cas d'une transformation quelconque s'obtient en effet (cf. infra) par l'ajout d'une étape d'estimation fonctionnelle à chaque itération.

D'autre part, le cas où la structure statistique présentée dans le modèle 2 est connue a priori, se ramène au cas métrique, après transformation des données. Soit donc :

$$\sigma(\mathbf{X}) = \sum_j \sum_k w_{jk} (r_{jk} - d_{jk}(\mathbf{X}))^2$$

Nous posons alors les notations,

$$\eta^2(\mathbf{X}) = \sum_j \sum_k w_{jk} d_{jk}^2(\mathbf{X})$$

$$\rho(\mathbf{X}) = \sum_j \sum_k w_{jk} r_{jk} d_{jk}(\mathbf{X})$$

\mathbf{R} et \mathbf{W} sont les données du problème, de sorte que l'on peut, sans perte de généralité, supposer satisfaite la contrainte de normalisation :

$$\sum_j \sum_k w_{jk} r_{jk}^2 = 1$$

Une manipulation algébrique simple montre de plus que :

$$\begin{aligned} \eta^2(\mathbf{X}) &= \sum_{s=1}^p \sum_j \sum_k w_{jk} (x_j^s - x_k^s)^2, \\ &= \text{Trace}({}^t \mathbf{X} \mathbf{M} \mathbf{X}) = \text{Trace}(\mathbf{W}_X) \\ &= \text{Trace}(\mathbf{M} \mathbf{X} {}^t \mathbf{X}) = \text{Trace}(\mathbf{M} \mathbf{V}_X) \end{aligned}$$

où M est la matrice Laplacienne du graphe complet valué $(J, W_{J \times J})$ de terme général

$$\begin{aligned} m_{jk} &= -2w_{jk}, \\ m_{jj} &= \sum_{k \neq j}^p 2w_{jk} \end{aligned}$$

de sorte que $\eta^2(\mathbf{X})$ est une fonction objectif usuellement optimisée par les méthodes classiques d'Analyse linéaire des données. Mais on traite ici le problème inverse, où \mathbf{X} est une inconnue. De plus

$$\begin{aligned} \rho(\mathbf{X}) &= \sum_j \sum_k \frac{w_{jk} r_{jk}}{d_{jk}(\mathbf{X})} d_{jk}^2(\mathbf{X}), \\ &= \sum_{s=1}^p \sum_j \sum_k b_{jk} x_j^s x_k^s = \text{Trace}({}^t \mathbf{X} \mathbf{B}(\mathbf{X}) \mathbf{X}) \end{aligned}$$

où $\mathbf{B}(\mathbf{X})$ est la matrice symétrique de terme général

$$\begin{aligned} b_{jk}(\mathbf{X}) &= -2 \frac{w_{jk} r_{jk}}{d_{jk}(\mathbf{X})}, \\ b_{jj}(\mathbf{X}) &= 2 \sum_{k \neq j}^p \frac{w_{jk} r_{jk}}{d_{jk}(\mathbf{X})} \end{aligned}$$

$\mathbf{B}(\mathbf{X})$ définit donc une métrique Riemannienne sur E . Pour une matrice \mathbf{R} donnée, le problème s'écrit sous l'une des formes équivalentes

- (1) $\text{Min}\{\sigma(\mathbf{X}) ; \mathbf{X}\}$
- (2) $\text{Min}\{\eta^2(\mathbf{X}) - \rho(\mathbf{X}) ; \mathbf{X}, \mathbf{X}\mathbf{1} = \mathbf{0}\}$

dont la solution vérifie l'équation de point fixe

$$\mathbf{X} = \mathbf{M}^+ \mathbf{B}(\mathbf{X}) \mathbf{X}$$

où \mathbf{M}^+ est l'inverse généralisée de \mathbf{M} :

$$\mathbf{M}^+ = \left(\mathbf{M} + \frac{1}{p} \mathbf{1} \mathbf{1}^t \right)^{-1} - \frac{1}{p} \mathbf{1} \mathbf{1}^t$$

La forme (2) définit un problème $D - C$, ou de minimisation de la différence de deux fonctions convexes.

Dans le cas général, on doit aussi estimer la transformation $t = m(r)$. Par suite, pour une dimension n fixée, on résoud le problème numérique posé par les méthodes de codage multidimensionnel en itérant les deux phases suivantes:

(Ph1) Pour $\mathbf{X}^{(\nu)}$ fixée, trouver la transformation m optimale dans \mathcal{C} , le cône des transformations (monotones) admissibles de \mathbf{R} . On recherche alors $\tau^{(\nu)}$, c'est à dire le codage optimal des dissimilarités au sens des moindres carrés. Cette phase donne son nom à la méthode.

(Ph2) Pour $m^{(\nu)}$ fixé, c'est à dire pour une disparité fixée, trouver la configuration de rang q optimale, c'est à dire la configuration $\mathbf{X}^{(\nu+1)}$ solution du problème sous l'une

des formes équivalentes (1) et (2).

Remarque : La méthode numérique décrite ci-dessus appartient à la classe des méthodes de sous-gradient, classe à laquelle appartient aussi l'adaptation de l'algorithme EM proposée par Machmouchi (1992). Elle souffre de trois types de pathologies liées entre elles :

- (1) une vitesse de convergence lente,
- (2) une sensibilité au choix d'une initialisation de la procédure,
- (3) la convergence est assurée vers un optimum local.

On observe alors la dépendance de la vitesse de convergence et le choix de la configuration initiale.

Boughazi (1987), Chine (1982), de Leeuw (1977, 1988), Ramsay (1969) et Yassine (1989) apportent des réponses partielles au point (1). Ils produisent des procédures efficaces d'accélération de l'itération. Le point (3) a fait l'objet de nombreuses études théoriques, en particulier sous l'impulsion de de Leeuw (1980, 1988). Le point (2) est souvent ignoré des utilisateurs qui se contentent à tort d'un seul choix de configuration initiale, souvent même fourni de façon automatique par le logiciel. Par suite, nombreux sont les articles d'application qui produisent des solutions non optimales de façon patente. Les utilisateurs de logiciels commerciaux du marché devraient prendre au sérieux cette faiblesse commune à tous les algorithmes résolutoires utilisés et ne devraient jamais se contenter de retenir une seule configuration initiale, bien qu'il soit difficile de faire preuve d'imagination et de réalisme en la matière.

Machmouchi (1992) propose une procédure particulièrement convaincante de construction adaptative des configurations initiales, fondée sur un algorithme de perturbation des optima locaux à l'aide d'une procédure de recuit simulé.

La recherche de configurations assurant des valeurs aussi petites que possibles de la fonction STRESS doit de plus être menée avec prudence. En effet, en particulier lorsque le rapport p/q est trop petit ou lorsque la dissimilarité étudiée R est très peu Euclidienne. Les algorithmes de codage optimal produisent de façon systématique des solutions dites dégénérées, c'est à dire pour lesquelles les contraintes de type inégalité traduisant la monotonie de la fonction liant disparités et dissimilarités, sont réalisées sous forme d'égalités, repérables par des structures géométriques régulières obtenues par l'itération (triangles équilatéraux, carrés, pyramides équilatérales, etc...).

El Hadri (1988) (voir aussi Winsberg et Ramsay (1983)) montre l'apport de transformations monotones lisses (bases de splines monotones, etc...) pour protéger l'itération des risques de solutions dégénérées. Les propriétés de convergence de ces itérations sont étudiées dans une série de thèses soutenues à Grenoble et dans les travaux de de Leeuw (1980,1988) et de Mathar (1985).

3.3 Méthode optimale au sens MINRESS

La méthode classique résultant des travaux de Shepard, Kruskal et Guttman conduit à un problème numérique difficile, nous venons de le voir. Une voie possible d'amélioration du comportement des procédures numériques passe par une modification de la fonction objectif. C'est pourquoi Takane, Young et de Leeuw (1977)

proposent une méthode alternative qui tend à estimer au mieux non pas \mathbf{R} mais la matrice des carrés des dissimilarités \mathbf{H} où

$$h_{jk} = r_{jk}^2$$

L'amélioration espérée provient de deux caractéristiques du problème alors traité. Tout d'abord, du fait que le carré des distances Euclidiennes se décompose de façon additive

$$\delta_{jk}^2 = d_{jk}^2(\mathbf{A}) = \sum_{s=1}^n (\lambda_j^s - \lambda_k^s)^2$$

les analyses algébrique et géométrique du problème sont simplifiées par application du théorème de Pythagore. D'autre part, la mise en oeuvre numérique de la phase 2 de l'algorithme de résolution du problème d'optimisation sous-jacent est mieux contrôlée ici, puisqu'elle est interprétable comme une projection sur le cône des matrices de carrés de distances Euclidiennes de format $p \times p$. d'Aubigny (1975) et Fichet (1983) montrent que cet ensemble est convexe, contrairement au cas du cône des matrices de distances Euclidiennes de format $p \times p$. Dans les deux cas cependant, la difficulté majeure provient du caractère non polyédral de ce cône.

La fonction objectif correspondant à une résolution du problème au sens des moindres carrés conduit alors à minimiser la fonction SSTRESS.

Définition 7 Soient \mathbf{W} une pondération définie sur $J \times J$ et $(J, R_{J \times J}, W_{J \times J})$ le triplet analysé. On appelle SSTRESS la fonction de \mathbf{X}

$$\mathbf{X} \longrightarrow s(\mathbf{X}) = \sum_j \sum_k w_{jk} (t_{jk}^2 - d_{jk}^2(\mathbf{X}))^2$$

où la disparité $t_{jk} = m(r_{jk})$ est une transformation des données brutes r_{jk} par une fonction monotone m .

Les auteurs montrent qu'alors, la construction d'un algorithme en deux phases nécessite l'évaluation des zéros d'un polynôme de degré 3 à chaque étape, c'est à dire la résolution d'un problème numérique simple. Cette méthode est implantée dans le logiciel ALSCAL.

Cette approche du problème ne doit cependant pas être considérée comme le point ultime des recherches dans ce domaine, du fait du manque de robustesse accru de cette méthode aux données extrêmes. Les risques d'obtenir une configuration essentiellement déterminée par ces points extrêmes, c'est à dire situés à une distance relativement grande du barycentre (mesurées par des grandes valeurs $d_{jk}^2(\mathbf{X})$ pour $k \neq j$) au détriment des points rapprochés qui contribuent alors relativement peu à la fonction objectif sont sensiblement accrus par comparaison aux méthodes de types MINRES. Il convient donc de redoubler d'esprit critique lorsqu'on utilise un logiciel fondé sur l'optimisation de la fonction SSTRESS.

La comparaison des solutions obtenues respectivement par la minimisation du STRESS et du SSTRESS, montre que bien que proches en général (à une transformation du groupe des déplacements près), ces solutions ne sont pas identiques. Il est donc important d'identifier clairement l'approche qui est adoptée par le logiciel du marché utilisé. De plus, les premières explorations numériques effectuées conduisent aux deux

conjectures suivantes, exprimées en fonction des solutions optimales \mathbf{X}^S et \mathbf{X}^{SS} respectivement obtenues par minimisation du STRESS et du SSTRESS et lorsque l'on suppose déterminée de façon correcte la dimension q de l'espace de représentation :

(1) La résolution du problème MINRES à partir de la configuration \mathbf{X}^{SS} conduit en général à une amélioration de $\sigma(\mathbf{X})$: $\sigma(\mathbf{X}^S) \leq \sigma(\mathbf{X}^{SS})$

Par contre, la résolution du problème MINRESS à partir de la configuration \mathbf{X}^S ne semble pas conduire à une amélioration de $s(\mathbf{X})$: $s(\mathbf{X}^{SS}) = s(\mathbf{X}^S)$.

(2) L'optimum \mathbf{X}^{SS} semble appartenir à l'ensemble des optima locaux de la fonction σ .

Enfin d'Aubigny (1989) montre par un argument algébrique simple que la solution du problème MINRESS est plus proche de la solution du problème CCA que ne l'est la solution du problème MINRES.

3.4 Le Modèle général: Estimation au sens MINRES

Les deux méthodes précédentes diffèrent par le choix de la fonction objectif retenue et non par le modèle retenu de décomposition de la dissimilarité. Sans présenter d'analyse détaillée du problème, nous remarquons que de nombreux modèles concurrents sont possibles. En particulier d'Aubigny (1989) montre que le modèle classique de l'analyse factorielle (en facteurs communs et facteurs spécifiques) se traduit en termes de distances Euclidiennes par le modèle

$$r_{jk}^2 = d_{jk}^2(\mathbf{X}) + \alpha_j + \alpha_k + \epsilon_{jk} \quad (16)$$

Il correspond donc à un modèle particulier de décomposition du terme d'interaction dans un plan d'expérience factoriel d'ordre 2 sous la forme standard, adapté à l'hypothèse de symétrie de la mesure.

Le modèle (16) fut proposé en psychométrie par Carroll (1983) puis Weinsberg et Carroll (1985). Saito (1985, 1986), proposa d'intégrer des effets de contexte dans le modèle suivant, d'une mise en oeuvre beaucoup plus complexe, voir aussi Le Calvé (1985):

$$r_{jk} = d_{jk}(\mathbf{X}) + \alpha_j + \alpha_k + \epsilon_{jk}$$

Encore une fois, le caractère additif de la décomposition dimensionnelle d'une distance Euclidienne explique ces différences.

Dans les deux cas, on observe que l'on peut écrire la fonction objectif à minimiser sous la forme générale :

$$\Psi(\mathbf{X}) = \sum_{j < k} w_{jk} (m(r_{jk}) - \gamma(d_{jk}(\mathbf{X}); \theta))^2 \quad (17)$$

Par exemple, le modèle (16) conduit à poser

$$\gamma(d_{jk}(\mathbf{X}); \theta) = [d_{jk}^2(\mathbf{X}) + \alpha_j + \alpha_k]^{\frac{1}{2}}$$

et le paramètre θ représente les unicités $\{\alpha_j ; j = 1, 2, \dots, p\}$.

Cette forme générale suppose donnée la famille paramétrée de courbes $\gamma(\cdot; \theta)$. Elle

montre qu'une grande classe de modèles de Codage multidimensionnel sont estimables par une procédure de type MINRES, pourvu que soient résolus les problèmes d'identification des paramètres du modèle. D'autre part, le recours à une transformation monotone m des dissimilarités observées caractérise cette approche comme un élément de la classe des méthodes d'ajustement de modèles semi-paramétriques.

Remarque : La forme (17) permet d'étendre l'utilisation de ces méthodes à l'étude des processus stochastiques spatiaux discrets, non nécessairement stationnaires ou isotropes. La seule contrainte imposée par ce modèle est alors la finitude du semi-variogramme $\gamma(d; \theta)$. Cependant, l'expression de γ comme fonction de d définit un modèle isotrope et donc l'écart évalué par Psi est en partie lié à l'anisotropie du processus étudié.

La richesse des modèles de semi-variogrammes qu'il est possible d'ajuster par une telle approche ne s'arrête pas au niveau de généralité du modèle (16). En fait le recours à un logiciel tel que PolyCon (Polynomial Conjoint measurement) qui met en oeuvre une large classe de modèles dits de mesures conjointes permet une plus grande généralité dans le choix d'un modèle.

4 Extensions

Les logiciels présentés dans la section suivante proposent un grand nombre de développements du modèle de base présenté ci-dessus. Nous discutons (trop!) brièvement les deux principales voies d'extension.

4.1 Analyse d'un ensemble de triples $(J, \{R_{J \times J}^i, i = 1, 2, \dots, n\}, W_{J \times J})$ et Analyse des Différences Individuelles

Sous le nom générique d'*Analyse des Différences Individuelles* s'est développée une classe de méthodes d'analyse d'un n -échantillon de tableaux de dissimilarités, à l'initiative de Carroll et Chang (1970), Harshman (1970), Heiser et Van Stoop (1986), Krane (1978), MacCallum (1979), McGee (1968), Takane, Young et de Leeuw (1977). Le modèle sous-jacent est algébrique et décrit les différences individuelles sous la forme d'un n -échantillon de métriques définies sur l'espace vectoriel G :

$$\begin{aligned} d_{i,jk}^2(\Lambda) &= \|\lambda_j - \lambda_k\|_{D_{\pi_i}}^2 \\ &= {}^t(\lambda_j - \lambda_k) D_{\pi_i} (\lambda_j - \lambda_k) \\ &= \sum_{s=1}^q \pi_i^s (\lambda_j^s - \lambda_k^s)^2 \end{aligned} \quad (18)$$

où les n matrices D_{π_i} sont diagonales de terme diagonal π_i^s , ($s = 1, \dots, q^0$). q^0 est ici la dimension optimale inconnue de l'espace de représentation. Pour une dimension q fixée ($q \leq q^0$), on construit donc une approximation X de Λ et une approximation D_P de D_{π_i} .

L'interprétation des coefficients π_i^s en termes de métrique associée à chaque sujet ω_i , pour $i = 1, \dots, n$ dans l'espace G de représentation des objets n'est pas classique. Pourtant elle justifie les contraintes de positivité qui sont en général imposées aux π_i^s .

Cependant, d'Aubigny (1989) montre qu'en plus, pour $(\mathbf{X}, \mathbf{D}_{\pi_i})$ donné, on obtient pour image du vecteur $\mathbf{1}$ de G dans E :

$$\mathbf{y}_i = \mathbf{X} \mathbf{D}_{\pi_i} \mathbf{1}$$

\mathbf{y}_i est le barycentre du nuage des points \mathbf{x}_j de E munis des poids π_i^s , si ceux-ci sont positifs. C'est une combinaison affine des \mathbf{x}_j dans le cas contraire, c'est à dire lorsque le sujet ω_i traite certaines dimensions comme des facteurs inverses. Il y a alors identité entre l'Analyse des Différences Individuelles et l'Analyse des Points de Vue, plus classique en psychométrie.

Cette approche algébrique traite les sujets comme des éléments différenciés par l'importance qu'ils attachent à chaque axe de représentation de la configuration commune. On estime donc un paramétrage particulier de l'effet sujet. Elle s'oppose à une approche stochastique qui regarde les sujets comme les éléments échangeables d'un échantillon analysé comme n répliques de la mesure d'une matrice aléatoire de dissimilarités. Ce deuxième point de vue est adopté de façon informelle par McGee (1968) et explicitement dans un cadre probabiliste par McKay et Zinnes (1971), Ramsay (1977, 1978, 1982), Takane (1978, 1981), Zinnes et McKay (1981). On fait alors une hypothèse sur la loi de probabilité de la matrice aléatoire \mathbf{R} . L'étude des structures statistiques associées est menée par d'Aubigny (1993).

L'intérêt formel de ces approches est cependant contrebalancé par le nombre limité d'échantillons aléatoires de tableaux de dissimilarités collectés et publiés.

4.2 Analyse des préférences $(J, \{R_j^i, i = 1, 2, \dots, n\}, W_{J \times J})$

Une deuxième voie d'extension du modèle consiste à poser des contraintes de comparabilité des mesures r_{jk} . On parle alors d'analyse *locale* ou *conditionnelle* de \mathbf{R} .

On distingue en général deux cas. Le premier utilise une méthode de collecte de l'information par points d'ancrage: chaque sujet est pris comme référence à son tour. Les comparaisons se font alors par référence à ce point d'ancrage, de sorte que les r_{jk} sont comparables uniquement au sein d'une même ligne. On précise alors le conditionnement en notant la mesure faite par référence à l'objet j par r_k^j . Par suite, en termes de comparaisons par paires prises en compte, il y a perte d'information par rapport au cas général et la fonction à optimiser doit être modifiée en conséquences. On calcule dans ce cas

$$STRESS(\mathbf{X}/J) = \sum_{j=1}^p \frac{\sum_{k \neq j} w_{jk} (m(r_k^j) - d_{jk}(\mathbf{X}))^2}{\sum_{k \neq j} w_{jk} d_{jk}^2(\mathbf{X})} \quad (19)$$

Ici encore, $I = J$ et $n = p$.

Le deuxième cas suppose que les objets définissant les lignes et les colonnes du tableau \mathbf{R} sont distincts. On parle alors de sujets ($i = 1, 2, \dots, n$) et d'objets comparés ($j = 1, 2, \dots, p$) et le tableau \mathbf{R} d'ordre $p \times n$ s'interprète dans le cadre d'un modèle distancié comme la donnée (en case (i, j)) des mesures de dissimilarité entre l'objet j et le *point idéal* du sujet i .

On décrit ainsi un cadre d'analyse des préférences, caractérisé par la recherche de deux matrices de configuration \mathbf{X} et \mathbf{Y} d'ordre respectif $p \times q$ et $n \times q$, telles que

$$r_j^i = d_{ij}(\mathbf{Z}) + \epsilon_j^i \quad (20)$$

$${}^t\mathbf{Z} = ({}^t\mathbf{Y}; {}^t\mathbf{X})$$

$$d_{ij}^2(\mathbf{Z}) = \sum_{s=1}^n (y_i^s - x_j^s)^2$$

y_i est le point idéal du sujet i et x_j est le point de G représentant l'objet j . La matrice des paramètres z_k^s du modèle est alors ajustée au sens des moindres carrés et la qualité de l'ajustement est évalué à l'aide de la fonction objectif (19) où l'on change \mathbf{X} en \mathbf{Z} . d'Aubigny (1979) présente une étude détaillée d'une plus large classe de modèles de description géométrique des données de préférence.

5 Les logiciels disponibles

Nous donnons dans cette section une liste des principaux logiciels du marché, qui traitent le problème de codage multidimensionnel.

5.1 Logiciels traitant un tableau de données de proximités

M-D-SCAL-5: MultiDimensionnal SCALing; 1968 ; Kruskal; Bell laboratories. Evaluation brutale: Non interactif, écrit en Fortran 66, avec une rigueur limitée dans le respect des types et des graphiques en mode texte. Il date et sa mise aux normes de programmation modernes nécessiterait un très gros travail. Mais si votre compilateur Fortran l'accepte, cela reste une bonne boîte noire.

POLYCON: Young; Thurstone Psychometric Laboratory, University of North Carolina; 1973; Destiné a faire du Conjoint Measurement il accepte aussi des modèles plus généraux que ceux qui sont abordés dans le présent texte et est donc très utile. Sur le plan de la programmation cependant, il est soumis à la même évaluation que M-D-SCAL 5.

KYST: Kruskal Young Shepard Torgerson; 1973; Kruskal, Young and Seery, Bell laboratories and Thurstone Psychometric Laboratory, University of North Carolina. Réécriture actualisée de M-D-SCAL 5: cela reste la référence en matière de structures de données prises en charges et aides à l'interprétation disponibles à l'époque. Pour les aspects informatiques, même évaluation que pour M-D-SCAL 5. Il n'est plus accepté par les compilateurs Fortran modernes disponibles actuellement à Grenoble et nous avons dû renoncer à trouver les causes des problèmes.

MINISSA: Michigan Israël Netherlands Integrated SSA; 1973; Lingoes et Roskam; Nimegen University et The University of Michigan. Programme de comparaison de méthodes au départ, ce produit fournit en sortie l'un des produits les plus efficaces au plan numérique. Au plan informatique, même évaluation que pour M-D-SCAL 5.

MINITRI: Michigan Israël Netherlands Integrated TRIadic analysis; 1973; Roskam; Nimegen University. Ecrit par le même auteur que MINISSA, il en partage les qualités numérique et est spécialisé dans le traitement de données issues de comparaisons par triples. Même évaluation informatique que pour M-D-SCAL 5.

MRSCAL: MetRic SCALing; 1973; Roskam ; Nimegen University. Même évaluation informatique que pour M-D-SCAL 5.

MULTISCALE: MULTIdimensional SCALing; 1980; Ramsay; McGILL University. L'intérêt de ce programme est de traiter le Problème dans un cadre probabiliste suivant la méthodologie inférentielle proposée par J.O. Ramsay (1977, 1982). FORTRAN 77, écrit pour Micro-ordinateurs, avec un effort limité d'interface à la MS-DOS.

PROSCAL: PRObabilistic SCALing; 1982; McKAY and Zinnes; Indiana University. L'intérêt de ce programme est de traiter le Problème dans un cadre probabiliste suivant la méthodologie inférentielle proposée par McKay et Zinnes (1981). FORTRAN 77, écrit pour mini-ordinateurs, adapté pour micro-ordinateurs, avec un effort limité d'interface à la MS-DOS.

SMACOF: Scaling by Maximizing a Convex Function; 1980; Ecrit en Fortran 77, avec un effort limité d'interface, pas de menus, pas de graphiques haute résolution, de Leeuw et Heiser, 1977. University of Leiden. C'est le meilleur choix actuel en termes d'efficacité numérique.

SSA-I à SSA-IV: Smallest Space Analysis; 1965; Guttman et Lingoes; The University of Michigan. Les méthodes développées par Guttman et Lindoes sont originales sur de nombreux points (choix de fonctions critères, d'algorithmes) et chaque programme est adapté à un mode particulier de collecte des données. Au plan informatique, même évaluation que pour M-D-SCAL 5.

TORSCA: TORgerson SCALing; 1967; Young and Torgerson; Thurstone Psychometric Laboratory, University of North Carolina. Même évaluation que pour M-D-SCAL 5.

SAMSCAL: Simulated Annealing Multidimensional SCALing; 1992; Machmouchi; Université GRENOBLE 2; Programme de comparaison de méthodes intégrées avec menus déroulants, graphiques de suivi d'itération, graphiques en mode haute définition, etc... Ecrit en Turbo Pascal 5, il tourne sur base MS-DOS. Les versions New-S et C, plus interactives et avec des aides diagnostiques à l'interprétation, seront disponibles au courant de l'été. Ecrit avec des standards modernes d'algorithmique.

SEMSCAL: Stochastic Expectation Maximisation SCALing; 1992; Machmouchi; Université GRENOBLE 2. Même standard que SAMSCAL. Ecrit en Turbo Pascal 5, il tourne sur base MS-DOS. Les versions New-S et C seront disponibles au courant de l'été. Même évaluation de la qualité algorithmique.

5.2 Logiciels traitant seulement les échantillons de tableaux de données de proximités

INDSCAL: INdividual Differences SCALing; Carroll and Chang; Bell laboratories. Même évaluation informatique que pour M-D-SCAL 5.

SINDSCAL: Small INDSCAL; Carroll and Pruzanski; Bell laboratories. Même évaluation informatique que pour M-D-SCAL 5.

5.3 Logiciels pouvant traiter un tableau de données de proximités et des échantillons de tableaux de proximités

ALSCAL: Alternating Least squares SCALing; 1977; Takane, Young et de Leeuw; Thurstone Laboratory, University of North Carolina. Même évaluation que pour M-D-SCAL 5 des standards de programmation.

PARAFAC: Harshman; Western Ontario. Même évaluation que pour M-D-SCAL 5 des standards de programmation.

3-MODE SCALING: ; Tucker and McCallum; Ohio State University. Adaptation du modèle de Tucker au cas de données de dissimilarités. Même évaluation informatique que pour M-D-SCAL 5.

5.4 Logiciels généralistes de traitement statistique de données

SPSS: on trouve une version largement étendue de ALSCAL, qui offre ainsi l'un des logiciels les plus complets de Codage Multidimensionnel simple, multiples, ... La prudence est cependant nécessaire, car l'ouvrage récent de Tournois et Dickes (1993) montre que certains utilisateurs peu formés peuvent être dupes d'un programme qui édite les valeurs du STRESS mais optimise le SSTRESS, conduisant ainsi à des solutions non identiques aux autres logiciels. Les sources ne sont pas accessibles.

SYSTAT: la place faite au codage Multidimensionnel est limitée au cas d'un tableau de données de proximités, traité suivant l'approche programmée dans M-D-SCAL. On peut cependant rentrer sa propre configuration de départ. Les sources ne sont pas accessibles.

MDS-X: est le résultat d'un travail de ré-écriture des programmes des laboratoires BELL (KYST exclu), effectuée sous la direction de Coxon à l'université d'Edimbourg. Les entrées sont désormais unifiées et de nombreux programmes utilitaires viennent compléter les outils d'interprétation des résultats (régression multiple pour projeter des variables externes sur la configuration obtenue, classification hiérarchique, etc. ...); Le tout en Fortran non dépourvu de bugs ! et peu structuré.

SAS: On trouve comme add-on à cet environnement une bibliothèque de procédures très proche de celle qui est disponible dans SPSS, puisqu'écrite aussi par F. W. Young et alii, dans la logique MINRESS, et donc fondée sur ALSCAL. Le produit ne semble plus suivi par SAS.

Références

- [1] Arabie, P. (1973) Concerning Monte Carlo Evaluations of Nonmetric Multidimensional Scaling. *Psychometrika* 38: 607-608.
- [2] d'Aubigny, G. (1975). Description Statistique des Données Ordinales: Analyse Multidimensionnelle. Thèse de 3ème cycle. Université de Grenoble I.
- [3] d'Aubigny, G. (1979). Analyse factorielle et visualisation des préférences. In Batteau, P., Jacquet-Lagrange, E., Monjardet, B. (eds.): *Analyse et agrégation des préférences dans les Sciences Sociales, économiques et de Gestion*:177-212. Economica, Paris.
- [4] d'Aubigny, G. (1989). L'Analyse Multidimensionnelle des données de dissimilarités. Thèse d'état. Université de Grenoble I.
- [5] d'Aubigny, G. (1993). A probabilistic modeling of the multidimensional scaling method and its relationship to shape analysis. In Bock, H.H., Lenski, W., Richter,

- M.M. (eds.): Information systems and data analysis. Studies in Classification, Data Analysis and Knowledge Organization, Vol. 4. Springer-Verlag, Heidelberg, 1993.
- [6] Barlow, R.E., D.J. Bartholomew, J.M. Bremer and H.D. Brunk (1972). Statistical Inference under Order Restrictions: The Theory of Isotonic Regression. John Wiley, New-York.
- [7] Barthélemy, J.-P. et A. Guénoche (1988) Les Arbres et les Représentations des Proximités. Masson. Paris.
- [8] Benzécri, J.-P. (1964). Sur l'Analyse Factorielle des Proximités. Publication de l'ISUP I, Vol 13.
- [9] Benzécri, J.-P. (1965). Sur l'Analyse Factorielle des Proximités. Publication de l'ISUP I, Vol 14.
- [10] Benzécri, J.-P. and al. (1973). L'Analyse des Données. Dunod. Paris.
- [11] Borg, I (eds.) (1981 b). Multidimensional data representations: When and Why? Mathesis Press.
- [12] Boughazi, M.A. (1987). Contribution à l'Etude des Algorithmes d'Optimisation en Analyse des Données. Thèse de 3ème cycle. Université de Grenoble I.
- [13] Bouroche, J.-M. (1977). Analyse des données en Marketing. Masson. Paris.
- [14] 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.
- [15] Carroll, J.D. (1983). Common Dimension Analysis. Paper presented at the Multidimensional Data Analysis Workshop. Paris.
- [16] Carroll, J.D. and J.J. Chang (1970) Analysis of Individual Differences in Multidimensional Scaling via a N-way Generalization of 'Eckart-Young' Decomposition. Psychometrika 35: 283-319.
- [17] Celeux, G., Diebolt, J. (1986). L'algorithme SEM: Un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités. Rev. stat. Appli. Vol 34 n°2.
- [18] Celeux, G., Diebolt, J. (1987). The EM and the SEM algorithms for mixture: statistical and numerical aspects. Proceedings of the 7th Franco-Belge meeting of Statistics. p.u.st Louis. Bruxelles and Rapport de recherche INRIA n°641.
- [19] Chandon J.L. et S. Pinson (1981). Analyse typologique, théorie et applications. Masson. Paris.
- [20] Chine, A. (1991). Algorithmes robustes en optimisation non convexe. Codes et simulations numériques en grandes dimensions. Thèse, Univ. J.F., Grenoble I.
- [21] Clark, A.K. (1976). Re-Evaluation of Monte Carlo Studies in Nonmetric Multidimensional Scaling. Psychometrika 41: 401-404.
- [22] Davy de Virville, M. (1972). l'Analyse des Proximités. Thèse de 3-ième cycle. Paris.

- [23] Ekman, G. (1954). Dimensions of Color Vision. *J. of Psychology*, 38: 467-474.
- [24] El Hadri, K. (1988) Lissage Monotone and Optimisation non Différentiable pour la résolution d'un problème de Codage Multidimensionnel. Thèse Université de Grenoble II.
- [25] Escoufier, Y. (1975) Le Positionnement Multidimensionnel. *Rev. de Statist. Appli.*, Vol XXIII, No 4: 5-14.
- [26] Fichet, B. (1983) Analyse Factorielle sur Tableaux de dissimilarités: Application aux Données sur Signe de Présence-Absence en Médecine. Thèse d'état, Univ. Marseille.
- [27] Gower, J.C. (1966). Some Distance Properties of Latent Roots and Vectors Used in Multivariate Analysis. *Biometrika* 53: 325-338.
- [28] 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.
- [29] Guttman, L. (1968). A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points. *Psychometrika* 33: 469-504.
- [30] 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.
- [31] Heiser, W.J. and de Leeuw, J. (1979). HOW to use SMACOF-I. *Departement of Data Theory*. University of Leiden.
- [32] Heiser, W.J., Stoop, I. (1986). Explicit Smacof Algorithms for Individual Differences Scaling. *Research Report RR-86-14*.
- [33] Isaac, P.D. and Poor, D.D.S. (1974). On the Determination of Appropriate Dimensionality in Data with Error. *Psychometrika*, 39: 91-109.
- [34] Klahr, D.A. (1969) Monte Carlo Investigation of the Statistical Significance of Kruskal's Scaling Procedure. *Psychometrika* 34: 319-330.
- [35] 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.
- [36] Kruskal, J.B. (1964 a) Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika* 29: 1-27.
- [37] Kruskal, J.B. (1964 b) Nonmetric Multidimensional Scaling: a Numerical Method. *Psychometrika* 29: 115-129.
- [38] Kruskal, J.B. and F. Carmone (1966). How to use M-D-SCAL (Version 5) and other useful informations. *Murray Hill, NJ: Unpublished manuscript, Bell Laboratories*.
- [39] Kruskal, J.B. and M. Wish. (1978) *Multidimensional Scaling* SAGE publ Beverly Hills.
- [40] Kruskal, J.B., F.W. Young and 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.*

- [41] Le Calvé, G. (1985). Distances à Centre. *Statistique and Analyse des Données*, Vol 10, No 2: 29-44.
- [42] de Leeuw, J. (1977). Applications of Convex Analysis to Multidimensional Scaling. *Recent Developments in Statistics*. J.R. BARRA and al. (eds.), North Holland. Amsterdam: 133-146.
- [43] de Leeuw, J. (1984). Differentiability of Kruskal's Stress at a local minimum. *Psychometrika* 49: 111-113
- [44] de Leeuw, J. (1988). Convergence of the majorization method for multidimensionnel scaling. *Journal of classification* 5 163-180.
- [45] de Leeuw, J. and I. Stoop. (1984). Upper Bounds for Kruskal's STRESS. *Psychometrika*, 439: 391-402.
- [46] de Leeuw, J. and W. Heiser. (1977). Convergence of Correction Matrix Algorithm for Multidimensional Scaling with Restrictions on the Configuration. In J. Lingoes(eds.), *Geometric Representation of Relational Data*. Mathesis Press. Ann Arbor: pp 735-752.
- [47] Levine, D.M. (1978). A Monte Carlo Study of Kruskal's Variance Based Measure on Stress. *Psychometrika*, 43: 307-315.
- [48] Lingoes, J.C. (1973). *The Guttman-Lingoes Nonmetric Program Series*. Mathesis Press, Ann Arbor.
- [49] Lingoes, J.C. and E.E. Roskam. (1973). A Mathematical and Empirical Analysis of two Multidimensional Scaling Algorithms. *Psychometrika Monograph Supplement*, 38: 1-93.
- [50] Lingoes, J.C., Roskam, E.E. and I. Borg (eds) (1979). *Geometric representations of relational data*. (2-ième édition) Mathesis Press, Ann Arbor
- [51] 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.
- [52] Machmouchi, M. (1992). *Contributions à la mise en œuvre de méthodes d'Analyse des données de dissimilarités*. Thèse, Univ. P.M.F., Grenoble II.
- [53] Mailles, J.P. (1979). *Analyse Factorielle des Tableaux de dissimilarités*. Thèse de 3ème cycle. Univ. Pierre and Marie Curie, Paris.
- [54] Mardia, K.V. (1978b). Some Properties of Classical Multidimensional Scaling. *Commun. Statist. A. Theory and Methods*, A7: 1233-1241.
- [55] Mardia, K.V., Kent, J.T. and J.M. Bibby. (1979). *Multivariate Analysis*. Academic Press. New-York.
- [56] Mathar, R. (1985). The Best Euclidean Fit to a given Distance Matrix in Prescribed Dimensions. *Linear Algebra and its Applications*, 67: 1-6.
- [57] McGee, V.E. (1968). Multidimensional Scaling of N Sets of Similarity Measures: A Nonmetric Individual Differences Approach. *Multivariate Behavioral Research*. 3: 233-248.

- [58] McKay, D.B. and J.L. Zinnes. (1981). Probabilistic Scaling of Spatial Judgments. *Geographical Analysis*, Vol 19, No. 1: 21-37.
- [59] Ramsay, J.O. (1977). Maximum Likelihood Estimation in Multidimensional Scaling. *Psychometrika* 42: 241-266.
- [60] Ramsay, J.O. (1978). MULTISCALE: Four programs for multidimensional scaling by the method of maximum likelihood. International Education Service, Chicago.
- [61] Ramsay, J.O. (1982). Some Statistical Approaches to Multidimensional Scaling data. *J. Roy. Statist. Soc. A* 145, part 3: 285-311.
- [62] 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.
- [63] Saito, T. (1985). Multidimensional Scaling to Represent Context Effect in Dissimilarity Judgment. *Hokkaido Behavioral Science Report, Series M, No. 7*. Sapporo.
- [64] Saito, T. (1986). Multidimensional Scaling to Explore Complex Aspects in Dissimilarity Judgment. *Behaviormetrika*, No. 20: 35-62.
- [65] Schiffman, S.S., M.L. Reynolds and F.W. Young (1981). *Introduction to Multidimensional Scaling: Theory, Methods and Applications*. Academic Press. New York.
- [66] Schoenberg, I.J. (1937). On certain Metric Spaces arising from Euclidean Spaces by a Change of Metric and their Imbedding in Hilbert Space. *Annals of Mathematics*, Vol 38, No. 4: 787-793.
- [67] Shepard, R.N. (1962 a). The Analysis of Proximities. Multidimensional Scaling with Unknown Distance Function. *Psychometrika* 27: 125-140.
- [68] Shepard, R.N. (1962 b). The Analysis of Proximities. Multidimensional Scaling with Unknown Distance Function. *Psychometrika* 27: 219-246.
- [69] Sherman, C.R. (1972). Nonmetric Multidimensional Scaling: A Monte Carlo Study of the Basic Parameters. *Psychometrika* 37: 323-355.
- [70] Spence, I. (1972). A Monte Carlo Evaluation of Three Nonmetric Multidimensional Scaling Algorithms. *Psychometrika*, 37: 461-486.
- [71] Spence, I. (1974). On Random Rankings Studies in Nonmetric Scaling. *Psychometrika*, 39: 267-268.
- [72] Spence, I. and J. Graef. (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.
- [73] Spence, I. and J. Graef. (1974). The Determination of the Underlying Dimensionality of an Empirically Obtained Matrix of Proximities. *Multivariate Behavioral research*, 9: 331-342.
- [74] Spence, I. and J.C. Ogilvie. (1973). A Table of Expected Stress Values for Random Rankings in Nonmetric Multidimensional Scaling. *Multivariate Behavioral research*, 8: 511-517.

- [75] Spence, I and F.W. Young (1978). Monte Carlo Studies in Nonmetric Scaling. *Psychometrika*, 43: 115-117.
- [76] Stenson, H.H and R.L. Knoll. (1969) Goodness of Fit for Random Rankings in Kruskal's Nonmetric Scaling Procedure. *Psychol Bulletin*, 72: 122-126
- [77] 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.*
- [78] Takane, Y. (1981) Statistical Procedures for Non Metric Multidimensional Scaling *J. of Ecology* 69: 135-152.
- [79] Takane, Y., F.W. Young and J. de Leeuw (1977). Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika* 42: 7-68.
- [80] Torgerson, W.S. (1958). *Theory and Methods of Scaling*. John Wiley New York.
- [81] Tournois, J. and P. Dickes (1993). *Pratique de l'échelonnement multidimensionnel: De l'observation à l'interprétation*. De Boeck-Wesmael, s.a., Bruxelles.
- [82] Wagenaar, W.A. and P. Padmos. (1971). Quantitative Interpretation of Stress in Kruskal's Multidimensional Scaling Technique. *British J. of Math. and Statist. Psychology*, 24: 101-110.
- [83] Winsberg, S. and J.D. Carroll. (1984). A Nonmetric Method for a Multidimensional Scaling Model Postulating Common and Specific Dimensions. Paper presented at the Psychometric Society Meeting.
- [84] Winsberg, S. and J.O. Ramsay (1983) Monotone Spline Transformations for Dimension Reduction. *Psychometrika*, 48: 575-595.
- [85] Yassine, A. (1989). *Etudes Adaptatives and Comparatives de certaines algorithmes en optimisation. Implémentations effectives and applications*. Thèse. Université de Grenoble I.
- [86] Young, F.W. (1968). TORSCA-9: A Fortran IV program for nonmetric Multidimensional Scaling. *Behavioral Science*, 13: 343-344.
- [87] 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.
- [88] Young, F.W. and J.R. Lewycky. (1979). *ALSCAL 4 User's guide*. (2nd edition). Data Analysis and Theory Associates, Chapel Hill.
- [89] Young, F.W. and M.S. Torgerson. (1967). TORSCA, A Fortran IV Program for the Shepard-Kruskal Multidimensional Scaling Analysis. *Behavioral Science*, 12: 498.
- [90] Zinnes, J.L. and D.B. McKay (1981). Multidimensional Scaling Models: The Other Side. In I. Borg (eds), *Multidimensional Data Representations: When and Why*. Mathesis Press. Ann Arbor.