

Université de Montréal

Biplot linéaire général et biplot non linéaire

par

Mohamed Nassim

Département de mathématiques et de statistique

Faculté des arts et sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en Discipline

mai 1998

© Mohamed Nassim, Avril 1998



E.5555-1111

QA
3
U54
1999
V.013

Université de Montréal

Diplôme de maîtrise en génie civil

Mohamed Nassef

Département de génie civil
Université de Montréal

Le présent rapport a été soumis en vue de l'obtention du grade de
maîtrise en génie civil de l'Université de Montréal.



Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Biplot linéaire général et biplot non linéaire

présenté par

Mohamed Nassim

a été évalué par un jury composé des personnes suivantes :

Yve Lepage

(président-rapporteur)

David Sankoff

(directeur de recherche)

Pascale Rousseau

(co-directeur)

André Plante

(membre du jury)

Mémoire accepté le :

03 Mai 1998

REMERCIEMENTS

Je remercie tout d'abord mon directeur de recherche, Monsieur David Sankoff, qui m'a fait confiance et m'a accueilli parmi ses étudiants. Je remercie sincèrement ma co-directrice de recherche, Madame Pascale Rousseau qui m'a dirigé pendant toute la durée de ce projet, et dont la disponibilité était exceptionnelle. Je remercie également l'Université de Montréal pour la bourse d'exemption pour les frais majorés qu'elle m'a accordée. La présentation et la mise en page de ce mémoire n'aurait pu être menées à terme sans l'aide de Madame Gisèle Legault et Monsieur René Ferland, lesquels je remercie chaleureusement. Je tiens à remercier mes amis Chafik Bouhaddioui, Amal Tanouti, Mohamed Abouhouraira et Hassan Youness qui m'ont soutenu pendant toute la durée de ma maîtrise, et qui m'ont aidé à surmonter plusieurs obstacles. Je tiens aussi à remercier une bonne amie, Aslin Parunak, pour son soutien et ses encouragements pendant toute la durée de ma maîtrise. L'occasion m'étant donnée, je voudrais exprimer toute ma gratitude à mes parents ainsi qu'à tous les membres de ma famille.

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vi
LISTE DES FIGURES	vii
RÉSUMÉ	ix
INTRODUCTION	1
CHAPITRE I	
BIPLOT CLASSIQUE : THÉORIE ET APPLICATION	4
1.1 Introduction	4
1.2 Espace des individus	6
1.2.1 Premier système d'axes dans R_p	6
1.2.2 Deuxième système d'axes dans R_p	7
1.3 Espace dual	11
1.3.1 Premier système d'axes pour R_n	11
1.3.2 Deuxième système d'axes pour R_n	12
1.4 Biplot classique	14
1.4.1 Propriétés du biplot	16
1.4.2 Interpolation et prédiction	17
1.4.3 Utilité du biplot dans R_2	20
1.4.4 Biplot et analyse des correspondances	20
1.5 Analyse de données	22
1.5.1 Ensemble de données	22
1.5.2 Biplot des 172 fondateurs et des 6 variables confondantes	23
1.5.3 Biplot des 172 fondateurs et des 6 variables de parenté	26
1.5.4 Espace dual	28
CHAPITRE II	
RÉSULTATS GÉNÉRAUX SUR L'ALGÈBRE ET LA GÉOMÉTRIE DU BILOT LINÉAIRE	35
2.1 Introduction	35

2.2	Préliminaire	37
2.2.1	Système cartésien	38
2.2.2	Axes obliques	40
2.3	Résultats généraux	41
2.4	Axes biplot	49
2.5	Conclusion	52
CHAPITRE III		
	POSITIONNEMENT MULTIDIMENSIONNEL	53
3.1	Introduction	53
3.2	Distances non pythagorienne	54
3.3	Solution géométrique	57
3.4	Solution algébrique: Cas où δ_{rs} est pythagorienne	60
3.4.1	Introduction	60
3.4.2	Solution algébrique	61
3.5	Solution algébrique: cas général	64
3.6	Non unicité de la solution	67
3.6.1	Approximation à la représentation	69
3.6.2	Matrice définie positive	70
3.7	Algorithme	71
CHAPITRE IV		
	BILOT NON LINÉAIRE	72
4.1	Introduction	72
4.2	Préliminaire	74
4.2.1	Notation	74
4.2.2	Hypothèses de base	74
4.2.3	Interpolation algébrique d'un point dans \mathcal{L}_m	76
4.3	Axes non linéaires dans \mathcal{L}_{n-1}	79
4.3.1	Trajectoires dans \mathcal{L}_{n-1}	80
4.3.2	Prédiction conventionnelle	83
4.3.3	Interpolation conventionnelle	83

4.4	Axes biplot dans \mathcal{L}_m	84
4.4.1	Axes biplot d'interpolation	84
4.4.2	Axes biplot de prédiction	86
	CONCLUSION	91
	RÉFÉRENCES	92

LISTE DES TABLEAUX

1.1	Résultats de l'ACP sur les variables confondantes	24
1.2	Norme de la projection pour chaque variable confondante	25
1.3	Résultats de l'ACP sur les variables de parenté	27
1.4	Norme de la projection pour chaque variable de parenté	28
1.5	Moyennes et écarts types des variables confondantes et des variables de parenté	29

LISTE DES FIGURES

1.1	Interpolation du point $P(2, -3, 4)$, tiré de Gower (1996), page 14.	18
1.2	Relation entre les marqueurs de prédiction P et ceux d'interpolation I	19
1.3	Axes biplot : ANN , $NBGEN$, $IDESC$, $IUTIL$, $IREP$ et $NBFRAT$	30
1.4	Axes biplot : N_s , \bar{X}_s , S_s , N_n , \bar{X}_n , et S_n	31
1.5	Axes biplot : Variables confondantes (4) et de parenté (6)	32
1.6	Espace dual : Variables confondantes centrées et non réduites	33
1.7	Espace dual : Variables confondantes centrées réduites	34
2.1	Interpolation du point P_1 de coordonnées (2,3) et prédiction des coordonnées (3,4) du point P_2	39
2.2	Interpolation du point $P(3, 4)$ par rapport à des axes obliques	40
2.3	Représentation de la back-projection \mathbf{z}	42
2.4	Les marqueurs de la droite β sont les back-projections des marqueurs de l'axe cartésien e_k	50
2.5	Géométrie de la back-projection	51
3.1	Représentation géométrique de trois valeurs d'un coefficient de similarité	58
4.1	Représentation géométrique des données dans deux espaces distincts	73

4.2 Interpolation d'un point $P(4, 3, 1)$ dans le cas non linéaire. « <i>tiré de Gower et Hand 1996, page 109</i> »	87
4.3 Utilisation de la trajectoire non linéaire β comme axe biplot. « <i>tiré de Gower 1996, page 111</i> »	89

RÉSUMÉ

Un ensemble de données constitué de p observations sur chacun des n individus peut être représenté dans un espace de dimension p ou de dimension n . L'espace qui représente les individus par des points est un espace à p dimensions; l'espace qui représente les variables par des points est un espace à n dimensions. Afin de visualiser et d'interpréter la configuration des points (*individus*) dans l'espace à p dimensions, on la représente approximativement dans un espace de dimension $d < p$, d est généralement égal à 2. L'information donnée par les axes originaux peut être retrouvée géométriquement à l'aide d'axes biplot qui représentent les variables originales. L'interpolation d'un nouveau point et la prédiction des valeurs des variables à partir d'un nouveau point peut aussi être fait géométriquement à partir des axes biplot. L'interprétation du nuage de points dans R_2 est ainsi facilitée par les axes biplot. Le traçage et l'utilisation de ces axes biplot dans ces sous-espaces de dimension d sont étudiés dans ce mémoire. Remarquons que ces axes biplot peuvent être linéaires ou non linéaires selon le cas. La visualisation et l'interprétation des points (*variables*) est faite dans un espace de dimension n .

INTRODUCTION

Dans un contexte d'analyse multivariée, nous observons p mesures sur chacun des n individus. Classiquement, il y a deux façons de représenter ces observations. Chaque individu peut être représenté par un point dans un espace euclidien de dimension p noté R_p : espace dans lequel les axes cartésiens représentent les variables. Cet espace est nommé espace des individus. La deuxième façon est de représenter chaque variable par un point dans un espace euclidien de dimension n noté R_n : les axes représentent les individus. Il paraît bizarre de représenter les individus par des axes, mais comme on le verra, cet espace possède des propriétés intéressantes. Remarquons que les points de R_p sont des axes dans R_n et les axes de R_p sont des points dans R_n .

Ce mémoire porte sur la représentation simultanée des individus et des variables dans un même espace R_p ou dans un sous-espace R_d de R_p ($d < p$). L'intérêt de cette représentation simultanée peut être très bien expliqué à l'aide du cas spécial où le nombre de mesures p sur chaque individu est égal 2 ou 3. En effet, lorsque nous avons deux variables mesurées sur chacun des individus, il est facile de faire un graphique dans l'espace des individus R_2 . On peut alors visualiser les tendances linéaires ou autres des deux variables, y repérer des regroupements d'individus (clusters) ou des valeurs aberrantes. En un mot, on peut donner un sens à la configuration de points obtenus. On peut aussi interpoler un nouveau point pour chaque individu qui vient se rajouter à l'ensemble de données original. On peut retrouver les valeurs des variables qu'aurait un point quelconque dans ce plan.

Ces mêmes idées restent pertinentes lorsqu'il y a p mesures observées sur chaque individu. L'analyse en composantes principales (ACP) permet d'approximer le nuage de points dans un sous-espace $R_d, d < p$. Mais il est difficile de donner un sens à la configuration obtenue car les axes représentent des transformations linéaires des va-

riables originales; il n'est pas facile d'interpoler de nouvelles valeurs; il est encore plus difficile d'y prédire des valeurs à partir d'un point de R_d . On développera au chapitre 1 la méthodologie pour cette représentation simultanée des individus par des points et des variables par des droites dans un sous-espace R_d . Ces droites représentant les variables seront appelées axes biplot. C'est avec ces axes biplot qu'il sera possible de retrouver des valeurs approximatives aux valeurs originales des variables pour chaque individu. Cette méthodologie s'appelle le biplot classique. On pourra aussi interpoler et prédire géométriquement avec ces axes biplot. Il sera ainsi aisé de donner un sens à la configuration du nuage de points.

Le nuage de points peut aussi être représenté dans des sous-espaces autres que celui donné par l'analyse en composantes principales. La recherche de tels sous-espaces $R_d \subset R_p$ est faite par exemple, avec l'analyse de la « recherche d'une projection » (projection pursuit). Le sous-espace R_d est trouvé en optimisant d'autres critères que celui de l'ACP. Il devient donc intéressant de généraliser la technique classique du biplot à un sous-espace quelconque. On y développera cette généralisation au chapitre 2. Les axes biplot trouvés sont encore des axes linéaires.

Ce nuage de points peut aussi être représenté dans un espace \mathcal{L} quelconque, $\mathcal{L} \not\subset R_p$. Nous élaborerons davantage sur cette représentation des individus au chapitre 3: il y sera question de la méthodologie du positionnement multidimensionnel (multidimensional scaling). La motivation de cette technique réside dans le fait que la métrique utilisée pour comparer deux individus peut être autre que pythagorienne. L'ACP et la recherche d'une projection ne peuvent plus être utilisées car celles-ci utilisent une distance pythagorienne.

Le problème de la représentation simultanée des variables et des individus dans ces espaces \mathcal{L} demeure toujours important. Au chapitre 4, nous développerons la méthodologie pour tracer des axes biplot dans de tels sous-espaces, pour interpoler et prédire à l'aide de ces axes biplot. Remarquons que ces axes biplot sont en général non linéaires.

L'outil mathématique le plus souvent employé est la décomposition en valeurs singulières d'une matrice de dimensions $n \times p$ ainsi qu'une approximation à cette matrice qui est la meilleure au sens des moindres carrés. Cette approximation est due à Eckart-Young-Householder (1936). Un exemple en épidémiologie sera étudié en utilisant les axes biplot linéaires.

Notre contribution originale a été de présenter d'une façon unifiée la démarche mathématique du biplot linéaire classique et généralisé, de la méthode du positionnement multidimensionnel métrique. Tous les détails mathématiques ont été faits. Ces techniques étaient publiées dans de nombreux articles. L'analyse sur les données en épidémiologie était une analyse originale dont le but était d'apporter des éclaircissements sur les liens entre les fondateurs-ancêtres sur l'île de Saint Barthélemy et des individus atteints de surdit  qui vivent actuellement sur cet  le.

CHAPITRE I

BIPLOT CLASSIQUE : THÉORIE ET APPLICATION

1.1 Introduction

Le but de ce chapitre est d'étudier une représentation simultanée et approximative des individus par des points, des variables par des droites dans un même sous-espace R_d . Cette technique de représentation simultanée s'appelle la technique du « biplot ». Ce chapitre portera sur la technique du biplot classique ainsi que ses propriétés tel que développées par Rubin Gabriel (1971, 1995) et John Gower (1992, 1993).

Tout d'abord, nous définirons l'espace des individus R_p et l'espace des variables R_n comme espaces de représentation de l'ensemble de données constitués de p observations faites sur chacun des n individus. L'espace R_p est celui dans lequel chaque individu est représenté par un point : cet espace est nommé espace des individus. Le second espace R_n est celui dans lequel chaque variable est représentée par un point : cet espace est nommé l'espace des variables ou espace dual. A la section 1.2, deux systèmes d'axes pour chacun de ces sous-espaces sont proposés. Le premier système d'axes est cartésien; le second système offre d'excellentes propriétés pour fins d'approximation dans R_d , $d < p$.

A la section 1.4, nous verrons une représentation simultanée des variables et des individus dans l'espace des individus R_p : les variables sont représentées par des droites dans R_p et les individus par des points dans R_p . Cette double représentation nous amène à une représentation simultanée et approximative dans R_d , $d < p$: celle-ci se nomme technique du biplot. Evidemment si la dimension p du sous-espace est 2 alors une

représentation graphique simultanée des individus et des variables peut être faite d'une façon exacte dans le plan. La représentation simultanée des variables et des individus peut aussi être faite dans le sous-espace des variables. Les techniques étant semblables pour R_p et R_n , nous nous concentrerons sur la représentation simultanée dans l'espace des individus : celle-ci offre d'ailleurs les propriétés les plus intéressantes. Remarquons que les propriétés de ces représentations simultanées dans l'espace des individus et dans l'espace des variables sont duales.

Ensuite nous établirons des propriétés des biplots et feront le lien avec l'analyse en composantes principales, l'ACP. Louis-Paul Rivest et Gaétan Daigle (1992) ont étudié la robustesse du biplot classique. Nous ne discuterons pas de cet aspect.

La section 1.5 portera sur une analyse de données provenant du domaine de l'épidémiologie : cet exemple illustrera dans R_2 les différents concepts étudiés.

Notation

La matrice contenant les p observations faites sur chacun des n individus est notée $\mathbf{Y}_{n \times p}$; l'élément y_{ij} représente l'observation faite sur la j -ième variable du i -ième individu. Nous ne discuterons que des cas où les p variables sont continues.

Le vecteur colonne \mathbf{y}_i contient les p observations faites sur le i -ième individu : \mathbf{y}'_i est donc la i -ième ligne de la matrice $\mathbf{Y}_{n \times p}$. Le vecteur colonne $\mathbf{y}_{(j)}$ contient les n observations faites sur la j -ième variable : $\mathbf{y}_{(j)}$ est donc la j -ième colonne de la matrice $\mathbf{Y}_{n \times p}$.

Nous supposerons, sans perte de généralité que les n observations faites sur la j -ième variable sont centrées, c'est-à-dire

$$\sum_{i=1}^n y_{ij} = 0, \quad \text{pour tout } j = 1, \dots, p.$$

Le vecteur unité noté \mathbf{e}_j a tous ses éléments nuls excepté le j -ième élément qui est égal à 1. Le vecteur $\mathbf{1}$ a tous ses éléments égaux à 1. La matrice identité est notée \mathbf{I} . Le contexte rend évident la dimension du vecteur \mathbf{e}_j , du vecteur $\mathbf{1}$ et de la matrice \mathbf{I} .

1.2 Espace des individus

1.2.1 Premier système d'axes dans R_p

Les deux premières sections développent des idées de base pour l'analyse multivariée. Nous avons trouvé intéressant d'établir un lien entre celles-ci et les récents développements du biplot. La décomposition en valeurs singulières est l'outil qui permet de faire ce lien.

Axes cartésiens

La façon la plus immédiate est de représenter les p observations faites sur un individu par des points dans un espace cartésien R_p : la direction des axes est alors $\mathbf{e}_{j_{(p \times 1)}}$. Chaque axe représente une variable; en effet les valeurs $\rho \mathbf{e}_j$, $-\infty < \rho < +\infty$, représentent toutes les valeurs que peut prendre la j -ième variable. Remarquons que ce choix d'axes orthogonaux est arbitraire et n'a aucune signification statistique.

Le i -ième individu est représenté par un point P_i dont les p coordonnées sont les p composantes du vecteur \mathbf{y}_i .

Représentation graphique

Si le nombre de variables p est 2 alors la représentation graphique des individus se fait dans le plan cartésien. Celle-ci est une nécessité élémentaire : elle nous permet de juger les tendances linéaires, quadratiques ou autres, ou tout simplement de visualiser des regroupements d'individus, des valeurs aberrantes, etc. De plus, une modélisation adéquate de la dépendance entre les variables est plus rapidement faite à l'aide d'un graphique.

La représentation est toujours possible si le nombre p de variable est 3 mais elle est moins aisée. Si p est plus grand que 3 alors la représentation graphique des n individus est impossible. La visualisation graphique devra se faire avec une approximation du

nuage de points dans R_2 ou R_3 . On se contentera d'une approximation aux positions relatives des point-individus P_i dans R_p .

Dans le but de trouver, dans R_d où $d < p$, une approximation au nuage de points qui est dans R_p , nous allons discuter d'un deuxième système d'axes.

1.2.2 Deuxième système d'axes dans R_p

Décomposition en valeurs singulières

Un autre choix d'axes peut être fait en utilisant la décomposition de la matrice \mathbf{Y} en valeurs singulières :

$$\begin{aligned} \mathbf{Y}_{(n \times p)} &= \mathbf{U}_{(n \times n)} \mathbf{\Gamma}_{(n \times p)} \mathbf{V}'_{(p \times p)} \\ &= (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) \begin{pmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_p \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v}'_1 \\ \mathbf{v}'_2 \\ \vdots \\ \mathbf{v}'_p \end{pmatrix} \end{aligned}$$

où \mathbf{U} et \mathbf{V} sont des matrices orthogonales et la diagonale de la matrice $\mathbf{\Gamma}$ contient les valeurs singulières γ_i , $i = 1, \dots, p$ triées selon un ordre décroissant; remarquons que les éléments hors diagonale de $\mathbf{\Gamma}$ sont nuls.

Les vecteurs colonnes $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ de la matrice \mathbf{V} peuvent servir de base pour représenter les lignes de la matrice \mathbf{Y} . Les coordonnées du vecteur \mathbf{y}_i dans la base $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p)$ sont $(\mathbf{y}'_i \mathbf{v}_1, \mathbf{y}'_i \mathbf{v}_2, \dots, \mathbf{y}'_i \mathbf{v}_p)$: celles-ci sont données par la i -ième ligne de $\mathbf{U}\mathbf{\Gamma}$. En effet on a

$$\mathbf{Y}\mathbf{V} = \mathbf{U}\mathbf{\Gamma}\mathbf{V}'\mathbf{V} = \mathbf{U}\mathbf{\Gamma}.$$

Sans perdre de généralité, on suppose que le rang de la matrice \mathbf{Y} est p . Le vecteur \mathbf{y}_i s'écrit dans la nouvelle base comme suit :

$$\mathbf{y}_i = a_{i1}^* \mathbf{v}_1 + a_{i2}^* \mathbf{v}_2 + \cdots + a_{ip}^* \mathbf{v}_p \quad (1.1)$$

où $a_{ij}^* = \gamma_j u_{ij} = \mathbf{y}'_i \mathbf{v}_j$.

Remarquons que la coordonnée a_{ij}^* de l'individu i sur l'axe dont la direction est \mathbf{v}_j est une combinaison linéaire des p mesures faites sur cet individu. Les poids $(v_{j1}, v_{j2}, \dots, v_{jp})$ affectés aux différentes mesures $(y_{i1}, y_{i2}, \dots, y_{ip})$ sont donnés par les coordonnées du vecteur $\mathbf{v}'_j = (v_{j1}, v_{j2}, \dots, v_{jp})$. Chaque axe de ce système cartésien peut donc être considéré comme une combinaison linéaire des variables originales. Le vecteur des coordonnées de l'individu i dans ce nouveau système d'axes est noté $\mathbf{a}_i^* = (a_{i1}^*, a_{i2}^*, \dots, a_{ip}^*)$.

Représentation graphique

La représentation graphique exacte dans un espace de dimension d égale à 2 peut être faite uniquement si le rang de la matrice \mathbf{Y} est 2. Dans le paragraphe suivant nous discuterons d'approximation $\mathbf{y}_{i[d]}$ à \mathbf{y}_i dans un sous-espace R_d , $d < p$. Nous établirons certaines propriétés pour cette approximation.

Représentation approximative dans un sous-espace R_d , $d < p$

La projection orthogonale \hat{P}_i du point P_i sur les d premiers vecteurs $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ associés aux valeurs $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_d$ est :

$$\mathbf{y}_{i[d]} = a_{i1}^* \mathbf{v}_1 + a_{i2}^* \mathbf{v}_2 + \dots + a_{id}^* \mathbf{v}_d$$

c'est-à-dire, les coordonnées du point \hat{P}_i sont $(a_{i1}^*, a_{i2}^*, \dots, a_{id}^*)$. Cette représentation approximative du nuage de points P_i , $i = 1, \dots, n$ par le nuage de points \hat{P}_i , $i = 1, \dots, n$ a de nombreuses propriétés. Avant de discuter celles-ci, nous allons énoncer le théorème d'Eckart-Young 1936. Ce théorème fournit une propriété intéressante du point de vue statistique et algébrique.

Théorème d'Eckart-Young

Théorème 1.2.1 Soit $\mathbf{Y}_{n \times p}$ une matrice quelconque dont la décomposition en valeurs singulières est $\mathbf{Y} = \mathbf{U}\mathbf{F}\mathbf{V}'$, où les vecteurs colonnes de \mathbf{V} et de \mathbf{U} sont triés selon l'ordre décroissant des valeurs γ_i ($i = 1, \dots, p$). Soit $\mathbf{L}_{p \times d}$ une matrice dont les vecteurs colonnes sont orthonormés. Alors $\|\mathbf{Y} - \mathbf{Y}\mathbf{L}\mathbf{L}'\|$ est minimale par rapport à tous les espaces de dimension d si $\mathbf{L} = \mathbf{V}_d$, où \mathbf{V}_d sont les d premières colonnes de \mathbf{V} . La norme matricielle utilisée est définie par : $\|\mathbf{Y}\| = \sum_{i=1}^n \sum_{j=1}^p y_{ij}^2$.

Corollaire 1.2.1 La matrice de rang d qui minimise $\|\mathbf{Y} - \mathbf{Y}\mathbf{L}\mathbf{L}'\|$ est donnée par $\mathbf{Y}_{[d]} = \mathbf{U}\mathbf{\Gamma}_d\mathbf{V}'_d$, où $\mathbf{\Gamma}_d$ est la matrice diagonale dont les éléments de la diagonale sont $\gamma_1 \geq \dots \geq \gamma_d$.

En effet, d'après le théorème, la norme $\|\mathbf{Y} - \mathbf{Y}\mathbf{V}_d\mathbf{V}'_d\|$ est minimale; donc

$$\begin{aligned} \mathbf{Y}\mathbf{V}_d\mathbf{V}'_d &= \mathbf{U}\mathbf{F}\mathbf{V}'\mathbf{V}_d\mathbf{V}'_d \\ &= \mathbf{U}\mathbf{F}\mathbf{J}_d\mathbf{V}'_d \\ &= \mathbf{U}\mathbf{\Gamma}_d\mathbf{V}'_d \\ &= \mathbf{Y}_{[d]} \end{aligned}$$

où \mathbf{J}_d est la matrice formée des d premières colonnes de la matrice \mathbf{I} de dimension p . Les vecteurs lignes de la matrice $\mathbf{Y}_{[d]}$ sont notés $\mathbf{y}_{i[d]}$.

Remarque 1.2.1 Les résultats du théorème et du corollaire sont connus sous le nom de théorème d'Eckart-Young. Gower et d'autres statisticiens ont beaucoup utilisé ce théorème dans le but de lui donner un sens statistique. Notre prestation est selon cette idée. Plusieurs preuves de ce théorème ont été faites. Gower (1995) présente une de ces preuves; nous n'en donnerons pas dans ce mémoire.

Propriétés de l'approximation du nuage de points dans R_d

1. Distance euclidienne $\|\mathbf{y}_i - \mathbf{y}_j\|^2$

$$\|\mathbf{y}_{i_{[d]}} - \mathbf{y}_{j_{[d]}}\| \lesssim \|\mathbf{y}_i - \mathbf{y}_j\|$$

où \lesssim signifie que le terme de gauche approxime inférieurement le terme de droite : en effet $\mathbf{y}_{i_{[d]}}$ est la projection orthogonale de \mathbf{y}_i sur \mathbf{V}_d .

On dira qu'un individu est bien représenté, c'est-à-dire que l'information sur cet individu a été assez bien conservée si $\|\mathbf{y}_i\| \sim \|\mathbf{y}_{i_{[d]}}\|$ pour tout $i = 1, \dots, n$, où le symbole \sim signifie que le membre de droite est à peu près égale au membre de gauche.

2. Direction des composantes principales $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$

Les directions des vecteurs $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ sont les directions des composantes principales (ACP) exécutée en vue d'une représentation des individus. Rappelons que les directions des composantes principales sont celles qui offrent une variance échantillonnale maximale sur toutes les directions orthogonales possibles. Ces directions sont données par les vecteurs propres de $\mathbf{Y}'\mathbf{Y}$. En fait ces vecteurs propres sont $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$. En effet :

$$\begin{aligned} \mathbf{Y}'\mathbf{Y} &= \mathbf{V}\mathbf{\Gamma}\mathbf{U}'\mathbf{U}\mathbf{\Gamma}\mathbf{V} \\ &= \mathbf{V}\mathbf{\Gamma}^2\mathbf{V}' \end{aligned}$$

Ainsi la projection du nuage de points sur le premier axe \mathbf{v}_1 donne une représentation des points qui offrent une variance maximale parmi toutes les directions possibles; \mathbf{v}_2 est la deuxième direction qui possède la variance échantillonnale maximale parmi toutes les directions orthogonales à \mathbf{v}_1 , etc.

3. Meilleure approximation au sens des moindres carrés

$$\sum_{i=1}^n \|\mathbf{y}_i - \mathbf{y}_{i_{[d]}}\|^2 \text{ est minimale}$$

En effet : $\sum_{i=1}^n \|\mathbf{y}_i - \mathbf{y}_{i_{[d]}}\|^2 = \|\mathbf{Y} - \mathbf{Y}_{[d]}\|^2$ est minimale par rapport à tous les sous-espaces de R_p de dimension d en vertu du théorème d'Eckart-Young.

1.3 Espace dual

1.3.1 Premier système d'axes pour R_n

Axes cartésiens

La représentation des p variables par p points distincts dans l'espace R_n possède des propriétés intéressantes. Celle-ci est faite à l'aide des vecteurs $\mathbf{y}_{(j)}$, $j = 1, \dots, p$, qui contiennent chacun les n observations faites sur chaque variable; ce vecteur est de dimension $n \times 1$. La façon la plus immédiate de représenter ce vecteur $\mathbf{y}_{(j)}$ par un point dans l'espace R_n est de choisir n axes cartésiens de direction $\mathbf{e}_{i_{(n \times 1)}}$, $i = 1, \dots, n$. Les coordonnées du point qui représente la j -ième variable sont les composantes du vecteur $\mathbf{y}_{(j)}$. Il est d'usage de représenter géométriquement une variable j par un vecteur reliant le point $\mathbf{y}_{(j)}$ à l'origine $\mathbf{0}$. Nous considérerons donc $\mathbf{y}_{(j)}$ comme un vecteur de l'espace R_n .

L'axe cartésien de R_n de direction \mathbf{e}_i représente le i -ième individu; il semble bizarre qu'un axe représente un individu : les valeurs $\rho \mathbf{e}_i$, $-\infty < \rho < +\infty$, n'ont aucune signification. Toutefois la représentation des variables dans cet espace des variable R_n a des propriétés géométriques très intéressantes : la variance, covariance et corrélation des variables ont des interprétations géométriques dans R_n .

Propriétés géométriques

1. La variance échantillonnale s_j^2 est égale au carré de la longueur du vecteur $\mathbf{y}_{(j)}$ à un facteur n près :

$$ns_j^2 = \|\mathbf{y}_{(j)}\|^2 = \sum_{i=1}^n (y_{ij})^2.$$

Si les variables sont centrées et réduites alors la longueur du vecteur est égale à \sqrt{n} , et les p vecteurs sont situés sur une hypersphère de rayon \sqrt{n} située dans R_n .

2. Le produit scalaire $\mathbf{y}'_{(j)} \cdot \mathbf{y}_{(k)}$ représente la covariance ($\text{cov}(\mathbf{y}_{(j)}, \mathbf{y}_{(k)})$) entre les variables j et k à un facteur n près :

$$n(\text{cov}(\mathbf{y}_{(j)}, \mathbf{y}_{(k)})) = \sum_{i=1}^n y_{ij}y_{ik} = \mathbf{y}'_{(j)} \cdot \mathbf{y}_{(k)}.$$

3. Le cosinus de l'angle θ entre les vecteurs $\mathbf{y}_{(j)}$ et $\mathbf{y}_{(k)}$ représente géométriquement la corrélation entre ces deux variables. La corrélation r_{jk} s'écrit :

$$r_{jk} = \frac{\mathbf{y}'_{(j)} \cdot \mathbf{y}_{(k)}}{\|\mathbf{y}_{(j)}\| \cdot \|\mathbf{y}_{(k)}\|} = \cos \theta.$$

La corrélation est grande si et seulement si l'angle θ est petit. La corrélation $r_{jk} = 0$ si et seulement si les vecteurs $\mathbf{y}_{(j)}$, $\mathbf{y}_{(k)}$ sont orthogonaux, etc.

Représentation graphique

Etant donné qu'un sous-espace à deux dimensions suffit pour représenter deux vecteurs, on peut visualiser géométriquement ces propriétés statistiques pour tout couple de variables. Les p vecteurs $\mathbf{y}_{(j)} \in R_n$ étant contenus dans un sous-espace de dimension p : il est impossible de visualiser simultanément les propriétés statistiques des p variables, pour $p > 3$. Il faudra avoir recours à un sous-espace R_2 ou R_3 , qui contient p vecteurs qui approximent les p vecteurs $\mathbf{y}_{(j)}$.

Dans la sous-section suivante, nous allons discuter d'un deuxième choix d'axes dans R_n pour représenter ces p vecteurs $\mathbf{y}_{(j)}$. Ce nouveau choix d'axes nous amènera à une approximation $\mathbf{y}_{(j)[d]}$ de $\mathbf{y}_{(j)}$, $\mathbf{y}_{(j)[d]} \in R_d$, $d < p \leq n$.

1.3.2 Deuxième système d'axes pour R_n

Décomposition en valeurs singulières

La décomposition de \mathbf{Y} en valeurs singulières nous fournira encore une fois, les outils pour choisir ce deuxième système d'axes. La façon de procéder est analogue à celle utilisée pour l'espace des individus. Nous avons : $\mathbf{Y} = \mathbf{U}\mathbf{T}\mathbf{V}'$ où les vecteurs colonnes $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ de la matrice \mathbf{U} peuvent servir de base pour représenter les colonnes de la matrice \mathbf{Y} . Les coordonnées de $\mathbf{y}_{(j)}$ dans cette nouvelle base sont :

$$(\mathbf{y}'_{(j)} \cdot \mathbf{u}_1, \mathbf{y}'_{(j)} \cdot \mathbf{u}_2, \dots, \mathbf{y}'_{(j)} \cdot \mathbf{u}_n)'$$

Le j -ième vecteur colonne de $\mathbf{\Gamma V}'$ est le vecteur qui contient les n coordonnées de la variable j dans la nouvelle base. En effet :

$$\begin{aligned} (\mathbf{Y}'\mathbf{U})' &= (\mathbf{V}\mathbf{\Gamma}\mathbf{U}'\mathbf{U})' \\ &= \mathbf{\Gamma V}' \end{aligned}$$

$$\mathbf{y}_{(j)} = b_{1j}^* \mathbf{u}_1 + b_{2j}^* \mathbf{u}_2 + \dots + b_{nj}^* \mathbf{u}_n \quad (1.2)$$

J'obtiens $\mathbf{b}_{ij}^* = \gamma_i v_{ij} = \mathbf{y}'_{(j)} \cdot \mathbf{u}_i$.

Remarquons que la coordonnée \mathbf{b}_{ij}^* de la variable j sur l'axe dont la direction est \mathbf{u}_i est une combinaison linéaire des n mesures faites sur cette variable. Les poids $(u_{j1}, u_{j2}, \dots, u_{jp})$ attribués aux n mesures $(y_{1j}, y_{2j}, \dots, y_{nj})$ sont donnés par les coordonnées du vecteur $\mathbf{u}'_i = (u_{j1}, u_{j2}, \dots, u_{jp})$. Chaque axe du système cartésien est une combinaison linéaire qui pondère les mesures faites sur les n individus. Il n'y a aucun sens statistique associé à cette pondération. Les coordonnées de la variable j dans ce nouveau système d'axes sont notées $\mathbf{b}_j^* = (b_{j1}^*, b_{j2}^*, \dots, b_{jp}^*)'$

Approximation de la représentation des variables

La projection orthogonale du vecteur $\mathbf{y}_{(j)}$ sur les d premiers vecteurs $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$ associés aux valeurs $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_d$ est :

$$\mathbf{y}_{(j)[d]} = b_{1j}^* \mathbf{u}_1 + b_{2j}^* \mathbf{u}_2 + \dots + b_{dj}^* \mathbf{u}_d.$$

Le vecteur reliant le point $\mathbf{b}_j^* = (b_{1j}^*, b_{2j}^*, \dots, b_{id}^*)$ à l'origine représente la j -ième variable. Cette représentation approximative a de nombreuses propriétés.

Propriétés de l'approximation

1. Distance euclidienne

$$\|\mathbf{b}_i^* - \mathbf{b}_j^*\| \lesssim \|\mathbf{y}_{(i)} - \mathbf{y}_{(j)}\|$$

La préservation de la distance entre deux variables est intéressante mathématiquement mais n'a pas d'intérêt statistique

2. Une variable est bien représentée si

$$\sqrt{ns_j} = \|\mathbf{y}_{(j)}\| \approx \|\mathbf{b}_j^*\|.$$

Si les variables sont centrées et réduites alors $\|\mathbf{y}_{(j)_{[d]}}\| \approx \sqrt{n}$: les variables les mieux représentées sont celles dont les valeurs $\mathbf{y}_{(j)_{[d]}}$ sont le plus près de l'hypersphère de rayon \sqrt{n} .

3. Les directions des vecteurs $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$ sont les **directions des composantes principales** pour fins de représentation des variables. En effet, ces directions sont données par les vecteurs propres de $\mathbf{Y}\mathbf{Y}'$ et nous avons bien que ces vecteurs propres sont $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$ car $\mathbf{Y}\mathbf{Y}' = \mathbf{U}\mathbf{F}\mathbf{V}'\mathbf{V}\mathbf{F}' = \mathbf{U}\mathbf{\Gamma}^2\mathbf{U}$.
4. Par le théorème d'Eckart-Young, $\mathbf{y}_{(j)_{[d]}}$ est la **meilleure approximation** à $\mathbf{y}_{(j)}$ en ce sens que $\sum_{j=1}^p \|\mathbf{y}_{(j)} - \mathbf{y}_{(j)_{[d]}}\|^2$ est minimale.

Dans la prochaine section, on verra comment Gabriel (1971, 1995) et Gower et Hand (1996) ont représenté simultanément les individus et les variables.

1.4 Biplot classique

La représentation des n individus faite avec l'A.C.P dans le sous-espace R_d ne donne aucune information sur les variables originales car les axes du sous-espace R_d sont des combinaisons linéaires des variables. La projection orthogonale d'un point \hat{P}_i sur un de ces axes nous indique si la combinaison de ces variables est grande ou petite. Un des buts du biplot est de retrouver facilement l'information sur chacune des variables en représentant simultanément les n individus et les p variables dans un même espace de dimension $d < p$. Cette représentation simultanée peut être faite dans l'espace des individus ou dans l'espace des variables. Les propriétés de la représentation simultanée (biplot) dans l'espace des individus semblent plus intéressantes, c'est pourquoi nous discuterons seulement de celle-ci.

Les valeurs $\rho \mathbf{e}_{j_{(p \times 1)}}$, $-\infty < \rho < +\infty$, représentent le j -ième axe cartésien de R_p . Gower définit l'axe biplot β_j comme étant la projection des vecteurs $\rho \mathbf{e}_j$ sur l'espace

engendré par les vecteurs colonnes de \mathbf{V}_d , $-\infty < \rho < +\infty$:

$$\begin{aligned}
 \rho \mathbf{e}'_j \mathbf{V}_d &= \rho \mathbf{e}'_j (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d) \\
 &= \rho \mathbf{e}'_j \begin{pmatrix} v_{11} & \cdots & v_{d1} \\ \vdots & \ddots & \vdots \\ v_{1p} & \cdots & v_{dp} \end{pmatrix} \\
 &= \rho (v_{1j}, v_{2j}, \dots, v_{dj}) \\
 &= \rho \mathbf{b}_j^{0j} = \beta'_j
 \end{aligned} \tag{1.3}$$

où $\mathbf{b}_j^{0j} = (v_{1j}, v_{2j}, \dots, v_{dj}) = \beta'_j$.

La projection de l'axe $\rho \mathbf{e}_j$ sur \mathbf{V}_d est dans la direction $(v_{1j}, v_{2j}, \dots, v_{dj})$, c'est-à-dire dans la direction de la j -ième composante de chaque vecteur $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$. L'axe $\rho \mathbf{e}_j$ peut donc être représenté par la droite $\rho (v_{1j}, v_{2j}, \dots, v_{dj})$, $-\infty < \rho < +\infty$; cette droite est appelée axe biplot β_j .

Gabriel (1971) avait déjà suggéré que $(v_{1j}, v_{2j}, \dots, v_{dj})$ représente la j -ième variable dans R_d , $d < p$. Gabriel (1995) utilise une notation basée sur la décomposition en valeurs singulières :

$$\mathbf{Y} = \mathbf{U}\mathbf{T}\mathbf{V}' = \mathbf{A}^* \mathbf{B}^0$$

où $\mathbf{A}^* = \mathbf{U}\mathbf{T}$ et $\mathbf{B}^0 = \mathbf{V}'$. La i -ième ligne de la matrice \mathbf{A}^* est le vecteur colonne du i -ième individu; la j -ième colonne de la matrice \mathbf{B}^0 est la direction de l'axe biplot qui représente la j -ième variable. L'approximation $\mathbf{Y}_{[d]}$ de \mathbf{Y} est:

$$\mathbf{Y}_{[d]} = \mathbf{A}_{[d]}^* \mathbf{B}_{[d]}^0 \tag{1.4}$$

où $\mathbf{A}_{[d]}^*$ est une matrice constituée des d premiers vecteurs colonnes de la matrice \mathbf{A}^* , et $\mathbf{B}_{[d]}^0$ est la matrice formée des d premières lignes de \mathbf{B}^0 .

On obtient ainsi une représentation simultanée des individus et des variables sur le même graphique :

1. la projection des individus \mathbf{y}_i sur \mathbf{V}_d donne une représentation des individus dans

le sous-espace R_d par des points $\hat{P}_i \in R_d$ dont les coordonnées sont les composantes du vecteur $\mathbf{a}_{i[d]}^*$;

- la projection des axes $\rho \mathbf{e}_j$ sur \mathbf{V}_d donne une représentation des axes $\rho \mathbf{e}_j$ par des axes biplot β_j ($j = 1, \dots, p$) dans ce même sous-espace R_d .

Nous verrons que cette représentation simultanée permet d'interpoler géométriquement dans R_d de nouveaux points $\xi = (\xi_1, \xi_2, \dots, \xi_p)$ et de prédire géométriquement à partir d'un nouveau point $\hat{P} \in R_d$ des valeurs approximatives aux ξ_j ($j = 1, \dots, p$). Nous verrons aussi qu'il sera facile de retrouver l'information $\mathbf{y}_{i[d]}$ à partir des points \hat{P}_i ($i = 1, \dots, n$).

1.4.1 Propriétés du biplot

Nous allons étudier les propriétés de cette représentation simultanée.

- La **géométrie** des vecteurs \mathbf{b}_j^0 du biplot dans l'espace des individus R_p n'a pas conservé l'interprétation des concepts statistiques des vecteurs avec coordonnées $\mathbf{b}_j^* \in R_n$ représentant les variables dans l'espace des variables. La longueur du vecteur \mathbf{b}_j^0 , l'angle entre b_j^0 et b_k^0 ne renseignent nullement sur la variance s_j^2 et la corrélation $\text{cor}(\mathbf{y}_{(j)}, \mathbf{y}_{(k)})$.
- La valeur y_{ij} est approximée par le produit scalaire $\mathbf{a}_{i[d]}^{*'} \cdot \mathbf{b}_{j[d]}^0$ où $\mathbf{a}_{i[d]}^{*'}$ est le vecteur des coordonnées du point \hat{P}_i et $\mathbf{b}_{j[d]}^0$ est la direction de l'axe biplot β_j . D'après l'équation 1.4, un élément de $\mathbf{Y}_{[d]}$ noté par $y_{ij[d]}$ est donné par :

$$y_{ij[d]} = \mathbf{a}_{i[d]}^{*'} \cdot \mathbf{b}_{j[d]}^0 \quad (i = 1, \dots, n; j = 1, \dots, p)$$

- La **projection** du point-individu \hat{P}_i sur l'axe biplot β_j est la propriété la plus importante du biplot. Cette projection est :

$$\frac{\mathbf{a}_{i[d]}^{*'} \cdot \mathbf{b}_{j[d]}^0}{\|\mathbf{b}_{j[d]}^0\|^2} = \frac{y_{ij[d]}}{\|\mathbf{b}_{j[d]}^0\|^2}.$$

- (a) Si $\mathbf{b}_j^0 \in R_p$ alors $\|\mathbf{b}_j^0\| = 1$ et la projection du point individu P_i sur l'axe biplot β_j est exactement y_{ij} .
- (b) Si $\mathbf{b}_j^0 \in R_d$ alors la projection de chaque point-individu \hat{P}_i sur l'axe biplot β_j nous renseigne sur la valeur $y_{ij[d]}$ au facteur $\|\mathbf{b}_j^0\|$ près. Ce facteur est constant pour tout point \hat{P}_i projeté sur l'axe biplot β_j .

Afin de ne pas avoir à multiplier la longueur de la projection par $\|\mathbf{b}_j^0\|^2$ pour reconstituer $\mathbf{y}_{j[d]}$, Gower (1995) suggère d'établir une échelle appropriée sur l'axe biplot β_j . Nous discuterons de ce point dans ce qui suit.

1.4.2 Interpolation et prédiction

Les axes biplot ont une grande utilité : retrouver l'information $\mathbf{y}_{i[d]}$ à partir du point \hat{P}_i , interpoler géométriquement un point \hat{P} à partir du nouveau vecteur $\xi = (\xi_1, \xi_2, \dots, \xi_p)$, et prédire à partir d'un point \hat{P} quelconque les valeurs approximatives $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_p)$.

Interpolation géométrique

L'interpolation d'une nouvelle valeur $\xi \neq \mathbf{y}_i$ ($i = 1, \dots, n$), $\xi' = (\xi_1, \xi_2, \dots, \xi_p)$ dans le sous-espace de représentation R_d peut se faire en calculant algébriquement $\xi' \mathbf{V}_d$. Cette façon de procéder n'est pas aisée pour un utilisateur. L'interpolation de ξ peut aussi se faire géométriquement à l'aide des axes biplot β_j ($j = 1, \dots, p$). Nous allons décrire la procédure d'interpolation géométrique telle que proposée par Gower(1992).

La projection orthogonale de $\xi = \sum_{j=1}^p \xi_j \mathbf{e}_j$ sur \mathbf{V}_d est $\xi' \mathbf{V}_d = \sum_{j=1}^p (\xi_j \mathbf{e}'_j) \mathbf{V}_d = \sum_{j=1}^p \xi_j (\mathbf{e}'_j \mathbf{V}_d)$ où $\mathbf{e}'_j \mathbf{V}_d$ représente le vecteur unité le long de l'axe β_j .

On a ainsi que $\xi' \mathbf{V}_d$ est une somme de p vecteurs $\xi_j (\mathbf{e}'_j \mathbf{V}_d)$ ($j = 1, \dots, p$) où chaque vecteur $\xi_j (\mathbf{e}'_j \mathbf{V}_d)$ est un vecteur le long de l'axe β_j . Si les axes biplot sont balisés à l'aide des valeurs ξ_j , on peut alors positionner le point $\xi_j (\mathbf{e}'_j \mathbf{V}_d)$ par l'addition de ces p vecteurs avec la méthode du parallélogramme. Nous pourrions ainsi interpoler le point

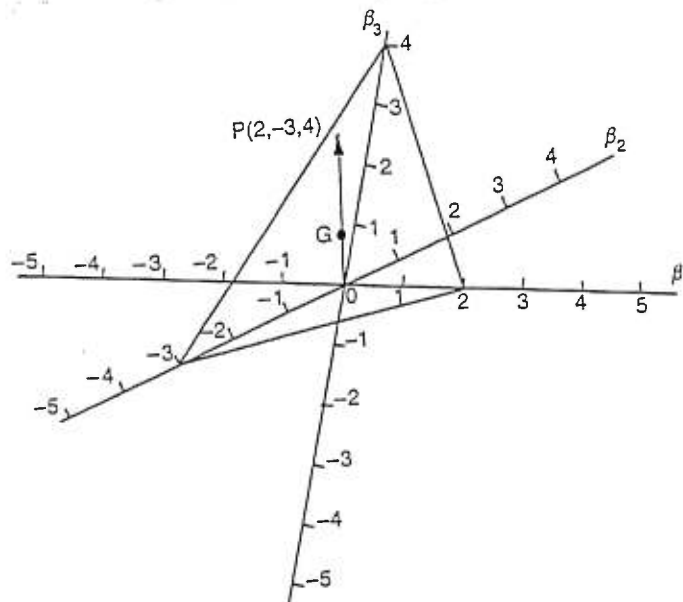


Figure 1.1: Interpolation du point $P(2, -3, 4)$, tiré de Gower (1996), page 14.

\hat{P} qui représente l'individu ξ .

L'échelle sur les axes biplot est établie en projetant ρe_j sur V_d pour $\rho = 0, \pm 1, \pm 2, \dots, \pm k$. Il n'est pas aisé d'utiliser la méthode du parallélogramme, il est plus pratique de projeter dans R_d le centre de gravité $\sum_{j=1}^p \left\{ \frac{1}{2} (\xi_j e_j) \right\}$ du parallélogramme dans R_p défini par les vecteurs $\xi_j e_j$, puis de multiplier la projection du centre de gravité dans R_p par p . La démarche d'interpolation est illustrée à la figure 1.1.

Sur la figure 1.1, on voit comment interpoler le point $P(2, -3, 4)$ par rapport à trois axes biplot β_1, β_2 , et β_3 . On calcule le centre de gravité G des points $(2, 0, 0)$, $(0, -3, 0)$ et $(0, 0, 4)$ puis on multiplie le vecteur \overrightarrow{OG} par 3 (le nombre d'axe biplot).

Prédiction

L'échelle des axes biplot qui sert à l'interpolation ne peut servir pour la prédiction. En effet le pied de la projection orthogonale du point P sur l'axe β_j ne coïncide pas avec l'échelle d'interpolation (voir figure 1.2).

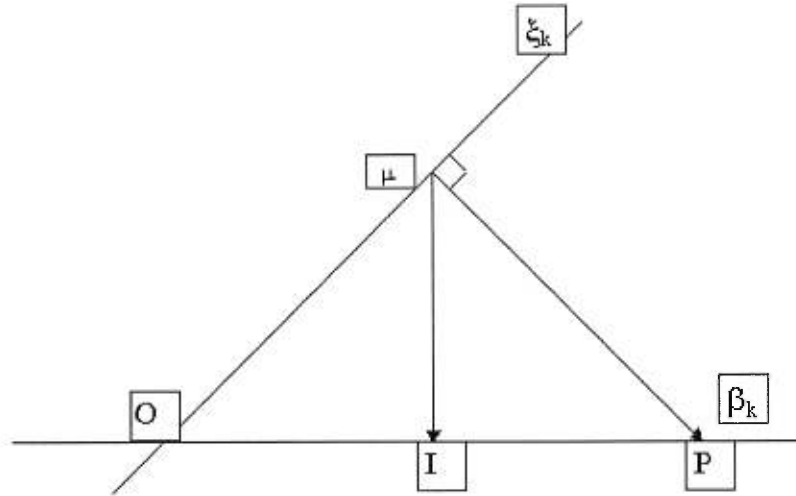


Figure 1.2: Relation entre les marqueurs de prédiction P et ceux d'interpolation I

Gower (1995) suggère de baliser les axes biplot β_j ($j = 1, \dots, p$) avec une autre échelle afin que l'ordre de grandeur des $\mathbf{y}_i \approx \mathbf{y}_{i[d]}$ puisse être jugé à l'aide du pied de la projection du point individu \hat{P}_i sur l'axe biplot β_j .

La propriété du produit scalaire nous permet d'écrire :

$$y_{ij[d]} = \mathbf{a}_i^{*t} \mathbf{b}_j^0$$

d'où

$$\frac{y_{ij[d]}}{\|\mathbf{b}_j^0\|^2} = \frac{\mathbf{a}_i^{*t} \mathbf{b}_j^0}{\|\mathbf{b}_j^0\|^2}$$

Le terme de droite est la projection du point \hat{P}_i sur l'axe β_j de direction \mathbf{b}_j^0 . Si le pied de la projection est marqué par $y_{ij[d]}$, la lecture de la valeur s'en suivra. L'échelle pour la prédiction est établie en marquant l'axe biplot β_j par des valeurs ρ pour $\rho = \pm 1, \pm 2, \dots$ lors de la projection de $\frac{\rho \mathbf{e}_j}{\|\mathbf{b}_j^0\|^2}$ sur \mathbf{V}_d .

1.4.3 Utilité du biplot dans R_2

Le biplot dans R_2 a une grande utilité. Outre la récupération de l'information, l'interpolation et la prédiction, on peut, par exemple, donner une interprétation aux regroupements des individus grâce à leur projection sur chaque axe β_j . En effet, si les projections de ces individus sur β_j sont assez rapprochées, alors ces individus ont des valeurs $y_{ij[d]}$ assez rapprochées. On ne pourrait détecter à l'oeil ces similitudes sur la matrice de données \mathbf{Y} : la dimension $n \times p$ nous en empêcherait même si celle-ci n'est pas trop grande.

1.4.4 Biplot et analyse des correspondances

Lien entre l'espace R_p et l'espace R_n

Gabriel (1995) a établi le lien mathématique entre l'espace des variables et l'espace des individus en se servant de la décomposition en valeurs singulières de la matrice \mathbf{Y} :

$$\begin{aligned} \mathbf{Y} &= \mathbf{U}\mathbf{\Gamma}\mathbf{V}' \\ &= \mathbf{A}^*\mathbf{V}' \end{aligned} \tag{1.5}$$

$$= \mathbf{U}\mathbf{B}^{*'} \tag{1.6}$$

où $\mathbf{A}^* = \mathbf{U}\mathbf{\Gamma}$ et $\mathbf{B}^{*'} = \mathbf{\Gamma}\mathbf{V}'$

L'équation 1.5 indique que la représentation des individus dans l'espace des individus est faite avec un système d'axes dont les directions sont données par les vecteurs lignes de \mathbf{V}' ; les coordonnées des point-individus sont données par les lignes de $\mathbf{A}^* = \mathbf{U}\mathbf{\Gamma}$.

L'équation 1.6 indique que la représentation des variables dans l'espace des variables est faite avec un système d'axes données par les colonnes de \mathbf{U} ; les coordonnées des vecteurs représentant ces variables sont données par les colonnes de $\mathbf{B}^* = \mathbf{\Gamma}\mathbf{V}'$.

Cette correspondance entre les deux espaces R_p et R_n a toujours été faite dans

le passé en utilisant une matrice de transition entre les vecteurs propres de $\mathbf{Y}'\mathbf{Y}$ et de $\mathbf{Y}\mathbf{Y}'$, c'est-à-dire entre les vecteur-colonnes de \mathbf{U} et de \mathbf{V} (voir Lebart, 1984). La notation de Gabriel (1995) nous permet d'établir un lien entre ces deux espaces qui est facile à visualiser.

Analyse des correspondances

Le lien entre les espaces des individus et des variables apporte un très grand éclaircissement sur l'analyse en correspondance développée par Benzécri (1992). La matrice \mathbf{X} de fréquence relative d'un tableau de contingence y est étudiée en fonction de la décomposition en valeurs spectrale de $\mathbf{X}\mathbf{X}'$ et de $\mathbf{X}'\mathbf{X}$ où \mathbf{X} est doublement centrée. La représentation des individus et des variables est faite en fonction de cette décomposition. Gabriel (1995) a explicité que cette représentation revient à utiliser les équations 1.5 et 1.6.

En analyse des correspondances, les coordonnées des points-individus sont données par les lignes de \mathbf{A}^* . Les coordonnées des variables sont données par les colonnes de \mathbf{B}^* . Cette analyse des correspondance telle que proposée par Benzécri (1992) est commentée par Gabriel (1995). Les arguments de Gabriel reposent sur le fait que la matrice \mathbf{A}^* fournit les coordonnées des individus dans la base \mathbf{V} et que la matrice \mathbf{B}^* fournit les coordonnées des variables dans la base \mathbf{U} . On ne peut pas utiliser \mathbf{A}^* et \mathbf{B}^* dans un même espace R_d car les bases sont différentes. Gabriel (1995), Goodman (1996) rejettent avec ces arguments mathématiques l'analyse des correspondances. Gabriel remarque toutefois que si les valeurs γ_1, γ_2 sont à peu près égales, alors la représentation obtenue par Benzécri pour les lignes et les colonnes du tableau de contingence est très semblable à celle décrite par le biplot. Ce qui explique pourquoi l'analyse des correspondances donnent de bons résultats.

1.5 Analyse de données

1.5.1 Ensemble de données

L'ensemble de données utilisé pour illustrer les différents concepts théoriques des biplots est mis à notre disposition par Madame Francine Mayer, du Département de biologie de l'Université du Québec à Montréal. F. Mayer et Y. Lavoie (1984) donne la description suivante de la population de laquelle a été tiré l'échantillon: « Les données ont été recueillies dans l'île de Saint-Barthélemy dans les Antilles françaises. Cette île possède des caractéristiques spécifiques qui tiennent d'une part à sa formation du nombre restreint de fondateurs, d'autre part à son isolement biologique quasi total. Dans cette communauté, la surdit  est un handicap  prouv  par une tr s grande proportion de la population ». L' tude de ce m moire est faite sur un sous- chantillon de 61 habitants de l' le dont 31 sont sourds et 30 sont non sourds. L'objectif de Madame Mayer est de d celer des liens entre ces 61 habitants et les 172 anc tres-fondateurs de la population de l' le. Une mesure P_{ij} du degr  de parent  de l'habitant j avec le fondateur i a  t  calcul e sur la base des donn es originales. A l'aide de ces mesures de parent , nous avons d fini six variables Ns , $\bar{X}s$, Ss , Nn , $\bar{X}n$, et Sn :

1. le nombre Ns_i d'habitants sourds n'ayant aucun lien de parent  avec le fondateur i (c'est- -dire le nombre de P_{ij} , $1 \leq j \leq 31$ avec valeur z ro);
2. le nombre Nn_i d'habitants non sourds ayant aucun lien de parent  avec le fondateur i (c'est- -dire le nombre de P_{ij} , $32 \leq j \leq 61$ avec valeur z ro);
3. le degr  de parent  moyen $\bar{X}s_i$ des habitants sourds avec le fondateur i : $\bar{X}s_i = \frac{1}{31} \sum_{j=1}^{31} P_{ij}$. L' cart type $Ss_i = \left(\frac{1}{30} \sum_{i=1}^{31} (P_{ij} - \bar{X}s_i)^2 \right)^{\frac{1}{2}}$;
4. le degr  de parent  moyen $\bar{X}n_i$ des habitants non sourds avec le fondateur i : $\bar{X}n_i = \frac{1}{30} \sum_{j=32}^{61} P_{ij}$. L' cart type $Sn_i = \left(\frac{1}{29} \sum_{i=32}^{61} (P_{ij} - \bar{X}n_i)^2 \right)^{\frac{1}{2}}$.

On dispose aussi de six variables confondantes (« confounding variables ») nécessaires pour cette étude; celles-ci sont entreposées dans les 6 premières colonnes de la matrice \mathbf{Y} de données de dimension (172×12) . Ces six variables sont définies comme suit :

- ANN : indique l'année d'arrivée du fondateur sur l'île
- $NBGEN$: indique sur combien de générations la descendance s'étend
- $IDESC$: indique le nombre d'individus dans la descendance
- $IUTIL$: indique le nombre de descendants qui se sont reproduits
- $IREP$: indique le nombre de descendants qui le sont par leur père et leur mère
- $NBFRAT$: indique le nombre de fratries de descendants

La matrice de données $\mathbf{Y}_{172 \times 12}$ est constituée des observations sur les 12 variables ($ANN, NBGEN, IDESC, IUTIL, IREP, NBFRAT, N_s, \bar{X}_s, S_s, N_n, \bar{X}_n, S_n$) faites sur chacun des 172 fondateurs. La visualisation de cet ensemble de données en utilisant la technique du biplot est réalisée : *i*) avec les variables confondantes c'est-à-dire sur une matrice $\mathbf{Y}_{172 \times 6}$ (voir figure 1.3); *ii*) avec les six variables ($N_s, \bar{X}_s, S_s, N_n, \bar{X}_n, S_n$) c'est-à-dire sur une autre matrice $\mathbf{Y}_{172 \times 6}$ (voir figure 1.4); *iii*) avec les dix variables $NBGEN, IDESC, IUTIL, IREP, NBFRAT, \bar{X}_s, S_s, N_n, \bar{X}_n,$ et S_n c'est-à-dire sur une autre matrice $\mathbf{Y}_{172 \times 10}$ (voir figure 1.5).

1.5.2 Biplot des 172 fondateurs et des 6 variables confondantes

Les six variables confondantes ayant des échelles différentes, nous les avons centrées et réduites avant de faire la représentation des individu-fondateurs dans un plan. Le lien entre les six variables et les 172 fondateurs est très bien visualisé sur le biplot (voir figure 1.3). Les fondateurs sont représentés par des points à l'aide de l'ACP exécuté sur la matrice centrée et réduite $\mathbf{Y}_{172 \times 6}$.

Les axes principaux \mathbf{v}_1 et \mathbf{v}_2 de cette représentation n'ont aucun intérêt. Notre analyse ne se fera pas relativement à ces deux axes : les axes biplot seront étudiés

explicitement car ils nous donnent de l'information sur les variables originales qui n'est pas donné par les composantes principales : celles-ci sont des combinaisons linéaires des variables originales. Notons que les deux premières composantes ou les deux premiers axe principaux sont très explicatifs. La variance expliquée est de 89,6% ($\sum_{i=1}^2 \lambda_i / \sum_{i=1}^6 \lambda_i$) (voir tableau 1.1).

	<i>ANN</i>	<i>NBGEN</i>	<i>IDESC</i>	<i>IUTIL</i>	<i>IREP</i>	<i>NBFRAT</i>
\mathbf{v}'_1	-0,3417	0,3621	0,4376	0,4359	0,4249	0,4357
\mathbf{v}'_2	-0,6424	0,5795	-0,2080	-0,2376	-0,3068	-0,2395
<i>valeurs propres λ_i</i>	29,17	12,76	3,07	1,60	0,14	0,02
<i>cumulatifs</i>	62.38	89.67	96.23	99.65	99.95	1

Tableau 1.1: Résultats de l'ACP sur les variables confondantes

La direction des axes biplot de chaque variable est donnée par le vecteur colonne du tableau 1.1. Par exemple la direction de l'axe biplot qui représente *ANN* est $(-0,3417, -0,6424)$. Remarquons que l'axe biplot *ANN* est de direction opposée à celui de *NBGEN*. L'axe biplot correspondant à chaque variable est étiqueté sur la figure 1.3 par le nom de la variable : l'emplacement du nom indique le sens positif de l'axe biplot. Ainsi, les axes biplot de *ANN* et de *NBGEN* sont de sens opposé. Les symboles indiquent ± 1 ou ± 2 écart type de la moyenne.

On remarque deux nuages de points sur la figure 1.3, le premier s'échelonne le long des axes biplot (*ANN* et *NBGEN*). Les vieux fondateurs qui sont arrivés sur l'île dans les années 1800 et moins ont des petites valeurs sur l'axe *ANN*; ils ont de nombreux descendants donc des grandes valeurs sur l'axe *NBGEN* (*les axes ANN et NBGEN ont des directions opposées*). Le deuxième nuage de points est celui qui s'échelonne le long des quatre autres axes biplot (*IDESC, IUTIL, IREP, NBFRAT*) : ces axes ont même direction. Les fondateurs avec des grandes valeurs sont ceux qui sont arrivés très tôt sur l'île.

D'après la propriété du biplot classique à la section 4 du chapitre 1, l'approximation d'Eckart-Young est $y_{ij[2]} = \mathbf{a}_i^* \mathbf{b}_j^0$, où les \mathbf{b}_j^0 sont les directions des axes biplot données par les colonnes du tableau 1.1 pour chaque variable.

Posons $\mathbf{V}_2 = \begin{pmatrix} \mathbf{v}'_1 \\ \mathbf{v}'_2 \end{pmatrix} = (\mathbf{b}_1^0, \mathbf{b}_2^0, \dots, \mathbf{b}_6^0)$, alors l'approximation d'Eckart-Young est $\mathbf{y}_{i[2]} = \mathbf{a}_i^* \mathbf{V}_2$. Par exemple, le sixième fondateur a comme coordonnées $\mathbf{a}_6^* = (4, -0.2)$, alors l'approximation Eckart-Young est :

$$(4, -0.2)\mathbf{V}_2 = (-1.2386, 1.3327, 1.7921, 1.7912, 1.7612, 1.7909).$$

Sur la figure 1.3, on peut visualiser ces valeurs par les projections orthogonales du point représentant le sixième fondateur sur chacun des axes biplot; remarquons que sur l'échelle originale des variables les valeurs approchées sont

$$(1706.843, 11.267, 5117.46, 1673.106, 3035.872, 1753.63);$$

la vraie valeur \mathbf{y}_i est $(1700, 11, 5193, 1684, 2889, 1767)$. On a ainsi vu, par un exemple, comment faire ces calculs visuellement avec les axes biplot.

Une grande perte d'information sur une variable est due à une mauvaise approximation du vecteur unité $\mathbf{e}_{j(6 \times 1)}$ par sa projection sur l'espace engendré par \mathbf{v}_1 et \mathbf{v}_2 c'est-à-dire $(\mathbf{e}'_j \mathbf{v}_1, \mathbf{e}'_j \mathbf{v}_2)$. Si la longueur du vecteur $(\mathbf{e}'_j \mathbf{v}_1, \mathbf{e}'_j \mathbf{v}_2)$ est près de la valeur un, alors le vecteur unité est bien représenté, ce qui entraîne évidemment que la variable est bien représentée.

Le tableau 1.2 donne la norme de chaque vecteur unité \mathbf{e}_j projeté sur l'espace engendré par $(\mathbf{v}_1, \mathbf{v}_2)$.

<i>ANN</i>	<i>NBGEN</i>	<i>IDESC</i>	<i>IUTIL</i>	<i>IREP</i>	<i>NBFRAT</i>
0,5294	0,4669	0,2347	0,2464	0.2746	0,2471

Tableau 1.2: Norme de la projection pour chaque variable confondante

On remarque que les variables ANN et $NBGEN$ sont les mieux représentées : par exemple les approximations d'Eckart-Young pour ces variables sont près de la valeur réelle pour le 6^e fondateur. Les variables $IDESC$, $IUTIL$, $IREP$ et $NBFRAT$ ne sont pas bien représentées : par exemple les valeurs prédites sont assez différentes des valeurs réelles. On dit alors que les axes relatifs à ces quatre variables sont presque orthogonaux au plan de projection.

Les axes biplot ANN et $NBGEN$ sont presque confondus (en sens opposé); les projections à partir de n'importe quel point individu (fondateur) sur ces deux axes vont prédire des valeurs telles que chacune est une transformation linéaire de l'autre. Remarquons que la corrélation entre la variable ANN et $NBGEN$ est très élevée : la corrélation est -0.94 . C'est la raison pour laquelle cette variable est éliminée des autres analyses : ce qui nous permet d'alléger le graphique.

Les axes biplot sont aussi utilisés pour interpoler un nouveau point, mais comme on a déjà vu à la section 1.4.2 au chapitre 1, l'échelle des axes d'interpolation et ceux de prédiction sont différentes. Un exemple d'interpolation d'un nouveau point est donné à la section 1.4.2.

1.5.3 Biplot des 172 fondateurs et des 6 variables de parenté

Le biplot des 172 individus et des 6 variables centrées et réduites Ns , $\bar{X}s$, Ss , Nn , $\bar{X}n$, et Sn est donnée à la figure 1.4.

On visualise que les fondateurs ayant un degré de parenté élevé avec les 61 habitants (grande valeur pour $\bar{X}s$ et $\bar{X}n$) sont en général parent avec presque tous les 61 habitants (Ns et Nn ont de petites valeurs) : il suffit de projeter, par exemple, les point-fondateurs 10, 4, 12, etc. sur les axes biplot pour s'en rendre compte. Le fondateur 65 a un degré de parenté élevé avec seulement certains habitants : $\bar{X}s$, $\bar{X}n$ élevés Ns et Nn moyen. Le graphique fait ressortir les deux fondateurs 152 et 158. En effet, ces fondateurs ont un degré de parenté faible avec les habitants sourds et un degré de parenté élevé avec les habitants non sourds : $\bar{X}s$ faible, $\bar{X}n$ élevé, Ns et Nn élevé. Il

faudrait investiguer davantage pour expliquer ce comportement de ces deux fondateurs.

On remarque aussi sur la figure 1.4 deux nuages de points; le premier est celui qui contient un très grand nombre d'individus (fondateurs) et qui sont bien concentrés autour du centre de gravité de la configuration : ce sont des jeunes fondateurs ; l'autre nuage de points est constitué de fondateurs qui sont arrivés très tôt sur l'île : ils ont un très grand score sur les variables \bar{X}_s , S_s , et \bar{X}_n (petits scores sur Nn et Ns).

L'ACP exécutée sur cette matrice $\mathbf{Y}_{172 \times 6}$ donne les deux premiers axes principaux \mathbf{v}_1 et \mathbf{v}_2 ; la variance expliquée par ces deux axes est 69.89%. (voir le tableau 1.3).

	Ns	\bar{X}_s	S_s	Nn	\bar{X}_n	S_n
\mathbf{v}'_1	0,4372	-0,4588	-0,3579	0,4344	-0,4574	-0,2687
\mathbf{v}'_2	-0,3303	0.0926	-0,1214	-0,3025	-0,0853	-0,8768
<i>valeurs propres λ_i</i>	27,76	11,83	9,44	4,46	2,20	0,95
<i>cumulatifs</i>		0,7	0,86	0,95	0,98	1

Tableau 1.3: Résultats de l'ACP sur les variables de parenté

Posons $\mathbf{V}_2 = \begin{pmatrix} \mathbf{v}'_1 \\ \mathbf{v}'_2 \end{pmatrix} = (\mathbf{b}_1^0, \mathbf{b}_2^0, \dots, \mathbf{b}_6^0)$, où \mathbf{b}_j^0 $j = 1, \dots, 6$ est la direction du j -ième axe biplot (voir tableau 1.3); l'approximation selon le théorème d'Eckart-Young de l'individu $i = 65$ de coordonnées $\mathbf{a}_i^* = (-4.3, -2.8)$ est :

$$\mathbf{y}_{i[2]} = \mathbf{a}_i^* \mathbf{V}_2 = (-0.9551, 1.7149, 1.8790, -1.0208, 2.2056, 3.6108),$$

c'est-à-dire sur l'échelle originale des variables (15.31, 0.022, 0.023, 13.88, 0.025, 0.043); la vraie valeur est : $\mathbf{y}_i = (18, 0.027, 0.043, 22, 0.017, 0.038)$. Les valeurs prédites ne sont pas très différentes des valeurs réelles; la perte d'information est due à la position du plan de projection par rapport aux axes représentant les variables.

En effet, la projection de chaque vecteur unité \mathbf{e}_j , $j = 1, \dots, 6$ est donnée par le tableau suivant :

N_s	\bar{X}_s	S_s	N_n	\bar{X}_n	S_n
0,300	0,219	0,143	0,280	0,216	0,840

Tableau 1.4: Norme de la projection pour chaque variable de parenté

La variable la mieux représentée est S_n : la projection du vecteur \mathbf{e}_6 (S_n est la 6-ième variable) sur \mathbf{v}_1 et \mathbf{v}_2 est $(\mathbf{e}'_6 \mathbf{v}_1, \mathbf{e}'_6 \mathbf{v}_2)$ dont la norme est 0.84; sur la figure 1.4, l'axe biplot associé à cette variable S_n est presque orthogonal aux cinq autres axes relatifs aux variables $N_s, \bar{X}_s, S_s, N_n, \bar{X}_n$, ce qui est normal car la norme de $(\mathbf{e}'_j \mathbf{v}_1, \mathbf{e}'_j \mathbf{v}_2)$ $j = 1, \dots, 5$ est faible : les axes relatif à ces variables sont presque orthogonaux au plan de projection.

Les axes biplot permettent de faire les calculs de prédiction visuellement en projetant orthogonalement par exemple le point individu $i = 65$ sur chaque axe biplot j . La figure 1.5, est une représentation des 172 individus (fondateurs) et des 10 variables $NBGEN, IDESC, IREP, NBFRAT, N_s, \bar{X}_s, S_s, N_n, \bar{X}_n$ et S_s . Pour alléger la figure 1.5, on n'a pas représenté les variables ANN et $IUTIL$ car elles sont très corrélées respectivement avec les variables $NBGEN$ et $NBFRAT$. La disposition des point-individus n'est pas différente de celle de la figure 1.4 : on remarque que les fondateurs qui sont arrivés très tard sur l'île sont regroupés autour du centre de gravité, et qu'un petit nuage de points est formé des fondateurs qui sont arrivés très tôt sur l'île. Ces derniers ont des scores très élevés sur les variables $IDESC, IREP, NBFRAT, \bar{X}_s$ et \bar{X}_n .

1.5.4 Espace dual

La représentation des variables confondantes centrées est faite à la figure 1.6. La longueur des différents vecteurs représente approximativement la variance de chacune des variables. On voit que la longueur de la variable $NBGEN$ est à peu près zero : la variance de cette variable est très petite comparée à celles de $IDESC, ANN, IREP, NBFRAT$ et $IUTIL$.

	<i>ANN</i>	<i>NBGEN</i>	<i>IDESC</i>	<i>IUTIL</i>	<i>IREP</i>	<i>NBFRAT</i>
<i>moyenne</i>	1774,116	8,2441	1392,18	423,3314	624,5116	442,4477
<i>écart type</i>	54,3121	2,2681	2078,681	697,7146	1369,138	732,1268

	<i>ANN</i>	<i>NBGEN</i>	<i>IDESC</i>	<i>IUTIL</i>	<i>IREP</i>	<i>NBFRAT</i>
	N_s	\bar{X}_s	S_s	N_n	\bar{X}_n	S_n
<i>moyenne</i>	23,8604	0,0058	0,0079	22,7558	0,0058	0,0092
<i>écart type</i>	8,9490	0,0096	0,0081	8,6874	0,0088	0,0095

Tableau 1.5: Moyennes et écarts types des variables confondantes et des variables de parenté

La représentation des variables centrées réduites dans la figure 1.7 illustre les relations des différentes variables entre elles, c'est à dire, les différentes corrélations. Le cosinus de l'angle entre deux vecteurs approxime la corrélation entre les deux variables. La figure 1.7 montre que l'angle entre les vecteurs représentant les variables *ANN* et *NBGEN* est presque 180 degrés, le cosinus de l'angle est presque égale à (-1) , ce qui n'est autre que le coefficient de corrélation entre les variables *ANN* et *NBGEN*. Les variables bien représentées devraient avoir une représentation par un vecteur de longueur égale à l'unité. On voit donc que certaines variables sont moins bien représentées.

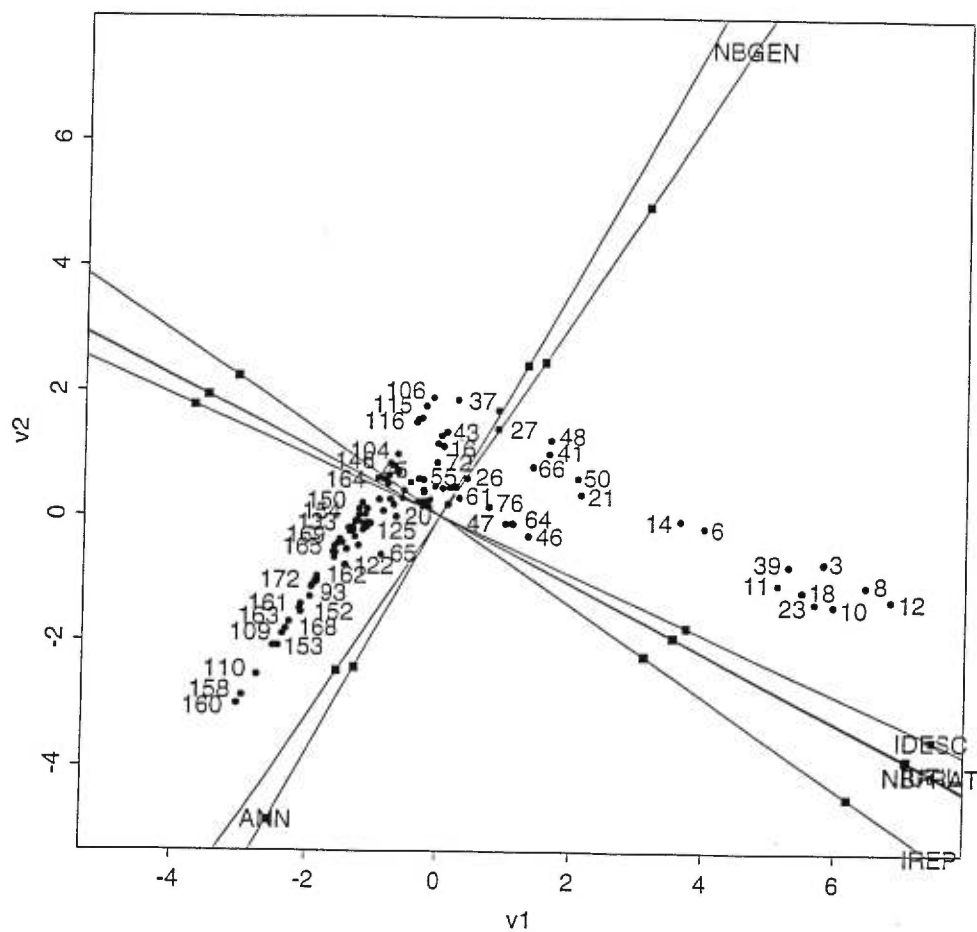


Figure 1.3: Axes biplot : ANN, NBGEN, IDESC, IUTIL, IREP et NBFAT

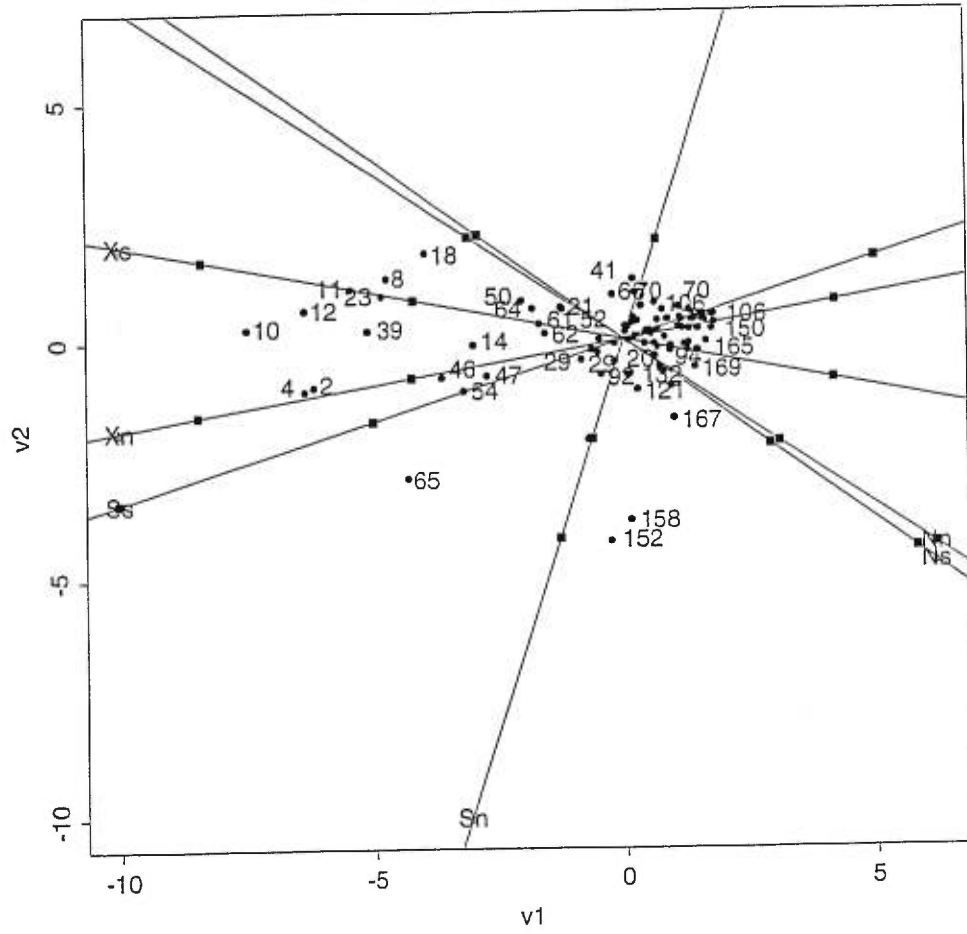


Figure 1.4: Axes biplot : Ns , $\bar{X}s$, Ss , Nn , $\bar{X}n$, et Sn

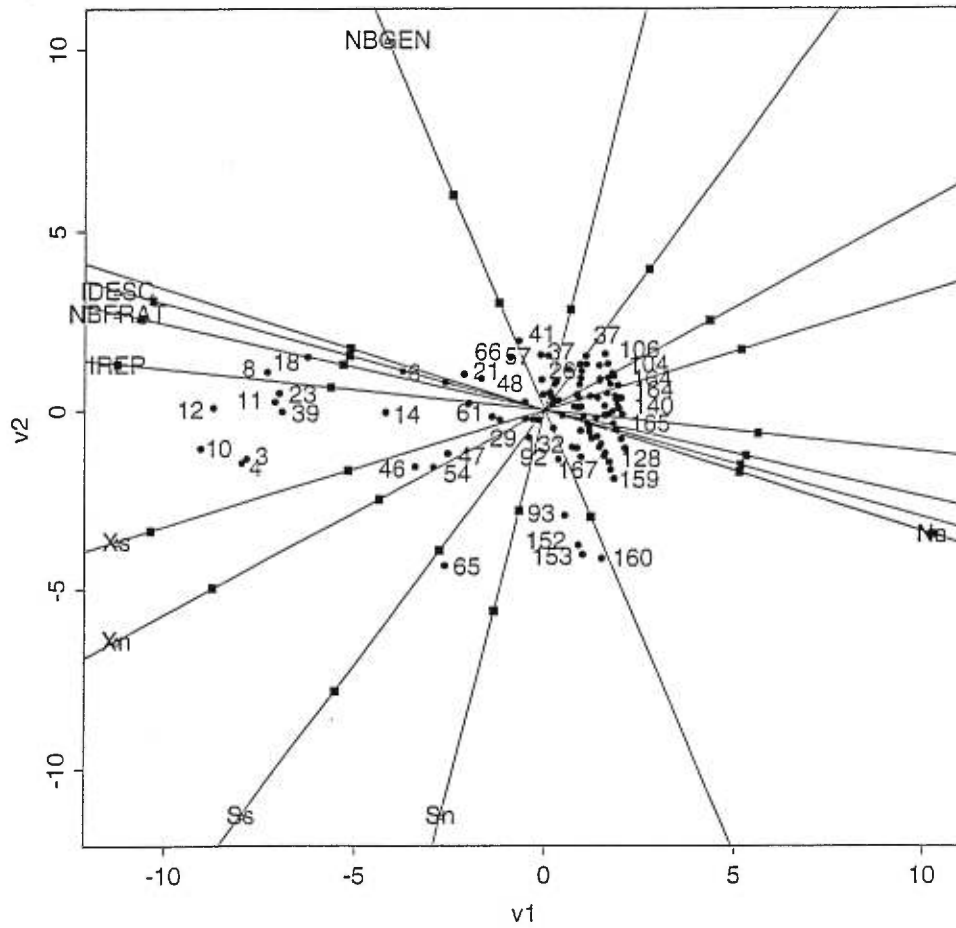


Figure 1.5: Axes biplot : Variables confondantes (4) et de parenté (6)

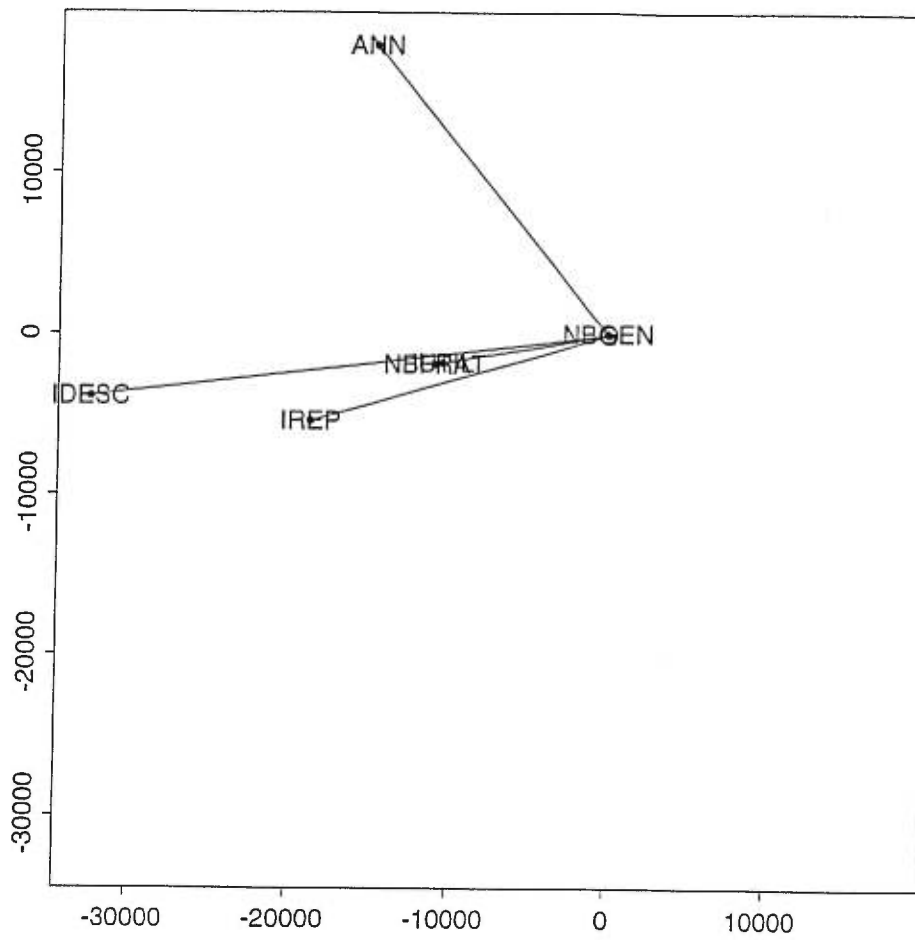


Figure 1.6: Espace dual : Variables confondantes centrées et non réduites

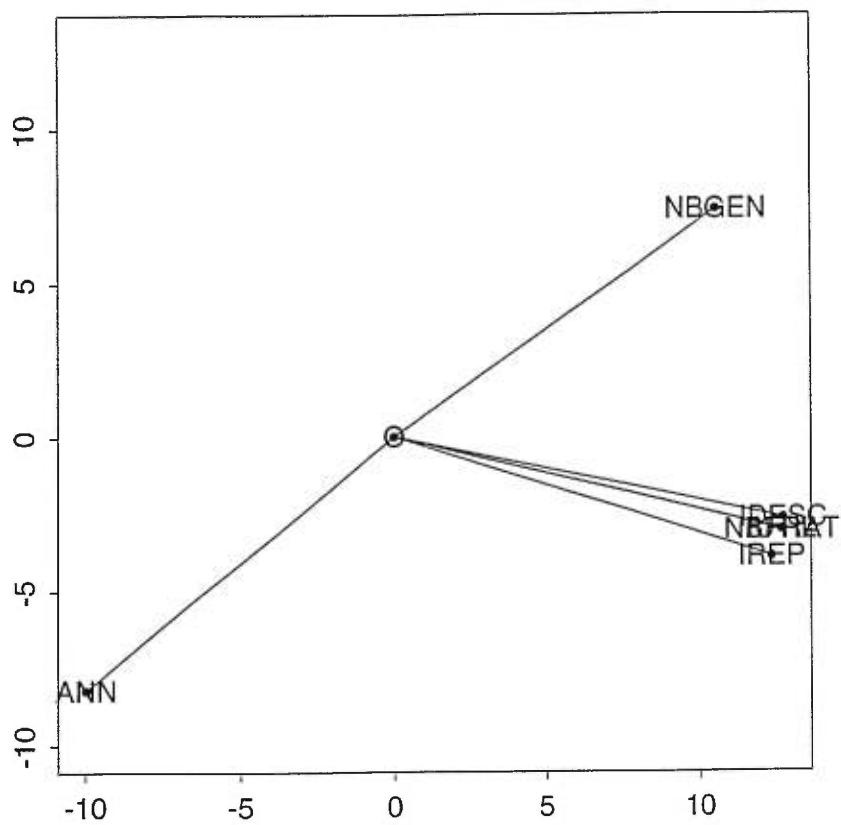


Figure 1.7: Espace dual : Variables confondantes centrées réduites

CHAPITRE II

RÉSULTATS GÉNÉRAUX SUR L'ALGÈBRE ET LA GÉOMÉTRIE DU BIPLLOT LINÉAIRE

2.1 Introduction

Rappelons que la discussion du biplot classique se faisait dans le sous-espace des individus R_d , $d < p$ défini par l'ACP et de plus, les distances inter-individus dont il était question étaient mesurées par la métrique euclidienne. L'élaboration de la théorie générale du biplot est doublement motivée par ces deux aspects: espace de représentation des individus et la métrique utilisée pour mesurer la proximité des individus.

Le sous-espace \mathcal{L} de représentation des point-individus peut être différent de R_d , $d \leq p$: le sous-espace \mathcal{L} peut être tout à fait quelconque ($\mathcal{L} \subseteq R_p$ ou $\mathcal{L} \not\subseteq R_p$).

En effet les vecteurs qui engendrent le sous-espace $\mathcal{L} \subset R_p$ peuvent avoir des directions autres que celles données par les vecteurs propres de $\mathbf{Y}'\mathbf{Y}$. Krzanovski (1993 p.81 – 83) donne un aperçu de ce problème. Nous en présenterons un bref résumé:

Beaucoup d'intérêts ont été manifestés depuis quelques temps, dans la recherche de sous-espace \mathcal{L} de dimension $d \leq p$ qui font ressortir d'autres aspects spécifiques de la structure existant dans l'ensemble de données. Toutes ces méthodes nécessitent d'établir des critères quantitatifs de telle sorte que le sous-espace \mathcal{L} optimise un de ces critères. On recherche donc la direction des axes de \mathcal{L} qui optimisent ce critère. Cette technique

est appelée recherche d'une projection (projection poursuit).

Un exemple de ce critère est la recherche d'une direction qui offre le minimum d'entropie. Etant donné que l'entropie est maximale lorsque la distribution est normale (variance constante), on trouve ainsi une direction selon laquelle l'ensemble de données a un comportement qui s'éloigne le plus d'une normale.

La métrique utilisée pour mesurer la distance entre les vecteur-individus peut être « non pythagorienne » ce qui entraîne, comme nous le verrons, un sous-espace de représentation $\mathcal{L} \subsetneq R_p$. Nous discuterons des situations où la distance est non pythagorienne dans les chapitres 3 et 4. Nous ne discuterons pas dans ce mémoire de la technique de la recherche d'une projection. Remarquons que la nécessité de tracer des axes biplot dans ces sous-espaces \mathcal{L} quelconques est toujours primordiale et c'est ce dont nous allons discuter dans ce chapitre.

La théorie générale du biplot linéaire se rapporte aux situations où la représentation des vecteur-individus est faite dans un sous-espace \mathcal{L} quelconque, $\mathcal{L} \subset R_p$ où \mathcal{L} n'est plus le sous-espace fourni par l'A.C.P., mais où l'on continue d'utiliser la distance pythagorienne pour mesurer la proximité des points individus.

Dans ce chapitre, nous allons discuter de la théorie générale du biplot linéaire sous trois aspects: unicité des axes biplot pour l'interpolation; infinité d'axes biplot pour la prédiction; notion de prédiction appelée back-projection dans $\mathcal{L} \subset R_p$ et prédiction dans un sous-espace $M \subset R_p$. Il faut noter que l'espace \mathcal{L} contient les projections \hat{P}_i des points $P_i \in R_p$, et que le sous-espace M contient de l'information originale sur les points P_i .

A la section 2.2, on fera les rappels élémentaires des principes de base de l'interpolation et prédiction dans R_d : une bonne visualisation du cas général sera facilitée. La théorie algébrique du cas général sera développée aux sections 2.3 et 2.4

2.2 Préliminaire

Interpolation et prédiction dans R_p

Dans cette section, nous verrons les étapes élémentaires de l'interpolation et prédiction dans $R_d, d \leq p$, qui serviront d'introduction aux principes de la théorie générale du biplot linéaire et non linéaire. Nous pourrons ainsi visualiser géométriquement les opérations faites dans les cas généraux. Les principes de base sont ceux des plans orthogonaux et du point le plus près.

Tout d'abord, nous décrirons les opérations élémentaires et intuitives que nous faisons lors de

1. L'interpolation de $\mathbf{y} = (y_1, y_2)$ dans R_2
2. La recherche des coordonnées (y_1, y_2) d'un point P à partir de sa position dans R_2 , c'est à dire la prédiction des coordonnées de P .

Ces opérations élémentaires seront décrites lorsque R_2 est muni

1. d'axes orthogonaux
2. d'axes obliques.

La généralisation de ces opérations à R_p est immédiate.

Ces opérations élémentaires ainsi décrites, nous permettront de dégager les principes de base et de mieux comprendre leurs généralisations pour l'interpolation et la prédiction dans un sous-espace \mathcal{L} quelconque de R_p .

Il pourrait apparaître que la description de ces principes dans R_2 est inutile, mais elle est indispensable pour la compréhension de leurs généralisations dans $\mathcal{L} \subset R_p$.

2.2.1 Système cartésien

Remarquons que l'interpolation et la prédiction sont des opérations très élémentaires dans le plan cartésien car les axes sont orthogonaux et de direction \mathbf{e}_j . Nous donnons quand même la description de ces opérations dans R_2 car l'interpolation dans \mathcal{L} et la prédiction depuis \mathcal{L} se feront selon ces mêmes principes élémentaires.

Interpolation

Le vecteur $\mathbf{y} = (y_1, y_2)$ est interpolé graphiquement dans R_2 par la méthode des plans normaux (*droites dans R_2*). On trace une droite N_1 orthogonale au premier axe passant par le marqueur y_1 , puis une deuxième droite N_2 orthogonale au deuxième axe passant par le marqueur y_2 . Le point d'intersection de ces deux droites est considéré comme le point P de coordonnées (y_1, y_2) .

L'interpolation est décrite algébriquement et géométriquement en faisant la somme des deux vecteurs $y_1\mathbf{e}_1$ et $y_2\mathbf{e}_2$: $y_1\mathbf{e}_1 + y_2\mathbf{e}_2$ (voir figure: **2.1**)

Prédiction

Afin de trouver graphiquement les coordonnées (y_1, y_2) d'un point P du plan, on utilise de nouveau la méthode des plans normaux. On trace une droite N_1 orthogonale au premier axe passant par le point P : l'intersection de N_1 avec cet axe est marquée y_1 . Puis on trace une droite N_2 orthogonale au deuxième axe passant par le point P : l'intersection de N_2 avec le deuxième axe est marquée y_2 . Remarquons que le point P est le point le plus près de chacun des axes.

Ces deux propriétés (le point le plus près et les plans orthogonaux) jouent un rôle très important dans la théorie générale (voir figure **2.1**), la généralisation à un système d'axes cartésien dans R_p est immédiate.

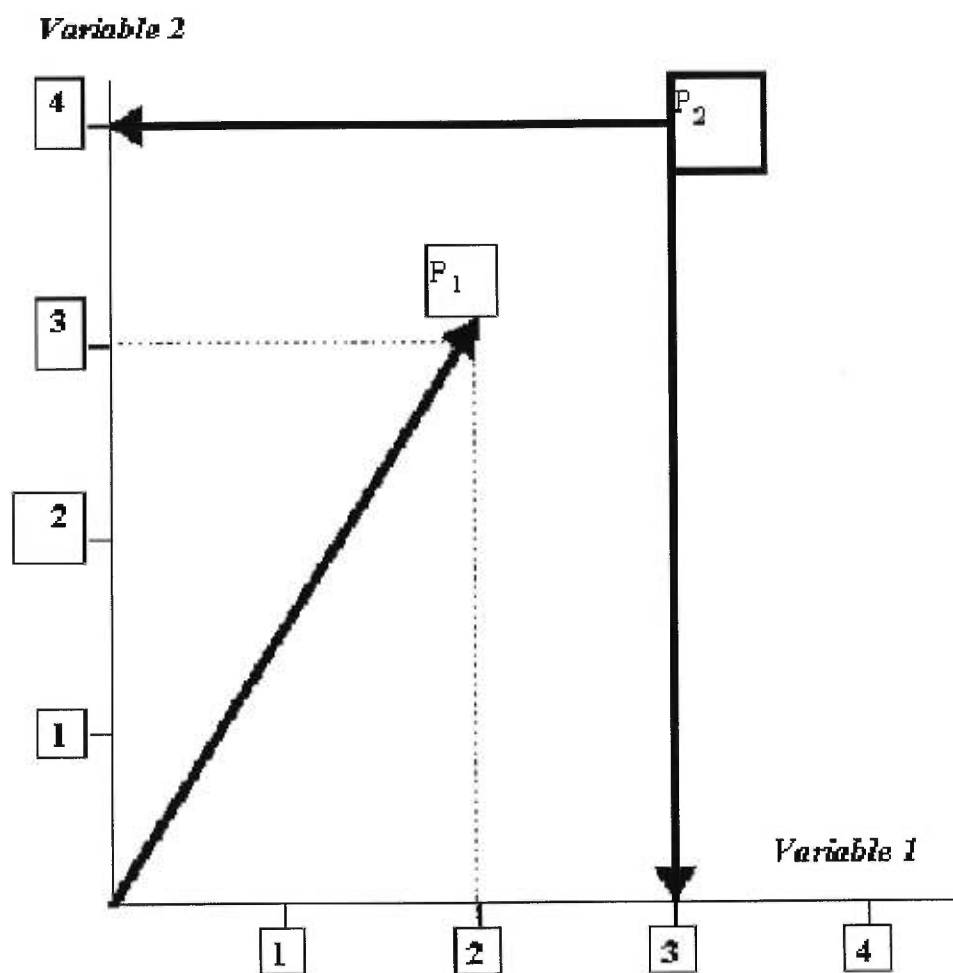


Figure 2.1: Interpolation du point P_1 de coordonnées (2,3) et prédiction des coordonnées (3,4) du point P_2

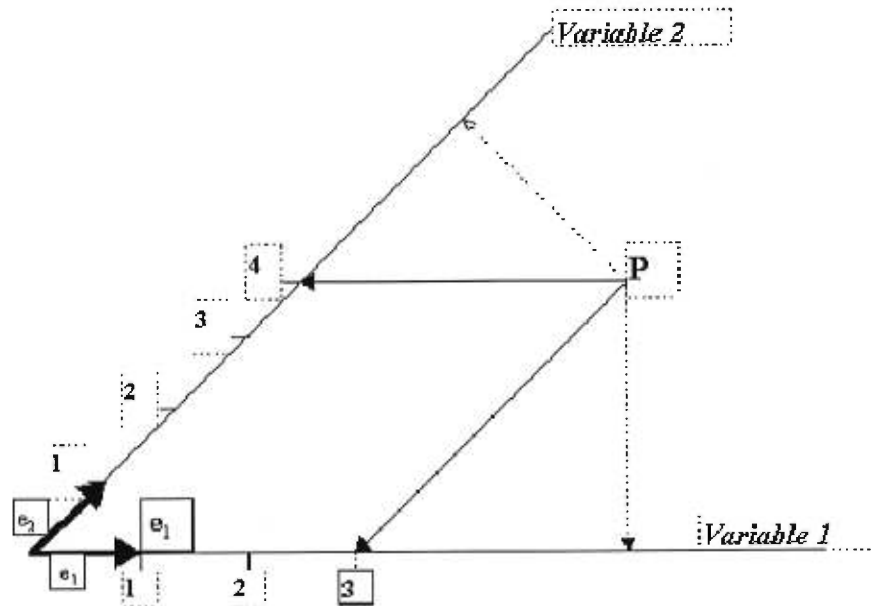


Figure 2.2: Interpolation du point $P(3, 4)$ par rapport à des axes obliques

2.2.2 Axes obliques

L'interpolation dans un plan avec systèmes d'axes obliques est basée sur la méthode des axes parallèles. Le point P dont les coordonnées sont (y_1, y_2) est obtenu en traçant les vecteurs $(y_1, 0)$ et $(0, y_2)$ et en complétant le parallélogramme. La prédiction des coordonnées du point P se fait en traçant des parallèles aux axes et en marquant par y_1 et y_2 les intersections avec ces axes (2.2)

On remarque tout de suite que l'utilisation de droites normales aux axes obliques prédit des valeurs différentes que celles interpolées avec la méthode du parallélogramme. Une échelle qui sert pour la prédiction est donc requise. On a d'ailleurs vu cette propriété de double échelle des axes biplot classique d'interpolation et de prédiction.

Il est évident que la généralisation à l'espace R_p de ces techniques est immédiate.

2.3 Résultats généraux

Dans cette section nous allons distinguer deux façons de faire une prédiction à partir d'un point $\hat{P} \in \mathcal{L}$.

Soit $\xi' = (\xi_1, \xi_2, \dots, \xi_p)$ un vecteur-individu représenté par le point \hat{P} dans le sous-espace \mathcal{L} . Rappelons que lorsque \mathcal{L} est le sous-espace engendré par l'A.C.P, l'interpolation géométrique du point $\hat{P} \in \mathcal{L}$ et la prédiction de $\hat{\xi}' = (\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_p)$ à partir de \hat{P} est faite avec les axes biplot $\beta_k \in \mathcal{L}$ (voir chapitre 1). La prédiction de $\hat{\xi}_k$ aurait pu se faire en érigeant des plans normaux à $\mathbf{e}_k \in R_p$. On voit ainsi qu'il y a deux façons de prédire $\hat{\xi}_k$: d'une part avec un axe biplot $\beta_k \in R_d$ et d'autre part avec $\mathbf{e}_k \in R_p$ ou plus généralement avec un sous-espace $M \subset R_p$.

Nous allons maintenant établir un lien entre la prédiction faite avec le sous-espace M et la prédiction faite avec le sous-espace \mathcal{L} . Remarquons que le sous-espace M contient l'information originale sur le vecteur individu, tandis que le sous-espace \mathcal{L} contient les points \hat{P}_i . Nous allons commencer par définir le terme back-projection.

Définition 2.3.1 Soient les matrices $\mathbf{L}_{(p \times r)}$ et $\mathbf{M}_{(p \times s)}$ dont les vecteurs colonnes engendrent les sous-espaces \mathcal{L} et M respectivement. Soit $\mathbf{x} \in M$, alors tous les points du l'hyperplan N orthogonal à M en \mathbf{x} vont prédire la même valeur \mathbf{x} (*méthode du plan normal*). L'hyperplan $N \cap \mathcal{L}$ contient tous les points de \mathcal{L} qui prédisent \mathbf{x} , et parmi tous ces points on prend \mathbf{z} le vecteur le plus près de \mathbf{x} qu'on appellera *back-projection de \mathbf{x} sur \mathcal{L}* .

La figure 2.3 donne une représentation géométrique de la back-projection $\mathbf{z} \in N \cap \mathcal{L}$ du point \mathbf{x} . Une expression algébrique de \mathbf{z} en fonction de $\mathbf{x} \in M$ sera développée au théorème 2.3.1

On discutera donc du lien entre deux façons de prédire le vecteur $\hat{\xi}$ à partir d'un point $\hat{P} \in \mathcal{L}$: la prédiction de $\hat{\xi}$ faite à l'aide d'un sous-espace $M \neq \mathcal{L}$ ou, avec les seules connaissances de \mathcal{L} (*back-projection*)

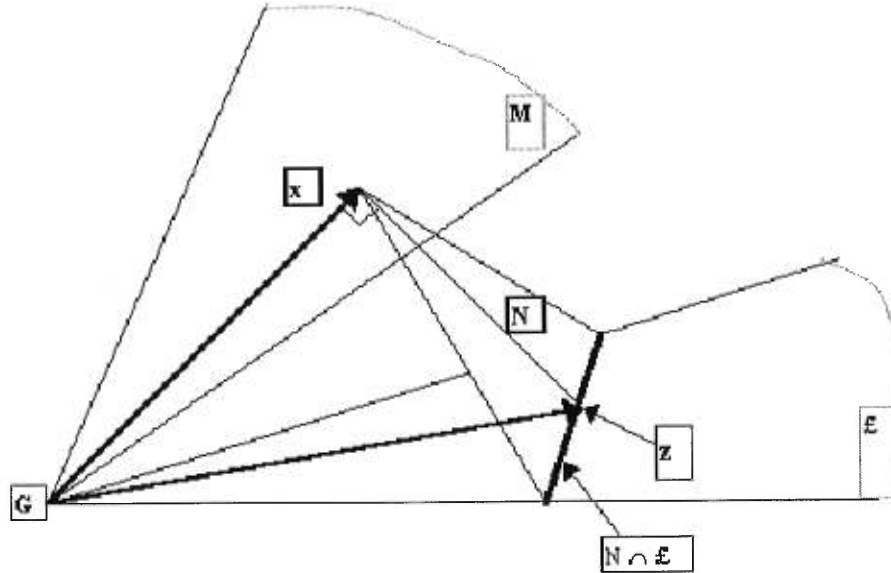


Figure 2.3: Représentation de la back-projection \mathbf{z}

ACP et plans normaux

La première application des principes de base énoncés à la section 1.3 est de donner un sens géométrique aux approximations $\mathbf{z}_{i[d]}$ trouvées au chapitre 1.

Soit le vecteur-individu \mathbf{z}_i représenté par le point P_i dans le système d'axes cartésien de R_p et représenté par le point $\hat{P}_i \in R_d$. Dans la notation du chapitre précédent, nous avons que \hat{P}_i avait comme coordonnée $\mathbf{a}_i^* = (a_{i1}^*, a_{i2}^*, \dots, a_{id}^*)$ dans la base \mathbf{V} . La valeur du point P_i sur l'axe \mathbf{e}_k est z_{ik} . Remarquons que le sous-espace M coïncide avec l'axe \mathbf{e}_k . Si l'on dirige un plan normal à \mathbf{e}_k passant par $\hat{P} \in \mathcal{L}$, alors l'intersection de ce plan avec \mathbf{e}_k est marquée \hat{z}_{ik} : La valeur \hat{z}_{ik} est la valeur prédite par \hat{P}_i selon les opérations élémentaires décrites à la section précédente. La différence résiduelle entre P_i et \hat{P}_i est:

$$r_i^2 = \sum_{k=1}^p (z_{ik} - \hat{z}_{ik})^2$$

La somme résiduelle pour les n points P_i est donnée par $\sum_{i=1}^p r_i^2 = \|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2$. En vertu du théorème d'Eckart-Young, celle-ci est minimisée lorsque $\widehat{\mathbf{Y}} = \mathbf{Y}\mathbf{V}_d$, c.à.d. lorsque $\widehat{\mathbf{Y}} = \mathbf{Y}_{[d]}$. On a donc que $\widehat{\mathbf{z}}_i = (\widehat{z}_{i1}, \widehat{z}_{i2}, \dots, \widehat{z}_{id})' = \mathbf{z}_{i[d]}$

Une nouvelle propriété du sous-espace R_d fournit par l'ACP est ainsi trouvée: au sens des moindres carrés, la méthode des plans normaux donne la meilleure récupération de l'information.

Remarquons par ailleurs, que l'information récupérée par la projection du point \widehat{P}_i sur les axes biplot étaient $\mathbf{z}_{i[d]}$. Nous verrons plus loin que le pied B de la projection de \widehat{P}_i sur l'axe biplot β_k (marqué par \widehat{z}_{ik}) est en fait la back-projection d'un point C sur ρe_k ; ce point C étant marqué de la valeur \widehat{y}_{ik} .

Nous allons maintenant démontrer un résultat général algébrique qui relie la prédiction faite avec le sous-espace M et la prédiction faite avec la back-projection dans \mathcal{L} . Sans perdre de généralité, nous allons supposer que les colonnes de \mathbf{L} sont orthogonales, ainsi que celles de \mathbf{M} c'est à dire $\mathbf{L}'\mathbf{L} = \mathbf{I}_r$ et $\mathbf{M}'\mathbf{M} = \mathbf{I}_s$.

Si $\mathbf{z} \in \mathcal{L}$ et $\mathbf{x} \in M$ alors on peut écrire:

$$\mathbf{K}\mathbf{K}'\mathbf{z} = \mathbf{0} \quad \text{et} \quad \mathbf{N}\mathbf{N}'\mathbf{x} = \mathbf{0} \quad (2.1)$$

où les colonnes de $\mathbf{K}_{(p \times (n-r))}$ et $\mathbf{N}_{(p \times (n-s))}$ sont respectivement les bases orthogonales des espaces complémentaires aux sous-espaces \mathcal{L} et N , c.à.d.: $\mathbf{K}\mathbf{K}' = \mathbf{I} - \mathbf{L}\mathbf{L}'$ et $\mathbf{N}\mathbf{N}' = \mathbf{I} - \mathbf{M}\mathbf{M}'$. Et puisque \mathbf{z} est le point le plus proche de \mathbf{x} alors ce point \mathbf{x} est la projection orthogonale de \mathbf{z} sur M c'est à dire :

$$\mathbf{x} = \mathbf{M}\mathbf{M}'\mathbf{z}$$

on a aussi $\mathbf{x} = \mathbf{M}\mathbf{M}'\mathbf{x}$. Ce qui entraîne que

$$\mathbf{M}\mathbf{M}'(\mathbf{z} - \mathbf{x}) = \mathbf{0} \quad (2.2)$$

Théorème 2.3.1 *Supposons que $s \leq r$ et $\mathcal{L} \cap M = \{O\}$, l'expression algébrique de la*

back-projection $\mathbf{z} \in \mathcal{L}$ du point $\mathbf{x} \in M$ est:

$$\mathbf{z} = (\mathbf{I} - \mathbf{N}\mathbf{N}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}')\mathbf{x} \quad (2.3)$$

où

$$\mathbf{z} = \mathbf{L}\mathbf{L}'(\mathbf{I} + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}')\mathbf{x} \quad (2.4)$$

Démonstration. On a par définition $\mathbf{z} \in N \cap \mathcal{L}$, et \mathbf{z} est le point le plus près de \mathbf{x} , c'est à dire la distance entre \mathbf{x} et \mathbf{z} est minimale. Le problème revient donc à solutionner:

$$\min(\mathbf{z} - \mathbf{x})'(\mathbf{z} - \mathbf{x})$$

sous les contraintes qui sont données par les équations 2.1 et 2.2:

$$\mathbf{M}\mathbf{M}'(\mathbf{z} - \mathbf{x}) = \mathbf{0} \quad \text{et} \quad \mathbf{K}\mathbf{K}'\mathbf{z} = \mathbf{0}$$

En appliquant la méthode des multiplicateurs de Lagrange, on obtient:

$$(\mathbf{z} - \mathbf{x})'(\mathbf{z} - \mathbf{x}) - \lambda\mathbf{M}\mathbf{M}'(\mathbf{z} - \mathbf{x}) - \mu\mathbf{K}\mathbf{K}'\mathbf{z} = \mathbf{0}$$

La dérivée de cette expression par rapport à \mathbf{z} est:

$$2\mathbf{z} - 2\mathbf{x} - \mathbf{M}\mathbf{M}'\mu - \mathbf{K}\mathbf{K}'\lambda = \mathbf{0}$$

donc

$$\mathbf{z} = \mathbf{x} + \mathbf{M}\mathbf{M}'\mu + \mathbf{K}\mathbf{K}'\lambda \quad (2.5)$$

et si on multiplie les deux termes par $\mathbf{N}\mathbf{N}'$ on obtient alors:

$$\mathbf{N}\mathbf{N}'(\mathbf{z} - \mathbf{x}) = \mathbf{N}\mathbf{N}'\mathbf{K}\mathbf{K}'\lambda \quad (2.6)$$

car $\mathbf{N}'\mathbf{M} = \mathbf{0}$. Puisque $\mathbf{I} = \mathbf{M}\mathbf{M}' + \mathbf{N}\mathbf{N}'$ on a:

$$\mathbf{N}\mathbf{N}'(\mathbf{z} - \mathbf{x}) = (\mathbf{z} - \mathbf{x}) - \mathbf{M}\mathbf{M}'(\mathbf{z} - \mathbf{x}) = (\mathbf{z} - \mathbf{x}) \quad (2.7)$$

La contrainte 2.2 entraîne la deuxième égalité. En substituant 2.7 dans 2.6:

$$(\mathbf{z} - \mathbf{x}) = \mathbf{N}\mathbf{N}'\mathbf{K}\mathbf{K}'\lambda \quad (2.8)$$

et en multipliant l'équation 2.8 par la matrice \mathbf{K}' on obtient:

$$\mathbf{K}'(\mathbf{z} - \mathbf{x}) = (\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})\mathbf{K}'\lambda$$

La matrice $(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})$ est inversible car $r \geq s$, on peut donc écrire:

$$\mathbf{K}'\lambda = (\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'(\mathbf{z} - \mathbf{x})$$

Or $\mathbf{K}'\mathbf{K} = \mathbf{I}$ et $\mathbf{K}\mathbf{K}'\mathbf{z} = \mathbf{0}$, donc $\mathbf{K}'\mathbf{z} = (\mathbf{K}'\mathbf{K})\mathbf{K}'\mathbf{z} = \mathbf{K}'(\mathbf{K}\mathbf{K}'\mathbf{z}) = \mathbf{0}$ alors :

$$\mathbf{K}'\lambda = -(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \quad (2.9)$$

afin d'éliminer μ de 2.5 on procède d'une façon analogue en multipliant 2.5 par $\mathbf{L}\mathbf{L}'$, alors :

$$\mathbf{L}\mathbf{L}'(\mathbf{z} - \mathbf{x}) = \mathbf{L}\mathbf{L}'\mathbf{M}\mathbf{M}'\mu \iff (\mathbf{z} - \mathbf{x}) - \mathbf{K}\mathbf{K}'(\mathbf{z} - \mathbf{x}) = \mathbf{L}\mathbf{L}'\mathbf{M}\mathbf{M}'\mu \quad (2.10)$$

on multiplie 2.10 par \mathbf{M}' , et puis que $\mathbf{K}\mathbf{K}'\mathbf{z} = \mathbf{0}$ alors:

$$\mathbf{M}'(\mathbf{z} - \mathbf{x}) + \mathbf{M}'\mathbf{K}\mathbf{K}'\mathbf{x} = \mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M}\mathbf{M}'\mu$$

Or on sait que $\mathbf{M}\mathbf{M}'(\mathbf{z} - \mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{M}'\mathbf{M}\mathbf{M}'(\mathbf{z} - \mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{M}'(\mathbf{z} - \mathbf{x}) = \mathbf{0}$ donc:

$$\mathbf{M}'\mathbf{K}\mathbf{K}'\mathbf{x} = \mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M}\mathbf{M}'\mu$$

$$\mathbf{M}'\mu = (\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}'\mathbf{x} \quad (2.11)$$

Si on substitue les deux expressions 2.9 et 2.11 dans 2.5 alors:

$$\begin{aligned} \mathbf{z} &= \mathbf{x} - \mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}'\mathbf{x} & (2.12) \\ &= \mathbf{x} - [\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1} - \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}]\mathbf{K}'\mathbf{x} \\ &= \mathbf{x} - [\mathbf{K} - \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K}](\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \\ &= \mathbf{x} - [\mathbf{I} - \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}'\mathbf{N}\mathbf{N}']\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \\ &= \mathbf{x} - [\mathbf{I} - \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'(\mathbf{I} - \mathbf{L}\mathbf{L}')\mathbf{N}\mathbf{N}']\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \\ &= \mathbf{x} - [\mathbf{I} + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{N}\mathbf{N}']\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} & \text{car } \mathbf{M}'\mathbf{N} = \mathbf{0} \\ &= \mathbf{x} - [\mathbf{I} + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{L}\mathbf{L}'(\mathbf{I} - \mathbf{M}\mathbf{M}')] \mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{x} - [\mathbf{I} + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{L}\mathbf{L}' - \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M}\mathbf{M}']\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \\
&= \mathbf{x} - [\mathbf{I} + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{L}\mathbf{L}' - \mathbf{M}\mathbf{M}']\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \\
&= \mathbf{x} - [\mathbf{N}\mathbf{N}' + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{L}\mathbf{L}']\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \\
&= \mathbf{x} - \mathbf{N}\mathbf{N}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{x} \quad \text{car } \mathbf{L}'\mathbf{K} = \mathbf{0}
\end{aligned}$$

On obtient ainsi la première expression 2.3:

$$\mathbf{z} = [\mathbf{I} - \mathbf{N}\mathbf{N}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}']\mathbf{x}$$

Nous allons maintenant démontrer l'expression 2.4. Tout d'abord nous allons établir une égalité dont nous aurons besoin. L'égalité des équations 2.3 et 2.12 donne:

$$-\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}' + \mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}' = -\mathbf{N}\mathbf{N}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'$$

En additionnant $\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'$ à chaque membre des équations on obtient:

$$\mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}' = (\mathbf{I} - \mathbf{N}\mathbf{N}')\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}'$$

d'où:

$$\mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}' = \mathbf{M}\mathbf{M}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}' \quad (2.13)$$

En substituant l'égalité $\mathbf{I} = \mathbf{L}\mathbf{L}' + \mathbf{K}\mathbf{K}'$ dans l'équation 2.13, on obtient:

$$\begin{aligned}
\mathbf{I} - \mathbf{N}\mathbf{N}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}' &= \mathbf{L}\mathbf{L}' + \mathbf{K}\mathbf{K}' - \mathbf{L}\mathbf{L}'\mathbf{N}\mathbf{N}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}' - \\
&\quad \mathbf{K}\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}' \\
&= \mathbf{L}\mathbf{L}' - \mathbf{L}\mathbf{L}'(\mathbf{I} - \mathbf{M}\mathbf{M}')\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}' + \mathbf{K}\mathbf{K}' - \mathbf{K}\mathbf{K}' \\
&= \mathbf{L}\mathbf{L}' - \mathbf{L}\mathbf{L}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1}\mathbf{K}' + \mathbf{L}\mathbf{L}'\mathbf{M}\mathbf{M}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1} \\
&= \mathbf{L}\mathbf{L}' + \mathbf{L}\mathbf{L}'\mathbf{M}\mathbf{M}'\mathbf{K}(\mathbf{K}'\mathbf{N}\mathbf{N}'\mathbf{K})^{-1} \quad \text{car } \mathbf{L}'\mathbf{K} = \mathbf{0} \\
&= \mathbf{L}\mathbf{L}' + \mathbf{L}\mathbf{L}'\mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}' \quad \text{d'après 2.13}
\end{aligned}$$

Ainsi la deuxième expression 2.4 est trouvée:

$$\mathbf{z} = [\mathbf{L}\mathbf{L}' + \mathbf{L}\mathbf{L}'\mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{K}\mathbf{K}']\mathbf{x}$$

L'équation 2.4 est définie pour un point \mathbf{x} de M tel que $\mathcal{L} \cap M = \{O\}$. Supposons que \mathbf{x} est un point quelconque de R_p , et $\mathcal{L} \cap M = \emptyset$, alors d'après la décomposition $\mathbf{x} = \mathbf{MM}'\mathbf{x} + \mathbf{NN}'\mathbf{x}$ qu'on interprète comme suit:

Le point \mathbf{x} se décompose comme une projection sur l'espace M défini par $\mathbf{MM}'\mathbf{x}$ puis une autre projection sur N , l'espace normal à M en \mathbf{x} . Cette dernière projection $\mathbf{q} = \mathbf{NN}'\mathbf{x}$ peut être vue comme une translation du sous-espace M par rapport à \mathcal{L} . Le point de M marqué par cette translation est la projection orthogonale de G sur M . La figure 2.3 illustre la géométrie de cette généralisation.

Remarque 2.3.1 Si $\mathbf{q} = \mathbf{0}$ alors l'intersection de \mathcal{L} et M n'est pas vide; mais l'inverse n'est pas vraie quand $\mathbf{q} \neq \mathbf{0}$.

Remarque 2.3.2 Si on pose $\mathbf{q} = \mathbf{NN}'\mathbf{x}$, alors $\mathbf{x} = \mathbf{MM}'\mathbf{x} + \mathbf{q}$. Et en substituant \mathbf{x} dans 2.4 on obtient alors:

$$\begin{aligned} \mathbf{z} = & \mathbf{LL}'\mathbf{MM}'\mathbf{x} + \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}(\mathbf{M}'\mathbf{KK}'\mathbf{M})\mathbf{M}'\mathbf{x} \\ & + \mathbf{LL}'\mathbf{q} + \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{KK}'\mathbf{q} \end{aligned}$$

Et en remplaçant \mathbf{KK}' par $\mathbf{I} - \mathbf{LL}'$ sans oublier que $\mathbf{M}'\mathbf{q} = \mathbf{0}$ alors:

$$\begin{aligned} \mathbf{z} = & \mathbf{LL}'\mathbf{MM}'\mathbf{x} + \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{MM}'\mathbf{x} - \\ & \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{LL}'\mathbf{MM}'\mathbf{x} + \mathbf{LL}'\mathbf{q} + \\ & \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{q} - \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{LL}'\mathbf{q} \\ = & \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{x} + (\mathbf{I} - \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}')\mathbf{LL}'\mathbf{q} \quad (2.14) \end{aligned}$$

L'équation 2.14 est une forme plus générale que celle donnée par les équations 2.3 et 2.4; elle n'est pas facile à manipuler mais elle nous donne un cas particulier très intéressant quand $\mathbf{q} = \mathbf{0}$. Dans ce cas les équations 2.3 et 2.4 se simplifient pour donner:

$$\mathbf{z} = \mathbf{LL}'\mathbf{M}(\mathbf{M}'\mathbf{LL}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{x} \quad (2.15)$$

Remarque 2.3.3 Pour que l'équation 2.15 soit vérifiée, il suffit que $\mathbf{L}'\mathbf{q}$ soit égale à zéro.

Un lien algébrique est ainsi établi à l'aide de l'équation 2.15 entre le point $\mathbf{x} \in M$ et sa back-projection $\mathbf{z} \in \mathcal{L}$. L'équation 2.15 indique que le point \mathbf{z} est une projection orthogonale du vecteur $\mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{x}$ sur le sous-espace \mathcal{L} engendré par les vecteurs colonnes de la matrice \mathbf{L} .

Cas particulier: axes biplot classique.

Les résultats du théorème 2.3.1 généralisent les résultats du biplot classique. En effet, si le sous-espace M est engendré par le vecteur unité \mathbf{e}_k et le sous-espace \mathcal{L} est engendré par les vecteurs \mathbf{V} , alors l'équation 2.15 s'écrit sous la forme suivante:

$$\mathbf{z} = \mathbf{V}\mathbf{V}'\mathbf{e}_k(\mathbf{e}_k'\mathbf{V}\mathbf{V}'\mathbf{e}_k)^{-1}\mathbf{e}_k'\mathbf{x}$$

mais $\mathbf{b}_k^{0'} = \mathbf{e}_k'\mathbf{V}$ d'après l'équation 1.3 du chapitre 1, et donc

$$\begin{aligned} \mathbf{z} &= \mathbf{V}\mathbf{b}_k^0(\mathbf{b}_k^{0'}\mathbf{b}_k^0)^{-1}\mathbf{e}_k'\mathbf{x} \\ &= \begin{pmatrix} \mathbf{V}\mathbf{b}_k^0 \\ \mathbf{b}_k^{0'}\mathbf{b}_k^0 \end{pmatrix} \mathbf{x}_k \quad \text{où } \mathbf{x}' = (x_1, \dots, x_p) \end{aligned} \quad (2.16)$$

Le point \mathbf{z} est la projection orthogonale de $\mathbf{M}(\mathbf{M}'\mathbf{L}\mathbf{L}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{x} = \mathbf{e}_k(\mathbf{e}_k'\mathbf{V}\mathbf{V}'\mathbf{e}_k)^{-1}\mathbf{e}_k'\mathbf{x} = \frac{\mathbf{e}_k\mathbf{x}_k}{\mathbf{b}_k^{0'}\mathbf{b}_k^0}$ sur \mathbf{V}_d . L'échelle de prédiction établie au chapitre 1 en projetant $\frac{\rho\mathbf{e}_k}{\mathbf{b}_k^{0'}\mathbf{b}_k^0}$ sur \mathbf{V}_d était établie de cette façon. L'équation 2.16 fait ressortir que les points \mathbf{z} sur l'axe biplot β_k dépendent linéairement des points $\rho\mathbf{e}_k$ sur l'axe de direction \mathbf{e}_k . Chaque point \mathbf{z} est en fait la back-projection du point $\rho\mathbf{e}_k$. Les résultats du théorème 2.3.1 généralisent donc les résultats obtenues pour le biplot classique.

Dans la section suivante, nous allons discuter de la multiplicité de l'axe biplot de prédiction et de l'unicité de l'axe biplot d'interpolation.

2.4 Axes biplot

Axes biplot de prédiction

La notion de back-projection attire notre attention sur le fait qu'il y a une infinité de points qui peuvent servir de prédiction. En effet, soient les points $\hat{y}_{ik}\mathbf{e}_k$, où \hat{y}_{ik} $i = 1, \dots, n$ sont les approximations aux observations faites sur la k ième variable, soient N_i les plans orthogonaux à \mathbf{e}_k en ces points alors tous les points dans $N_i \cap \mathcal{L}$ prédisent la valeur \hat{y}_{ik} ; On a bien sûr, que la back-projection est le point le plus près parmi tous ces points. Remarquons aussi que les plans orthogonaux N_i à \mathbf{e}_k en les points $\hat{y}_{ik}\mathbf{e}_k$ sont tous parallèles.

On peut tracer une droite β quelconque dans \mathcal{L} et choisir le point d'intersection $\beta \cap N_i \cap \mathcal{L}$ pour prédire \hat{y}_{ik} : ce point sera marqué \hat{y}_{ik} . Remarquons que l'angle d'intersection de β avec $N_i \cap \mathcal{L}$ est le même pour tout $i = 1, \dots, n$, car tous ces espaces sont parallèles. Connaissant l'angle d'intersection de la droite β avec $N_i \cap \mathcal{L}$, la valeur approximative \hat{y}_{ik} de la k ième variable d'un point $\hat{P}_i \in \mathcal{L}$ peut être prédite en projetant selon cet angle le point \hat{P}_i sur la droite β . Le marqueur au pied de cette projection est la valeur de prédiction cherchée. Ainsi toute droite peut servir d'axe biplot de la k ième variable. (voir figure 2.4)

Tout ce processus de prédiction doit être simplifié afin que son application soit facile à utiliser. Des restrictions doivent être imposées:

1. L'angle d'intersection entre β et les espaces $N_i \cap \mathcal{L}$ est choisi d'avance, et l'angle droit est l'angle le plus immédiat.
2. Pour bien localiser l'axe biplot β , on le fera passer par l'origine G et non par n'importe quel point arbitraire de \mathcal{L} .

Ainsi, l'angle que fait l'axe biplot β_k avec $N_i \cap \mathcal{L}$ est un angle de 90° , et, le pied de la projection orthogonale d'un point \hat{P}_i sur cet axe β_k prédit la valeur \hat{y}_{ik} . De plus

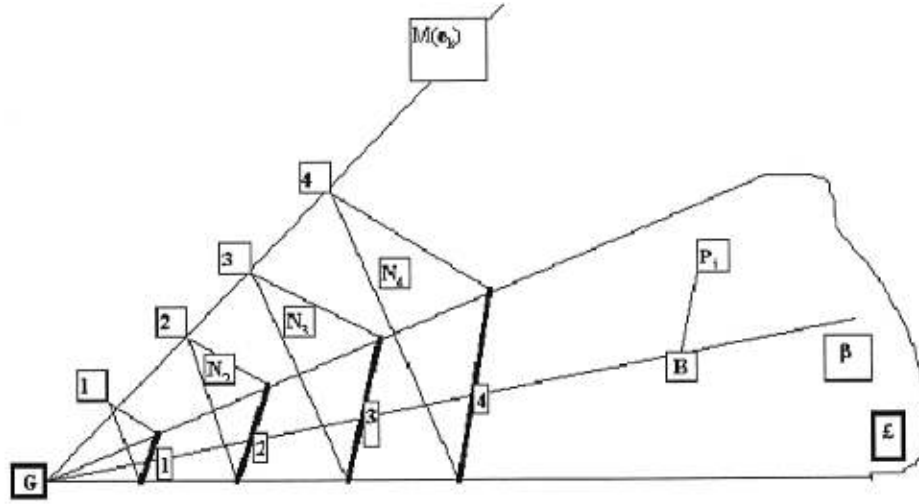


Figure 2.4: Les marqueurs de la droite β sont les back-projections des marqueurs de l'axe cartésien e_k .

cet axe biplot β_k passe par le centre de gravité G . Donc la projection orthogonale de G sur le plan $N_i \cap \mathcal{L}$ prédit aussi la valeur \hat{y}_{ik}

Sur la figure 2.4, M représente le k ième axe cartésien e_k . β est un axe arbitraire situé dans les espaces parallèles $N_i \cap \mathcal{L}$ passant par les points correspondant aux marqueurs de M . Pour prédire \hat{x}_{ik} la coordonnée de \hat{P}_i dans \mathcal{L} , il suffit de prendre l'espace parallèle à $N_i \cap \mathcal{L}$ qui va passer par l'axe β et qui va donner le marqueur \hat{x}_{ik} .

Remarque 2.4.1 La prédiction \hat{y}_{ik} est faite en utilisant l'axe biplot $\beta_k \in \mathcal{L}$. Cette prédiction serait la même que si l'on avait utilisé l'axe $e_k \in M$

Remarque 2.4.2 Les marqueurs sur l'axe biplot β_k obtenus en substituant ρe_k dans l'équation 2.15 dépendent linéairement des marqueurs sur e_k . Ainsi les marqueurs de distance unité sur β_k sont à égale distances.

Remarque 2.4.3 Puisque les axes biplot peuvent être n'importe quelle droite dans \mathcal{L} , alors une seule droite avec p échelles différentes vont suffire pour prédire les valeurs

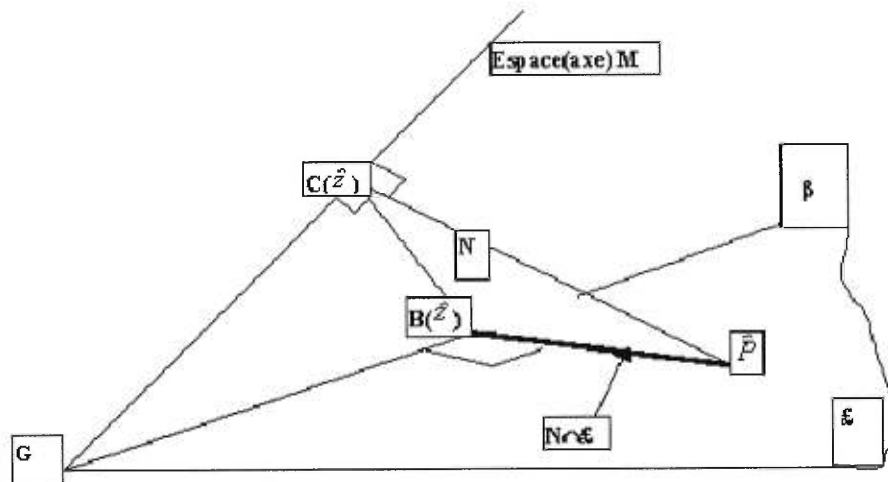


Figure 2.5: Géométrie de la back-projection

pour toutes les variables. Mais cette façon de procéder n'est pas très pratique.

Géométrie de la back-projection.

La figure (2.5) va bien expliquer géométriquement les résultats énoncés à propos de la back-projection:

La figure 2.5 explique géométriquement les résultats algébriques obtenus: notons $B(\hat{z})$, $C(\hat{z})$ les points B et C marqués respectivement par la valeur \hat{z} .

On a $B(\hat{z})$ est la projection orthogonale de \hat{P} sur β , donc $\hat{P}B \perp \beta$, on a aussi $B(\hat{z})$ est le back-projection de $C(\hat{z})$ dans M , donc $CB \perp M$. Et enfin $C\hat{P} \perp M$ puisque C est la projection orthogonale de \hat{P} sur M . Tous ces résultats illustrent que l'on peut faire toute la prédiction seulement à partir de l'espace \mathcal{L} et sans perdre d'information. En effet la prédiction C sur M à partir de \hat{P} dans \mathcal{L} est la même que la prédiction B sur β à partir de \hat{P} dans \mathcal{L} , puisque les points C et B sont tous les deux marqués par la valeur \hat{z} .

Axes biplot d'interpolation

Les axes biplot d'interpolation ne peuvent être autres que ceux obtenus par la projection de $\rho \mathbf{e}_k$ sur \mathcal{L} . En effet si $\xi' = (\xi_1, \dots, \xi_p)$ alors $\xi = \sum_{k=1}^p \xi_k \mathbf{e}_k$ et la projection de ξ sur \mathcal{L} est $\xi' \mathbf{L} \mathbf{L}' = \sum_{k=1}^p \xi_k \mathbf{e}'_k \mathbf{L} \mathbf{L}'$, où $\mathbf{e}'_k \mathbf{L} \mathbf{L}'$ est le vecteur unité sur la projection de $\rho \mathbf{e}_k$ sur \mathcal{L} . On a donc que le point $\hat{P} \in \mathcal{L}$ représentant le vecteur $\xi \in R_p$ s'écrit d'une façon unique comme une somme de p vecteurs, chacun de ces vecteurs étant sur un axe biplot d'interpolation. L'unité de l'axe d'interpolation en découle.

2.5 Conclusion

Nous avons ainsi discuté d'axes biplot de prédiction et d'interpolation dans un sous-espace quelconque $\mathcal{L} \subset R_p$. Notre attention a été attirée sur la multiplicité des axes biplot de prédiction et sur l'unicité des axes biplot d'interpolation.

Des résultats algébriques démontrent qu'il n'y a pas de perte d'information entre la prédiction fait d'une part avec un sous-espace M contenant de l'information originale et d'autre part avec un sous-espace \mathcal{L} qui ne contient que les points \hat{P}_i . Tous ces résultats étaient des généralisations des résultats du chapitre 1 et une introduction au chapitre 4.

Tous les résultats obtenus dans ce chapitre ne dépendent pas du choix de l'espace \mathcal{L} . Les axes biplot de prédiction et ceux d'interpolation sont définis pour n'importe quel sous-espace $\mathcal{L} \subset R_p$.

CHAPITRE III

POSITIONNEMENT MULTIDIMENSIONNEL

3.1 Introduction

Aux chapitres 1 et 2 nous avons discuté de la technique du biplot pour représenter un ensemble de données dans un sous-espace $\mathcal{L} \subset R_p$. Quatre aspects y étaient discutés: la représentation dans \mathcal{L} des vecteur-individus, le traçage dans \mathcal{L} des axes biplot, l'interpolation d'un nouveau point dans \mathcal{L} et la prédiction de nouvelles valeurs à partir d'un point de \mathcal{L} . Au chapitre 1, le sous-espace \mathcal{L} était le sous-espace R_d , $d < p$, engendré par les d premiers vecteur-colonnes de \mathbf{V} où $\mathbf{Y} = \mathbf{U}\mathbf{T}\mathbf{V}'$: ce sous-espace est en fait celui qui est engendré par les composantes principales. Au chapitre 2, nous avons généralisé la théorie du biplot classique appliquée à R_d au biplot appliqué dans $\mathcal{L} \subset R_p$, où \mathcal{L} est trouvé en optimisant d'autres critères que celui de l'ACP. Nous y avons aussi discuté brièvement du choix des autres sous-espaces $\mathcal{L} \subset R_p$. Dans ce chapitre nous allons étudier des techniques de représentation des individus par des points dans un espace euclidien quelconque \mathcal{L} qui n'est même pas inclus dans R_p .

Dans les chapitres 1 et 2, l'information utilisée pour représenter les individus par des points $P_i, i = 1..n$, étaient constituée de la matrice de données $\mathbf{Y}_{n \times p}$. Dans le chapitre 3, l'information utilisée sera constituée d'une matrice $\Delta_{n \times n}$ où chaque élément δ_{rs} est une mesure de proximité (dissimilarité ou similarité) entre le r ième et le s ième individu. La représentation des individus par des points ne sera pas basée sur la connaissance de la matrice $\mathbf{Y}_{n \times p}$ mais sur la connaissance de la matrice $\Delta_{n \times p}$. Les différentes

techniques qui traitent de ces représentations se nomment technique de positionnement multidimensionnel MDS (*Multidimensional scaling*)

Le but de la représentation de chaque individu par un point $Q_r \in \mathcal{L}, r = 1, \dots, n$, est de pouvoir visualiser les similarités (dissimilarités) des différents individus de l'échantillon par la proximité entre les points Q_r . La distance euclidienne d_{rs} entre les points Q_r et Q_s représentant respectivement les individus r et s , sera égale ou approximativement égale à la dissimilarité δ_{rs} . On pourra juger ainsi globalement des différentes dissimilarités entre les n individus par les proximités entre les n points Q_r . Il y aura, d'une part correspondance entre les individus i et les points Q_i $i = 1, \dots, n$, et d'autre part correspondance entre dissimilarités (similarités) δ_{rs} et distances euclidiennes d_{rs} . Évidemment, cette représentation n'est pas très utile lorsque la dimension de \mathcal{L} est supérieure à 3; nous discuterons tout de même d'une approximation à la représentation des individus dans un sous-espace \mathcal{L}_m , où la valeur de m est quelconque. Nous verrons que la dimension m nécessaire à la représentation de n individus avec la seule connaissance de Δ , est plus petite ou égale à $(n - 1)$.

Dans le but d'apporter des éclaircissements mathématiques, nous discuterons, dans la section 3.2, du positionnement multidimensionnel classique *CMDS*: la représentation des individus par ces points $Q_r \in \mathcal{L}$ est faite à partir d'une matrice de distances euclidiennes entre les individus. A la section 3.3, nous discuterons du positionnement multidimensionnel des individus par des points $Q_r \in \mathcal{L}$ fait à partir d'une matrice de similarités ou de dissimilarités.

3.2 Distances non pythagorienne

On sait que l'ACP utilise la distance euclidienne pour mesurer la proximité des individus. La technique du positionnement multidimensionnel a été développée en raison du fait qu'une distance euclidienne n'est pas toujours appropriée pour mesurer la distance entre les individus. Gower (1982) trouve que le terme distance euclidienne porte à confusion: il suggère que l'on emploie le terme « distance pythagorienne » afin de mettre

en relief la propriété importante de cette distance reliée au théorème de Pythagore.

Nous allons commencer par donner deux exemples de distance pythagorienne pour illustrer que la distance pythagorienne n'est pas toujours appropriée. Le premier exemple mettra en relief une motivation mathématique; quant au second exemple, l'interprétation d'une distance non pythagorienne sera jugée nécessaire dans un contexte des sciences de l'environnement.

Considérons un vecteur $\mathbf{y}'_i = (y_{i1}, \dots, y_{ip})$ où les $y_{ik}, k = 1, \dots, p$, représentent les proportions observées dans chacune des p catégories d'une variable multinômiale. La distance de Bhattacharyya δ_{ij} entre les individus i et j est définie comme suit:

$$\delta_{ij}^2 = \sum_{k=1}^p (y_{ik}^{1/2} - y_{jk}^{1/2})^2.$$

Cette distance peut être interprétée comme une distance asymptotique de Mahalanobis (voir Mardia, Kent et Bibby 1979, p379). Cette propriété est une *motivation mathématique* pour l'utilisation d'une autre distance que la distance de Pythagore. On peut évidemment avoir une représentation des individus à l'aide d'axes qui représentent la racine carrée de chaque variable. Remarquons alors que l'interprétation de la proximité des individus à l'aide de ces axes n'est pas très utile car ceux-ci ne représentent plus la valeur originale des variables.

Considérons un autre exemple en sciences de l'environnement. Supposons que chaque composante du vecteur \mathbf{y} indique une absence ou une présence d'un caractère. Supposons que l'on étudie deux caractères: la présence d'ailes pour voler et la présence de sabots. L'absence de ces deux caractères ne rend pas semblables un homme et un poisson; pourtant la proximité entre ceux-ci serait zéro si l'on utilisait une distance pythagorienne: ce qui est évidemment absurde. Sokal et Sneath (1973), Legendre et Legendre (1979), Mardia, Kent et Bibby (1979) discutent de différentes façons de calculer des proximités entre deux individus afin que cette proximité ait un sens dans le contexte d'application. On appelle cette proximité (non proximité), une similarité (dissimilarité)

entre deux individus.

Définition 3.2.1 Une matrice $\mathbf{A}_{n \times n}$ est dite une matrice de similarité si:

$$a_{ij} = a_{ji} \quad \text{et} \quad a_{ij} \leq a_{ii} \quad \text{pour tout } i, j$$

Définition 3.2.2 Une matrice $\mathbf{\Delta}_{n \times n}$ est une matrice de dissimilarité si

$$\begin{aligned} \delta_{ii} &= 0 && \text{pour tout } i \\ \delta_{ij} &\geq 0 && \text{pour tout } i, j \end{aligned}$$

Les cas d'assymétrie ($a_{ij} \neq a_{ji}$), peuvent être traités en posant $a_{ij}^* = \frac{a_{ij} + a_{ji}}{2}$; de même pour les dissimilarités.

Dans le cas des similarités (dissimilarités); les individus i et j sont d'avantages semblables selon que a_{ij} est grande (δ_{ij} est petite). Une distance non pythagorienne δ_{ij} peut être définie à partir de la similarité a_{ij} en utilisant les expressions $\delta_{ij} = 1 - a_{ij}$, si $a_{ij} \leq 1$, ou $\delta_{ij} = 1 - \log a_{ij}$ si $a_{ij} > 0$.

Ramsay(1982) fait remarquer qu'il est souhaitable que les mesures de similarités (dissimilarités) soient consistantes: si deux individus i et j sont semblables à un troisième individu k alors ces deux individus i et j devraient être semblables entre eux. Ce qui implique l'inégalité triangulaire $\delta_{ij} \leq \delta_{ik} + \delta_{jk}$. c'est à dire, si δ_{ik} et δ_{jk} sont petits alors δ_{ij} est petit. Nous prouverons que la propriété de consistance est une condition nécessaire et suffisante pour qu'une représentation des individus par des points soit possible.

Dans plusieurs domaines, seule la similarité ou dissimilarité entre deux individus est observée. Mardia, Kent et Bibby (1979) donnent l'exemple de l'écoute par des individus de signaux sonores morses des 10 chiffres de 0 à 9. La similarité a_{ij} est le comptage du nombre de fois que le signal i a été confondu avec le signal j , $i \neq j$; la

similarité a_{ii} est le nombre de fois qu'il a été bien compris. La matrice \mathbf{A} constituée des éléments a_{ij} est une matrice de similarité pour laquelle la proximité des signaux n'a pas été calculée à partir de variables.

Une représentation des chiffres de 0 à 9 par des points $Q_r \in \mathcal{L}$ $r = 1, \dots, 10$ qui refléterait la proximité des signaux morse nous permettrait de visualiser les signaux qui seraient fréquemment confondus (points voisins) de signaux peu souvent confondus (points distants).

La théorie de représentation des individus par des points Q_r dans un espace euclidien \mathcal{L}_m à partir d'une matrice de distances non pythagorienne ou à partir d'une matrice de similarité sera développée dans les sections suivantes. Le cas particulier de $m = 2$ ou 3 est un cas particulier intéressant car il permet la visualisation de la proximité des individus.

3.3 Solution géométrique

La représentation, à partir d'une matrice de dissimilarité Δ , des individus par des points Q_r dans un espace euclidien \mathcal{L} est exacte si les distances euclidiennes inter-points $d(Q_r, Q_s)$ notées d_{rs} calculées dans \mathcal{L} , sont égales aux dissimilarités δ_{rs} : remarquons que δ_{rs} n'est pas nécessairement pythagorienne. Gower (1982) utilise le terme " δ_{rs} is embeddable in an Euclidean space" si $\delta_{rs} = d_{rs}$. Nous utiliserons l'expression δ_{rs} est implantable dans un espace euclidien.

Nous verrons, dans le lemme suivant, une démonstration par construction géométrique d'une condition nécessaire et suffisante que les similarités doivent satisfaire pour que celles-ci soient euclidiennes implantables : la représentation exacte des n points sera possible dans un espace de dimension inférieure ou égale à $(n-1)$. Nous remercions monsieur André Plante qui nous a suggéré la démarche à suivre dans cette démonstration.

Lemme 3.3.1 *Soit une matrice $\Delta_{n \times n}$ constituée des dissimilarités δ_{rs} ; une représentation des n individus par des points $Q_r \in \mathcal{L}_{n-1}$, $r = 1, \dots, n$ telle que les distances d_{rs} entre*

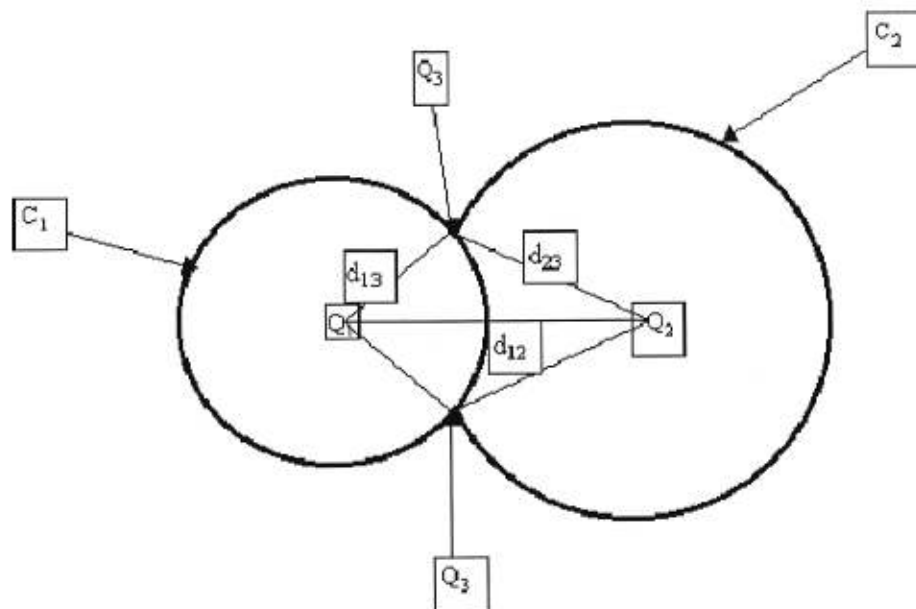


Figure 3.1: Représentation géométrique de trois valeurs d'un coefficient de similarité

les points Q_r et Q_s , sont égales aux dissimilarités δ_{rs} est possible si et seulement si les inégalités triangulaires $\delta_{rs} \leq \delta_{rk} + \delta_{ks}$ sont vérifiées pour tout triplet r, s, k .

Démonstration. Nous allons établir une démonstration par construction. Évidemment, un espace \mathcal{L}_1 contient des points Q_1 et Q_2 dont la distance est $d_{12} = \delta_{12}$. Ajoutons une deuxième dimension à \mathcal{L}_1 .

L'ajout du point Q_3 est faite en construisant deux cercles dans \mathcal{L}_2 . Soit $C_i(Q_i, \delta_{3i})$ les circonférences des cercles de centre Q_i et de rayon δ_{3i} , $i = 1, 2$. Afin que les distances entre le point Q_3 et chacun des points Q_1 et Q_2 soient égales à δ_{31} et δ_{32} respectivement, il faut que (voir figure 3.1):

$$Q_3 \in C_1(Q_1, \delta_{31}) \quad \text{et} \quad Q_3 \in C_2(Q_2, \delta_{32})$$

Une condition nécessaire et suffisante pour qu'une telle solution existe est que les

deux cercles aient une intersection non nulle. La distance entre les centres Q_1 et Q_2 doit donc être plus petite ou égale à la somme de la longueur des rayons: $\delta_{12} \leq \delta_{13} + \delta_{23}$.

L'ajout du point Q_4 est fait en construisant trois sphères $S_i(Q_i, \delta_{4i})$, de centre Q_i et de rayon δ_{4i} , $i = 1, 2, 3$. Afin que les distances entre le point Q_4 et les points Q_i soient δ_{4i} , il faut que:

$$Q_4 \in S_1(Q_1, \delta_{41}) , \quad Q_4 \in S_2(Q_2, \delta_{42}) , \quad Q_4 \in S_3(Q_3, \delta_{43})$$

Une condition nécessaire et suffisante pour qu'une telle solution existe est que les trois sphères aient simultanément une intersection non nulle. Celle-ci est non nulle si les inégalités triangulaires $\delta_{rs} \leq \delta_{rk} + \delta_{ks}$ sont satisfaites pour tout triplet $r, s, k = 1, 2, 3$.

L'ajout du point Q_r se fait en construisant $r - 1$ hypersphères dans \mathcal{L}_{r-2} : chacune étant centrée sur un point Q_i de \mathcal{L}_{r-2} et de rayon δ_{ri} , $i \leq r - 1$. Si nous continuons la construction d'une façon analogue pour tous les n points, nous obtenons un espace euclidien \mathcal{L}_{n-1} qui contient les n points $Q_r, r = 1, \dots, n$ tels que $d(Q_r, Q_s) = \delta_{rs}$. \square

Remarque 3.3.1 Il n'y a pas de solution unique. L'intersection des deux cercles tracés dans \mathcal{L}_2 peut avoir une, deux ou une infinité de solutions; de même pour l'intersection de $r - 1$ hypersphères.

Dans les sections suivantes, nous donnerons une solutions algébrique qui donnera les coordonnées des points Q_r en utilisant l'information donnée par *i*) une matrice Δ dont les éléments δ_{rs} sont pythagoriennes, *ii*) une matrice de similarité \mathbf{A} où $\mathbf{H}\mathbf{A}\mathbf{H}$ est définie positive, *iii*) une matrice Δ dont les éléments sont non pythagoriennes et $-\frac{1}{2}\mathbf{H}\Delta\mathbf{H}$ est défini positive.

3.4 Solution algébrique: Cas où δ_{rs} est pythagorienne

3.4.1 Introduction

Définition du problème

Dans cette section, nous allons donner une solution analytique au positionnement des points $Q_i, i = 1, \dots, n$ dans un espace euclidien \mathcal{L} à partir de la seule connaissance de la matrice Δ . Nous allons supposer que les éléments de cette matrice sont des distances pythagorienne c'est à dire $\delta_{rs}^2 = \sum_{k=1}^p (y_{rk} - y_{sk})^2$. Ce cas n'a d'intérêt que celui d'établir des liens mathématiques avec le cas plus général où δ_{rs} est un coefficient de similarité (dissimilarité). Avec cette optique mathématique, nous supposons donc que les distances ont été calculées à partir d'une matrice de données $\mathbf{Y}_{n \times p}$ et que le positionnement des points Q_r se fait sans la connaissance de la matrice $\mathbf{Y}_{(n \times p)}$. Cette matrice $\mathbf{Y}_{n \times p}$ n'est utilisée que dans le but de bien établir les relations mathématiques entre les différentes entités.

Chaque individu r est représenté par un point P_r dans R_p à l'aide du vecteur \mathbf{y}_r . Nous allons représenter chaque individu dans un autre espace \mathcal{L} , $\mathcal{L} \neq R_p$. Chaque individu sera représenté par $Q_r \in \mathcal{L} \neq R_p$, ($r = 1, \dots, n$) de telle sorte que les distances euclidiennes d_{rs} sont calculées avec les coordonnées des points Q_r et Q_s . Ces distances d_{rs} seront égales aux distances δ_{rs} calculées avec les coordonnées $y_{rs} \in R_p$. Remarquons que chaque axe de R_p représente une variable, tandis que les axes de l'espace \mathcal{L} sont des axes quelconques. Ces derniers n'ont aucun lien avec les axes de R_p : ils n'ont aucune signification. Ce sont des axes « bidons »

Les coordonnées des points Q_r de \mathcal{L}_m seront trouvées analytiquement par opposition à la solution géométrique de la section précédente. Ces coordonnées seront notées $(x_{r1}, x_{r2}, \dots, x_{rm})$ $r = 1, \dots, n$ et entreposées dans la matrice $\mathbf{X}_{n \times m}$.

Non unicité de la solution

Nous savons que la solution n'est pas unique. En effet, deux configurations différentes peuvent avoir les mêmes distances inter-points : dans la géométrie euclidienne, ceci peut se produire lorsqu'une configuration peut s'obtenir d'une autre par translation, rotation ou réflexion.

Soient $\tilde{\mathbf{A}}$ une matrice de rotation, et $\tilde{\mathbf{C}}$ une matrice de translation. Les coordonnées données par la r -ième ligne de la matrice $\tilde{\mathbf{A}}\mathbf{X} + \tilde{\mathbf{C}}$ sont les coordonnées d'un autre point Q_r^* : les distances entre les points Q_r^* auront les mêmes distances inter-points que celles entre les points Q_r .

Problème de translation

Pour éliminer le problème de translation on fixera l'origine au centre de gravité des points Q_r :

$$\sum_{r=1}^n x_{rj} = 0, \quad \text{pour tout } j = 1, \dots, m.$$

Le nuage de points de la configuration sera centré sur le centre de gravité du nuage de points : ainsi qu'au centre de gravité G du nuage de points de R_p correspond l'origine O de \mathcal{L}_m .

Nous discuterons du problème générale de la non unicité de la solution à la section 3.6

3.4.2 Solution algébrique

La connaissance des coordonnées des n points P_i dans l'espace euclidien R_p , mène d'une façon immédiate au calcul des distances pythagoriques inter-points. Les carrés de ces distances sont entreposés dans la matrice de distance $\mathbf{\Delta}_{(n \times n)}$. Nous sommes intéressés au cas inverse: retrouver une matrice $\mathbf{X}_{(n \times p)}$ des coordonnées des points Q_i à partir de la matrice $\mathbf{\Delta}_{(n \times n)}$. Nous allons en premier lieu, établir un lien entre la matrice

de distance Δ et la matrice $\mathbf{Q} = \mathbf{Y}_{(n \times p)} \mathbf{Y}'_{(p \times n)}$.

Lemme 3.4.1 Soit $\mathbf{Q} = \mathbf{Y}_{n \times p} \mathbf{Y}'_{p \times n}$, où $\mathbf{Y}'\mathbf{1} = \mathbf{0}$. Cette matrice \mathbf{Q} peut être calculée directement à partir de la matrice de distance Δ en utilisant l'expression suivante :

$$\mathbf{Q} = -\frac{1}{2} \mathbf{H} \Delta \mathbf{H} \quad (3.1)$$

où $\mathbf{H} = \mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}'$.

Démonstration. Soit $\mathbf{Q} = \mathbf{Y}_{(n \times p)} \mathbf{Y}'_{(p \times n)}$. On a alors $q_{rs} = \sum_{j=1}^p y_{rj} y_{sj}$. Le carré de la distance euclidienne δ_{rs}^2 entre deux points P_r et P_s dans R_p est :

$$\begin{aligned} \delta_{rs}^2 &= \sum_{j=1}^p (y_{rj} - y_{sj})^2 \\ &= \sum_{j=1}^p y_{rj}^2 + \sum_{j=1}^p y_{sj}^2 - 2 \sum_{j=1}^p y_{rj} y_{sj} \\ &= q_{rr} + q_{ss} - 2q_{rs}. \end{aligned} \quad (3.2)$$

Or la matrice \mathbf{Q} est centrée selon les lignes et les colonnes :

$$\sum_{r=1}^n q_{rs} = \sum_{s=1}^n q_{rs} = 0.$$

En effet,

$$\sum_{r=1}^n q_{rs} = \sum_{r=1}^n \sum_{j=1}^p y_{rj} y_{sj} = \sum_{j=1}^p y_{sj} \sum_{r=1}^n y_{rj} = 0,$$

car le nuage de points est centré à l'origine. De même,

$$\sum_{s=1}^n q_{rs} = \sum_{s=1}^n \sum_{j=1}^p y_{rj} y_{sj} = \sum_{j=1}^p y_{rj} \sum_{s=1}^n y_{sj} = 0,$$

pour la même raison. Donc

$$\sum_{r=1}^n \delta_{rs}^2 = \text{tr}(\mathbf{Q}) + nq_{ss}$$

$$\sum_{s=1}^n \delta_{rs}^2 = \text{tr}(\mathbf{Q}) + nq_{rr}$$

$$\sum_{r=1}^n \sum_{s=1}^n \delta_{rs}^2 = (2n) \text{tr}(\mathbf{Q})$$

où $\text{tr}(\mathbf{Q}) = \sum_{r=1}^n q_{rr}$ est la trace de la matrice \mathbf{Q} . Or, d'après l'équation 3.2, on a :

$$\begin{aligned} -2q_{rs} &= \delta_{rs}^2 - q_{rr} - q_{ss} \\ &= \delta_{rs}^2 - \left(\delta_r^2 - \frac{\text{tr}(\mathbf{Q})}{n}\right) - \left(\delta_s^2 - \frac{\text{tr}(\mathbf{Q})}{n}\right) \\ &= \delta_{rs}^2 - \delta_r^2 - \delta_s^2 + \delta_{..}^2 \end{aligned}$$

où le point en indice représente la moyenne des éléments de la ligne ou de la colonne selon la position du point. On a donc:

$$q_{rs} = -\frac{1}{2}(\delta_{rs}^2 - \delta_r^2 - \delta_s^2 + \delta_{..}^2)$$

qui s'écrit matriciellement $\mathbf{Q} = -\frac{1}{2}\mathbf{H}\Delta\mathbf{H}$. \square

Théorème 3.4.1 *Soit une matrice $\Delta_{n \times n}$ constituée des distances pythagoriennes δ_{rs} entre les points-individus P_r et P_s de R_p . Une représentation des n individus par des points $Q_r \in \mathcal{L}$, $\mathcal{L} \neq R_p$ est possible. Les coordonnées des points Q_r , $r = 1, \dots, n$, sont données par les vecteurs-lignes de la matrice $\mathbf{X} = \mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}}$ où les matrices \mathbf{V} et $\mathbf{\Gamma}$ sont obtenues par la décomposition en valeurs spectrales de la matrice $\mathbf{Q} = -\frac{1}{2}\mathbf{H}\Delta\mathbf{H}$.*

Démonstration. Nous avons, en vertu du lemme 3.4.1 que

$$\mathbf{Q} = -\frac{1}{2}\mathbf{H}\Delta\mathbf{H} = \mathbf{Y}\mathbf{Y}'.$$

La matrice \mathbf{Q} étant symétrique de rang p , ayant p valeurs propres strictement positives et $(n - p)$ valeurs propres nulles. La décomposition en valeurs spectrales de \mathbf{Q} est:

$$\mathbf{Q} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}'$$

où \mathbf{V} est constituée des vecteurs propres $(\mathbf{v}_1, \dots, \mathbf{v}_p)$ et les éléments λ_i de la matrice diagonale $\mathbf{\Gamma}$ sont triés selon l'ordre décroissant ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$). Posons $\mathbf{X} = \mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}}$ où

$$\mathbf{X} = \mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}} = \begin{pmatrix} \gamma_1 v_{11} & \gamma_2 v_{21} & \cdots & \gamma_p v_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_1 v_{1n} & \gamma_2 v_{2n} & \cdots & \gamma_p v_{pn} \end{pmatrix} \begin{array}{l} \longrightarrow Q_1 \\ \vdots \\ \longrightarrow Q_n \end{array}$$

avec $\gamma_r = \lambda_r^{\frac{1}{2}}$ pour tout r . A chaque ligne r de la matrice \mathbf{X} correspond les coordonnées d'un point Q_r . La distance pythagorienne d_{rs} entre les points Q_r et Q_s est égale à la distance δ_{rs} entre les points P_r et P_s . En effet:

$$\begin{aligned} d_{rs}^2 &= \sum_{i=1}^p (x_{ri} - x_{si})^2 \\ &= \sum_{i=1}^p x_{ri}^2 + \sum_{i=1}^p x_{si}^2 - 2 \sum_{i=1}^p x_{ri} x_{si} \end{aligned}$$

Or $\mathbf{Q} = \mathbf{V}\mathbf{T}\mathbf{V}' = \mathbf{X}\mathbf{X}'$, donc $\sum_{i=1}^p x_{ri}^2 = q_{rr}$ et $\sum_{i=1}^p x_{ri} x_{si} = q_{rs}$. Ainsi, on a:

$$d_{rs}^2 = q_{rr} + q_{ss} - 2q_{rs} = \delta_{rs}^2$$

Cette dernière égalité est obtenue par l'équation 3.2.

Nous avons donc trouvé algébriquement les coordonnées de n points $Q_i \in \mathcal{L}_p$ tels que les distances d_{rs} entre les points Q_r et Q_s sont égales aux distances δ_{rs} entre les points P_i de R_p . Ceci a été fait avec la seule connaissance de la matrice $\mathbf{\Delta}_{n \times n}$. \square

3.5 Solution algébrique: cas général

Soit A une matrice de similarité dont les éléments a_{ij} représentent les similarités entre le i ème individu et le j ème individu. On peut définir une dissimilarité δ_{ij}^2 entre les individus i et j en posant $\delta_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}$. En effet δ_{ij}^2 est grande si la similarité est petite, et la dissimilarité δ_{ij}^2 est petite si la similarité est grande. En effet, a_{ii} correspond à un maximum de similarité: posons $a_{ii} = Max$. On a alors

$$\delta_{ij}^2 = 2(Max - a_{ij}).$$

Si deux individus sont très semblables alors leur similarité est proche du Max et donc leur distance est près de 0. Si la similarité entre deux individus va en décroissant alors leur distance d_{ij}^2 va en s'accroissant. Si a_{ij} est la corrélation entre deux individus, alors $a_{ii} = 1$ et l'on obtient: $d_{ij}^2 = 2(1 - a_{ij})$

Théorème 3.5.1 Soient \mathbf{A} une matrice de similarité entre n individus, et m le rang de la matrice \mathbf{A} ; Soit \mathcal{L}_m un espace vectoriel réel. La matrice $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ est symétrique définie positive si et seulement si une représentation des n individus par des points $Q_i \in \mathcal{L}_m$, ($i = 1, \dots, n$) existe telles que les distances pythagoriennes d_{ij} entre les points Q_i et Q_j satisfont $d_{ij}^2 = \delta_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}$.

Démonstration. Supposons qu'il existe une représentation des n individus par n points $Q_i \in \mathcal{L}_m$, $i = 1, \dots, n$. Soit $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$ le vecteur des coordonnées du point Q_i entreposés dans la i -ième ligne de la matrice $\mathbf{X}_{n \times m}$. Montrons que $\mathbf{H}\mathbf{A}\mathbf{H}$ est symétrique définie positive. Par hypothèse, on a

$$d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}; \quad (3.3)$$

donc

$$d_{i.}^2 = a_{ii} + \frac{\sum_{j=1}^n a_{jj}}{n} - 2a_{i.} \quad (3.4)$$

$$d_{.j}^2 = a_{jj} + \frac{\sum_{i=1}^n a_{ii}}{n} - 2a_{.j} \quad (3.5)$$

$$d_{..}^2 = 2 \frac{\sum_{i=1}^n a_{ii}}{n} - 2a_{..} \quad (3.6)$$

Une manipulation algébrique sur les équations (3.3) – (3.4) – (3.5) + (3.6), donne l'équation suivante:

$$-\frac{1}{2}(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2) = a_{ii} - a_{i.} - a_{.j} + a_{..}$$

On a donc

$$-\frac{1}{2}\mathbf{H}\mathbf{\Delta}\mathbf{H} = \mathbf{H}\mathbf{A}\mathbf{H}$$

où les éléments $\delta_{rs}^2 = d_{rs}^2$. Etant donné que les distances d_{rs} entre les points Q_r et Q_s sont pythagoriennes, on a $-\frac{1}{2}\mathbf{H}\mathbf{\Delta}\mathbf{H}$ est symétrique définie positive en vertu du lemme 3.4.1. L'égalité des matrices $-\frac{1}{2}\mathbf{H}\mathbf{\Delta}\mathbf{H} = \mathbf{H}\mathbf{A}\mathbf{H}$ entraîne que $\mathbf{H}\mathbf{A}\mathbf{H}$ est symétrique définie positive.

Réciproquement, supposons que $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ est symétrique définie positive. La décomposition de \mathbf{B} en valeurs spectrales est

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}'$$

où les vecteurs colonnes de \mathbf{V} sont les vecteurs propres de \mathbf{B} et les éléments de la diagonale contiennent les valeurs propres triées dans l'ordre décroissant $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_m \geq 0$. Posons

$$\mathbf{X} = \mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}} = \begin{pmatrix} \gamma_1 v_{11} & \gamma_2 v_{21} & \cdots & \gamma_p v_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_1 v_{1n} & \gamma_2 v_{2n} & \cdots & \gamma_p v_{pn} \end{pmatrix} \begin{matrix} \longrightarrow Q_1 \\ \vdots \\ \longrightarrow Q_n \end{matrix}$$

Soient les points Q_i dont les coordonnées sont données par la i ème ligne de \mathbf{X} . La distance pythagorienne d_{ij}^2 est:

$$\begin{aligned} d_{ij}^2 &= d(Q_r, Q_s) = \sum_{k=1}^m (x_{ik} - x_{jk})^2 \\ &= \sum_{k=1}^m x_{ik}^2 + \sum_{k=1}^m x_{jk}^2 - 2 \sum_{k=1}^m x_{ik} x_{jk}. \end{aligned} \quad (3.7)$$

Or le i ème élément de la diagonale de la matrice $\mathbf{H}\mathbf{A}\mathbf{H} = \mathbf{X}\mathbf{X}'$ est:

$$\sum_{k=1}^m x_{ik}^2 = a_{ii} - 2a_{i.} + a_{..} \quad (3.8)$$

Le (i, j) -ième élément de la matrice $\mathbf{X}\mathbf{X}'$ est

$$\sum_{k=1}^m x_{ik} x_{jk} = a_{ij} - a_{i.} - a_{.j} + a_{..} \quad (3.9)$$

En substituant les équations 3.8 et 3.9 dans 3.7 on obtient:

$$d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij} \quad (3.10)$$

Si \mathbf{B} est symétrique définie positive, une configuration de n points $Q_i, i = 1, \dots, n$, est donc possible, de plus la distance entre deux points Q_i et Q_j est donnée par 3.10. \square

Remarque 3.5.1 On a $\mathbf{B}\mathbf{1} = \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{1} = \mathbf{0}$; donc \mathbf{B} admet une valeur propre nulle. L'espace euclidien de représentation des n individus est au plus de dimension $(n - 1)$.

Corollaire 3.5.1 *Soit Δ une matrice dont les éléments sont les dissimilarités δ_{ij}^2 . Soit \mathbf{A} la matrice de similarité où $a_{ij} = -\frac{1}{2}\delta_{ij}^2$ et $a_{ii} = 0$. En appliquant le théorème 3.5.1 à la matrice \mathbf{A} on obtient une condition nécessaire et suffisante pour avoir une représentation des n individus par n points Q_i où la distance inter-point est $d_{ij}^2 = \delta_{ij}^2$ (en vertu de l'équation 3.10)*

3.6 Non unicité de la solution

Dans la section 3.4, nous avons parlé brièvement de la non unicité de la solution à la représentation de n points dans un espace euclidien à partir d'une matrice de distances euclidiennes implantables. Dans cette section, nous allons discuter de ce problème de la façon dont Gower (1982) l'a traité.

Si la matrice Δ est euclidienne implantable, alors il existe n points Q_1, Q_2, \dots, Q_n dont les coordonnées sont entreposées dans les lignes de la matrice $\mathbf{X}_{n \times m}$. Le lemme 3.4.1 donne une relation entre les matrices $\mathbf{Q} = \mathbf{X}\mathbf{X}'$ et Δ : $\mathbf{Q} = \mathbf{X}\mathbf{X}' = -\frac{1}{2}\mathbf{H}\Delta\mathbf{H}$. De plus l'équation 3.2 donne $\delta_{ij}^2 = q_{rr} + q_{ss} - 2q_{rs}$.

Les points Q_1, Q_2, \dots, Q_n ne sont pas uniques. Soient Q'_1, Q'_2, \dots, Q'_n , n autres points qui sont une autre solution. Soit $\mathbf{Z}_{n \times m}$ la matrice de coordonnées de ces points. Posons $\mathbf{F}_{n \times n} = \mathbf{Z}\mathbf{Z}'$. Nous allons maintenant établir une relation \mathbf{F} et Δ . Remarquons tout d'abord que

$$\delta_{ij}^2 = d_{ij}^2 = \sum_{k=1}^m (z_{ik} - z_{jk})^2 = f_{ii} + f_{jj} - 2f_{ij};$$

cette équation est l'analogie de l'équation 3.2.

Sans perte de généralité, on peut écrire $\mathbf{F} = -\frac{1}{2}\Delta + \mathbf{T}$, où \mathbf{T} est une matrice quelconque de dimension $(n \times n)$. On a alors:

$$\begin{aligned} \delta_{ij}^2 &= f_{ii} + f_{jj} - 2f_{ij} \\ &= \left(-\frac{1}{2}\delta_{ii}^2 + t_{ii}\right) + \left(-\frac{1}{2}\delta_{jj}^2 + t_{jj}\right) - 2\left(-\frac{1}{2}\delta_{ij}^2 - t_{ij}\right) \\ &= \delta_{ij}^2 + t_{ii} + t_{jj} - 2t_{ij} \quad \text{car } \delta_{ii}^2 = \delta_{jj}^2 = 0. \end{aligned}$$

Donc $t_{ij} = -\frac{1}{2}(t_{ii} + t_{jj})$ (ou $\mathbf{T} = \mathbf{t}\mathbf{1}' + \mathbf{1}\mathbf{t}'$) avec $t_i = -\frac{1}{2}t_{ii}$ et $\mathbf{1}$ est le vecteur unité. Ainsi toute matrice $\mathbf{F} = \mathbf{Z}\mathbf{Z}'$ où les lignes de la matrice \mathbf{Z} sont les coordonnées des points Q'_i avec distances inter-points entreposés dans la matrice Δ est de la forme:

$$\mathbf{F} = -\frac{1}{2}\Delta + \mathbf{t}\mathbf{1}' + \mathbf{1}\mathbf{t}' \quad (3.11)$$

Inversement, soit \mathbf{t} un vecteur quelconque et $\mathbf{F} = -\frac{1}{2}\Delta + \mathbf{t}\mathbf{1}' + \mathbf{1}\mathbf{t}'$, alors on a :

$$\mathbf{H}\mathbf{F}\mathbf{H} = \mathbf{H}\mathbf{A}\mathbf{H} + \mathbf{H}\mathbf{t}\mathbf{1}'\mathbf{H} + \mathbf{H}\mathbf{1}\mathbf{t}'\mathbf{H} = \mathbf{H}\mathbf{A}\mathbf{H}.$$

car $\mathbf{H}\Delta\mathbf{H} = \mathbf{H}\mathbf{A}\mathbf{H}$. Et d'après le théorème 3.5.1 on trouve une solution Q_1, Q_2, \dots, Q_n telle que

$$d_{ij}^2 = \delta_{ij}^2 = f_{ii} + f_{jj} - 2f_{ij} = a_{ii} + a_{jj} - 2a_{ij}.$$

Gower (1982) a démontré que pour chaque vecteur \mathbf{t} il existe un vecteur \mathbf{s} telle que la matrice \mathbf{F} donnée par l'équation 3.11, admet une forme multiplicative

$$\mathbf{F}^* = -\frac{1}{2}(\mathbf{I} - \mathbf{1}\mathbf{s}')\Delta(\mathbf{I} - \mathbf{s}\mathbf{1}') \quad (3.12)$$

où $\mathbf{s}'\mathbf{1} = 1$ et $\Delta\mathbf{s} \neq \mathbf{0}$.

En effet, supposons que les coordonnées de \mathbf{Z} sont translatées et que les nouvelles coordonnées sont $\mathbf{Z} + \mathbf{1}\mathbf{m}'$. Posons $\mathbf{m} = -\mathbf{Z}'\mathbf{s}$ où $\mathbf{s}'\mathbf{t} = 0$, et $\mathbf{s}'\mathbf{1} = 1$ alors $(\mathbf{Z} + \mathbf{1}\mathbf{m}') = (\mathbf{I} - \mathbf{1}\mathbf{s}')\mathbf{Z}$ et correspondent à \mathbf{F} . Nous obtenons :

$$\begin{aligned} \mathbf{F}^* &= (\mathbf{I} - \mathbf{1}\mathbf{s}')\mathbf{Z}\mathbf{Z}'(\mathbf{I} - \mathbf{s}\mathbf{1}') \\ &= (\mathbf{I} - \mathbf{1}\mathbf{s}')\left(-\frac{1}{2}\Delta + \mathbf{t}\mathbf{1}' + \mathbf{1}\mathbf{t}'\right)(\mathbf{I} - \mathbf{s}\mathbf{1}') \\ &= -\frac{1}{2}(\mathbf{I} - \mathbf{1}\mathbf{s}')\Delta(\mathbf{I} - \mathbf{s}\mathbf{1}') \end{aligned}$$

Ainsi à chaque vecteur $\mathbf{t}_{n \times 1}$ de l'équation 3.11, nous trouvons une équation 3.12 qui lui est équivalente.

Remarque 3.6.1 Pour tout \mathbf{s} qui vérifie 3.12 on a:

$$\begin{aligned} \mathbf{s}'\mathbf{F}\mathbf{s} &= \mathbf{s}'(\mathbf{I} - \mathbf{1}\mathbf{s}')\Delta(\mathbf{I} - \mathbf{s}\mathbf{1}') \\ &= (\mathbf{s}' - \mathbf{s}')\Delta(\mathbf{I} - \mathbf{s}\mathbf{1}') = 0 \quad \text{car } \mathbf{s}'\mathbf{1} = 1 \\ &= \mathbf{s}'\mathbf{Z}\mathbf{Z}'\mathbf{s} \quad \text{car } \mathbf{F} = \mathbf{Z}\mathbf{Z}' \end{aligned}$$

Donc $\mathbf{s}'\mathbf{Z} = \mathbf{0}$. Le vecteur \mathbf{s} détermine l'origine de la configuration du nuage de points. Gower (1982), a présenté plusieurs choix de \mathbf{s} qui sont intéressants. Nous allons présenter deux cas particuliers qui sont très utiles.

Remarque 3.6.2 Si on prend $\mathbf{s} = \mathbf{e}_i$ où \mathbf{e}_i est le i ème vecteur unité, alors Q_i est situé à l'origine du nuage de points. Il est à noter que $\mathbf{e}_i'\mathbf{1} = 1$, et $\Delta\mathbf{e}_i \neq 0$ sauf si la i -ème colonne de Δ est nulle, ce qui est un cas de dégénérescence.

Le second cas est le plus utilisé . Soit $\mathbf{s} = \frac{1}{n}\mathbf{1}$ où n est le nombre de points. L'origine de la représentation graphique est alors au centre de gravité des n points.

3.6.1 Approximation à la représentation

La représentation des n individus par n points telle que la distance inter-points d_{rs} est égale à la dissimilarité δ_{rs} nécessite un espace de dimension plus petite ou égale à $n - 1$. Afin de visualiser cette configuration de points, nous allons projeter ces n points dans un sous-espace de dimension m : le cas $m = 2$ ou $m = 3$ est un cas intéressant de point de vue pratique.

Théorème 3.6.1 Soit Δ une matrice de distance. Une approximation à la représentation des individus par les points $Q_i \in \mathcal{L}_m$, $m < (n - 1)$ est faite en prenant les m premières composantes du vecteur \mathbf{x}_i , où \mathbf{x}_i est la i ème ligne de la matrice \mathbf{X} telle que $-\frac{1}{2}\mathbf{H}\Delta\mathbf{H} = \mathbf{X}\mathbf{X}'$

Démonstration. Etant donné que l'ordre d'importance des vecteurs propres est selon un ordre décroissant des valeurs propres, alors la meilleure dimension pour représenter les points Q_i est donnée par les vecteurs propres correspondant aux m premières valeurs propres. La mesure de la meilleure approximation par rapport à la vraie dimension de représentation m est donnée par:

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^n \lambda_i}$$

En effet, on a démontré que la matrice de coordonnées \mathbf{X} est égale à $\mathbf{V}\Gamma^{\frac{1}{2}}$, donc $\mathbf{X}'\mathbf{X} = (\Gamma^{\frac{1}{2}}\mathbf{V}')(\mathbf{V}\Gamma^{\frac{1}{2}}) = \Gamma = \text{diag}(\lambda_1, \dots, \lambda_p)$. Donc les distances entre les n points dans

l'espace euclidien de dimension $m \leq n - 1$ est donnée par:

$$\begin{aligned} d_{ij}^2 &= \sum_{r=1}^m (x_{ir} - x_{jr})^2 \\ &= \sum_{r=1}^m \gamma_r (v_{ir} - v_{jr})^2 \end{aligned}$$

Ainsi, la contribution $(x_{ir} - x_{jr})^2$ à la distance entre les points Q_i et Q_j sera petite si λ_r est petite. \square

3.6.2 Matrice définie positive

On a vu que pour avoir une représentation graphique des n individus par des points Q_i , $i = 1, \dots, n$ à partir d'une matrice de dissimilarité $\mathbf{A}_{n \times n}$, il faut et il suffit que $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ soit définie positive. Dans la pratique il y a des matrices \mathbf{B} qui ont des valeurs propres négatives. Ces valeurs propres négatives correspondent à des dimensions imaginaires pour \mathbf{X} . On peut se poser deux questions importantes : Quand pouvons nous ignorer ces valeurs, et que pouvons nous faire si on ne peut pas les ignorer?

D'après Sibson (1979), si les valeurs propres négatives sont très petites (*en valeurs absolue*) le résultat ne s'éloigne pas de la solution originale. En effet Sibson a pris une matrice de distance correspondant à une solution réelle, puis il a perturbé les valeurs après avoir utilisé les coordonnées principales. Alors si on ignore les valeurs propres négatives très petites ou les valeurs propres positives mais petites alors la solution est proche du résultat de la solution réelle originale.

Si les valeurs propres négatives ne sont pas négligeables, Caillier, (1983) a donné une solution analytique au problème de valeurs propres négatives. Il a cherché la plus petite valeur h^* telle que la dissimilarité définie par:

$$\delta_{ij}^h = \delta_{ij} + h(1 - \delta_{KR}^{ij})$$

ait une représentation dans un espace euclidien pour tout $h \geq h^*$, où $\delta_{KR}^{ij} = 0$ si $i = j$ et 1 si $i \neq j$.

3.7 Algorithme

Les différentes étapes pour la représentation des individus à partir d'une matrice de dissimilarité sont résumées dans ce qui suit:

Soit Δ la matrice de dissimilarité

(1) Calcul de la matrice \mathbf{A} tel que $a_{ij} = -\frac{1}{2}\delta_{ij}^2$ et $a_{ii} = 0$.

(2) Calcul de la matrice $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$.

(3) Calcul des valeurs propres $\lambda_1, \dots, \lambda_{n-1}$ et des vecteurs propres $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$, où les vecteurs propres sont de norme 1 pour tout i . Si \mathbf{B} n'est pas définie positive (il existe des valeurs propres négative) alors (i) on ignore les valeurs négative et on continue, ou (ii) on ajoute une constante appropriée h tel que $\delta_{ij}^* = \delta_{ij} + h(1 - \delta_{KR}^{ij})$ engendre une matrice de dissimilarité $\Delta^* \approx \Delta$.

(4) Les coordonnées de la i ème ligne de la matrice $\mathbf{X} = \mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}}$ nous donne le point Q_i .

(5) Calcul du coefficient $\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^n \lambda_i}$ pour juger la qualité de la représentation.

CHAPITRE IV

BIPLOT NON LINÉAIRE

4.1 Introduction

Au chapitre 1, nous avons représenté les individus dans le sous-espace des individus R_d , où $d < p$, R_d étant le sous-espace donné par l'ACP. Puis nous avons projeté, dans R_d , les droites que nous avons appelé axes biplot. Au chapitre 2, nous avons vu la théorie générale pour tracer des axes biplot dans un sous-espace \mathcal{L} quelconque de représentation des individus, $\mathcal{L} \subset R_p$, où \mathcal{L} est un sous-espace quelconque pouvant être obtenu en optimisant un critère quelconque. Nous présentons évidemment l'utilité de ces axes biplot linéaires: l'interprétation de la configuration du nuage de points, l'interpolation et la prédiction. Au chapitre 3, nous avons vu comment représenter les individus dans un sous-espace quelconque \mathcal{L} qui n'est pas nécessairement inclus dans R_p à l'aide de la technique du positionnement multidimensionnel. Le problème d'interpréter la configuration du nuage de points en retrouvant l'information sur chaque individu donnée par les variables se pose encore.

Dans ce chapitre, nous supposons que la proximité des individus est calculée à l'aide d'une distance euclidienne implantable. Cette distance δ_{ij} entre deux individus est calculée sur les vecteurs \mathbf{y}_i et \mathbf{y}_j . La représentation des individus \mathbf{y}_i par des points Q_i , $i = 1, \dots, n$ est faite dans un sous-espace \mathcal{L}_m , à partir de la matrice des distances entre les individus. La proximité des points-individus $Q_i \in \mathcal{L}_m$ et $Q_j \in \mathcal{L}_m$ donne de l'information sur δ_{ij} . Remarquons que le sous-espace approprié pour représenter les

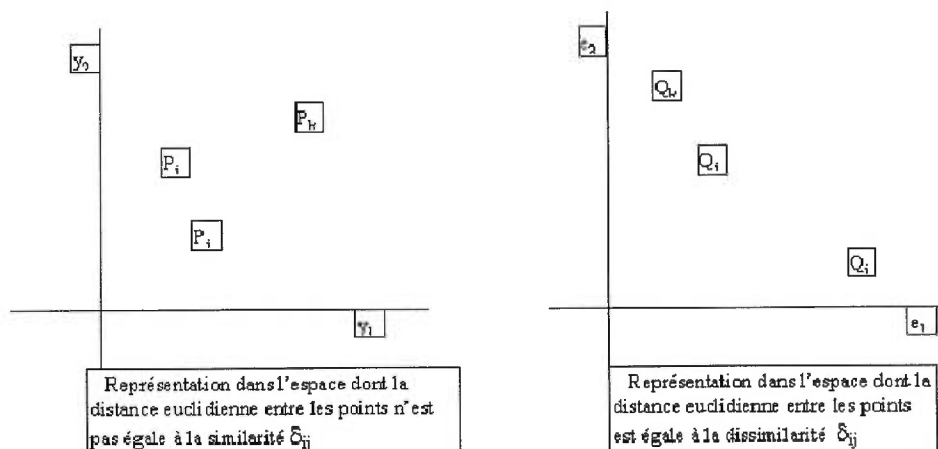


Figure 4.1: Représentation géométrique des données dans deux espaces distincts

individus est \mathcal{L}_m : celui-ci n'est pas le sous-espace R_d , $d \leq p$ qui contient les points P_i ou \hat{P}_i représentent les vecteurs \mathbf{y}_i ou $\mathbf{y}_{i[d]}$. La proximité des points $P_i \in R_d$, $P_j \in R_d$ renseigne sur la distance euclidienne $(\mathbf{y}_i - \mathbf{y}_j)'(\mathbf{y}_i - \mathbf{y}_j)$; cette dernière ne nous intéresse pas dans ce chapitre (voir figure 4.1) Le but du chapitre est de tracer des axes biplots dans \mathcal{L}_m afin que l'on puisse retrouver l'information y_{ij} ($i = 1, \dots, n$, $j = 1, \dots, p$) ou une approximation à celle-ci, et que l'on puisse interpoler un nouveau point dans \mathcal{L}_m .

Remarquons que la dimension de l'espace de représentation exacte \mathcal{L}_{n-1} est maintenant plus petite ou égale à $n - 1$ et non plus p comme dans les chapitres 1 et 2. Les axes de la configuration du nuage de points contenu dans \mathcal{L}_m sont des axes cartésiens arbitraires qui n'ont aucune signification.

Dans la section 4.2, nous établirons une notation, donnerons les hypothèses de base et discuterons de la technique de base qui est l'interpolation d'un point dans un espace quelconque \mathcal{L}_m . A la section 4.3 nous allons présenter la technique pour tracer des trajectoires dans l'espace \mathcal{L}_{n-1} qui serviront d'axes représentant les variables. A la section 4.4 nous verrons comment tracer deux types d'axes biplot dans \mathcal{L}_m , $m \leq n-1$: les axes biplot pour interpoler et les axes biplot pour prédire. Nous expliquerons comment se servir géométriquement des axes biplot ainsi définis pour interpréter, interpoler et

prédire. Remarquons une fois de plus que ces sous-espaces \mathcal{L}_m ne sont pas contenus dans \mathbb{R}_p . Nous verrons que ces axes biplot sont généralement non linéaires.

La notion de biplot non linéaire a été introduite par Gower et Harding (1988,1992a,1996) pour des variables quantitatives et par la suite, cette notion fut généralisée aux variables qualitatives. Nous ne discuterons pas du biplot pour variables qualitatives.

4.2 Préliminaire

4.2.1 Notation

1. G est le centre de gravité des n de points de la matrice $\mathbf{X}_{n \times m}$. Le point G est situé à l'origine des axes de \mathcal{L}_{n-1}
2. $d_i^2 = \mathbf{x}'_i \mathbf{x}_i$ est le carré de la distance du point $Q_i, i = 1, \dots, (n+1)$ au centre de gravité G .
3. $\mathbf{d}' = (d_1^2, d_2^2, \dots, d_n^2)$
4. $d_{n+1,i}^2$ est le carré de la distance entre les points Q_{n+1} et $Q_i, i = 1, \dots, n$
5. $\mathbf{d}'_{n+1} = -\frac{1}{2}(d_{n+1,1}^2, d_{n+1,2}^2, \dots, d_{n+1,n}^2)$
6. Le carré de la distance sera noté par \mathbf{d} distance. Nous suivons ainsi l'exemple de Gower et Hand (1996)
7. $\delta'_{n+1} = (\delta_{n+1,1}^2, \delta_{n+1,2}^2, \dots, \delta_{n+1,n}^2)$
8. $\mathbf{D} = (-\frac{1}{2}\delta_{ij}^2)_{i=1, \dots, n, j=1, \dots, p}$ la matrice dont les éléments sont $-\frac{1}{2}\delta_{ij}^2$.

4.2.2 Hypothèses de base

Dans ce chapitre, nous allons supposé que

1. la matrice $\mathbf{Y}_{n \times p}$ est connue

2. une distance δ_{ij} euclidienne implantable est calculée à partir de $\mathbf{Y}_{n \times p}$
3. la matrice de distance (δ_{ij}) est symétrique
4. $\delta_{ij}^2 = \sum_{k=1}^p f_k(y_{ik}, y_{jk})$
5. $f_k(y_{ik}, y_{jk}) > 0$; $f_k(y_{ik}, y_{jk}) = 0$ si et seulement si $y_{ik} = y_{jk}$
6. Les points Q_i dans l'espace \mathcal{L}_{n-1} ont été obtenus par la technique des coordonnées principales; c'est à dire par la méthode de positionnement multidimensionnel.

L'hypothèse 4 indique que chaque variable contribue indépendamment au carré de la distance. L'hypothèse 5 indique que δ est une distance. Nous allons supposer dans ce chapitre que la distance sur chacune des composantes est la même pour toutes les composantes, c'est à dire $f_k(y_{ik}, y_{jk}) = f(y_{ik}, y_{jk})$ pour tout k .

L'hypothèse 6 signifie que le vecteur des coordonnées du point Q_r est la r ième ligne de la matrice \mathbf{X} où $\mathbf{X} = \mathbf{L}\mathbf{\Gamma}^{\frac{1}{2}}$, $(\mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{D}(\mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}') = \mathbf{L}\mathbf{\Gamma}\mathbf{L}'$ (voir chapitre 3)

La dissimilarité δ_{ij} est calculée à partir des $\mathbf{y}_i \in R_p$, la distance d_{ij}^2 est calculée à partir des coordonnées $\mathbf{x} \in \mathcal{L}_m$. Etant donné que δ_{ij} est euclidienne implantable, on a que $\delta_{ij}^2 = d_{ij}^2$. On utilise cette double notation pour mieux expliciter lorsque nécessaire la provenance de la distance.

Proposition 4.2.1 *Le vecteur $\mathbf{d}' = (d_1^2, d_2^2, \dots, d_n^2)$ où d_i^2 est le carré de la distance de Q_i au centre de gravité s'écrit de la façon suivante:*

$$\mathbf{d} = \frac{1}{n^2}(\mathbf{1}'\mathbf{D}\mathbf{1})\mathbf{1} - \frac{2}{n}\mathbf{D}\mathbf{1} \quad (4.1)$$

En effet, Remarquons tout d'abord que le vecteur \mathbf{d} est formé des éléments diagonaux de la matrice $\mathbf{X}\mathbf{X}'$. Nous avons vu au chapitre 3 que:

$$\mathbf{X}\mathbf{X}' = (\mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{D}(\mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}')$$

d'ou

$$\mathbf{X}\mathbf{X}' = \mathbf{D} - \frac{1}{n}\mathbf{D}\mathbf{1}\mathbf{1}' - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{D} + \frac{1}{n^2}\mathbf{1}\mathbf{1}'\mathbf{D}\mathbf{1}\mathbf{1}'$$

on a:

$$\text{diag}(\mathbf{D}\mathbf{1}\mathbf{1}') = \text{diag}(\mathbf{1}\mathbf{1}'\mathbf{D}) = \mathbf{D}\mathbf{1}$$

$$\text{diag}(\mathbf{1}\mathbf{1}'\mathbf{D}\mathbf{1}\mathbf{1}') = (\mathbf{1}'\mathbf{D}\mathbf{1})\mathbf{1}$$

et $\text{diag}(\mathbf{D}) = \mathbf{0}$ car $\delta_{ii} = 0$, donc:

$$\mathbf{d} = \text{diag}(\mathbf{X}\mathbf{X}') = \frac{1}{n^2}(\mathbf{1}'\mathbf{D}\mathbf{1})\mathbf{1} - \frac{2}{n}\mathbf{D}\mathbf{1}$$

Remarque 4.2.1

$$\begin{aligned} \sum_{i=1}^n d_i^2 &= \mathbf{1}'\mathbf{d} \\ &= \frac{1}{n}(\mathbf{1}'\mathbf{D}\mathbf{1}) - \frac{2}{n}\mathbf{1}'\mathbf{D}\mathbf{1} \\ &= -\frac{1}{n}(\mathbf{1}'\mathbf{D}\mathbf{1}) \end{aligned} \tag{4.2}$$

4.2.3 Interpolation algébrique d'un point dans \mathcal{L}_m

Cette sous-section portera sur l'interpolation d'un nouveau point Q_{n+1} dans un sous-espace $\mathcal{L}_m, m \leq n - 1$. Nous supposons connues les coordonnées des n points Q_1, \dots, Q_n qui sont entreposées dans la matrice $\mathbf{X}_{n \times m}$; ces points sont centrés sur le centre de gravité G . Nous voulons trouver les coordonnées du point Q_{n+1} dans \mathcal{L} avec la seule connaissance des distances $d_{n+1,i}^2$ entre les points Q_i et Q_{n+1} ($i = 1, \dots, n$). Remarquons que l'ajout d'un point Q_{n+1} implique l'ajout d'une dimension à l'espace \mathcal{L}_m .

Théorème 4.2.1 Soient $\mathbf{x}' = (x_{n+1,1}, \dots, x_{n+1,m})$ les coordonnées inconnues de ce nouveau point Q_{n+1} dans $\mathcal{L}_m, m = n - 1$, et $x_{n+1,m+1}$ la coordonnée du point Q_{n+1} dans la dimension supplémentaire. Soit \mathbf{X} la matrice dont les vecteurs lignes sont les

vecteurs des coordonnées des points Q_1, \dots, Q_n . Soit $\mathbf{D}_{n \times n}$ la matrice des distances entre ces points. Les coordonnées du points Q_{n+1} sont données par les équations suivantes:

$$\mathbf{x}' = (x_{n+1,1}, \dots, x_{n+1,m})' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{d}_{n+1} - \frac{1}{n}\mathbf{D}\mathbf{1}). \quad (4.3)$$

$$x_{n+1,m+1}^2 = \frac{1}{n^2}\mathbf{1}'\mathbf{D}\mathbf{1} - \frac{2}{n}\mathbf{1}'\mathbf{d}_{n+1} - \mathbf{x}'\mathbf{x} \quad (4.4)$$

Démonstration. Les distances $d_{n+1,i}^2$ entre les points Q_{n+1} et Q_i ($i = 1, \dots, n$), doivent satisfaire l'équation:

$$d_{n+1,i}^2 = \sum_{k=1}^{m+1} (x_{ik} - x_{n+1,k})^2 \quad (4.5)$$

où $x_{i,m+1} = 0$ pour $i \neq n+1$ et $x_{n+1,k}$, $k = 1, \dots, m+1$ sont inconnus. On a:

$$\begin{aligned} d_{n+1,i}^2 &= \sum_{k=1}^m x_{ik}^2 + \sum_{k=1}^{m+1} x_{n+1,k}^2 - 2 \sum_{k=1}^m x_{ik}x_{n+1,k} \\ &= d_i^2 + d_{n+1}^2 - 2 \sum_{k=1}^m x_{ik}x_{n+1,k} \end{aligned} \quad (4.6)$$

où d_i^2 est la distance de Q_i au centre de gravité G (c.à.d. à l'origine des axes cartésiens de \mathcal{L}_{n-1}). L'équation 4.6 contient m inconnus $x_{n+1,k}$, $k = 1, \dots, m$.

On sait que les colonnes de la matrice \mathbf{X} sont centrées à l'origine c.à.d $\mathbf{X}'\mathbf{1} = 0$ (*chapitre 3*); on a donc que $\sum_{i=1}^n x_{ik} = 0$ pour tout $k = 1, \dots, m$.

En faisant la somme sur $i = 1, \dots, n$, l'équation 4.6 donne:

$$\begin{aligned} \sum_{i=1}^n d_{n+1,i}^2 &= \sum_{i=1}^n d_i^2 + nd_{n+1}^2 - 2 \sum_{k=1}^m \left(\sum_{i=1}^n x_{ik} \right) x_{n+1,k} \\ &= \sum_{i=1}^n d_i^2 + nd_{n+1}^2 \end{aligned}$$

d'où

$$d_{n+1}^2 = \frac{1}{n} \sum_{i=1}^n (d_{n+1,i}^2 - d_i^2) \quad (4.7)$$

$$= \frac{1}{n^2}\mathbf{1}'\mathbf{D}\mathbf{1} - \frac{2}{n}\mathbf{1}'\mathbf{d}_{n+1} \quad (4.8)$$

L'équation 4.8 est obtenue en utilisant l'équation 4.2 de la remarque de la proposition 4.2.1, et le fait que $-\frac{1}{2} \sum_{i=1}^n d_{n+1,i}^2 = \mathbf{1}'\mathbf{d}_{n+1}$.

En substituant l'équation 4.7 dans l'équation 4.6 on obtient:

$$2 \sum_{k=1}^m x_{ik} x_{n+1,k} = d_i^2 - d_{n+1,i}^2 + \frac{1}{n} \sum_{i=1}^n (d_{n+1,i}^2 - d_i^2) \quad (4.9)$$

La forme matricielle de l'équation 4.9 est:

$$2\mathbf{X}\mathbf{x} = \begin{pmatrix} d_1^2 \\ \cdot \\ \cdot \\ d_n^2 \end{pmatrix} - \begin{pmatrix} d_{n+1,1}^2 \\ \cdot \\ \cdot \\ d_{n+1,n}^2 \end{pmatrix} + \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n (d_{n+1,i}^2 - d_i^2) \\ \cdot \\ \cdot \\ \frac{1}{n} \sum_{i=1}^n (d_{n+1,i}^2 - d_i^2) \end{pmatrix}$$

d'où par la proposition 1, on obtient:

$$2\mathbf{X}\mathbf{x} = \frac{1}{n^2} (\mathbf{1}'\mathbf{D}\mathbf{1})\mathbf{1} - \frac{2}{n}\mathbf{D}\mathbf{1} + 2\mathbf{d}_{n+1} + \left(\frac{1}{n} \sum_{i=1}^n (d_{n+1,i}^2 - d_i^2) \right) \mathbf{1}$$

Et si on multiplie les deux termes de l'égalité par $\frac{1}{2}\mathbf{X}'$, alors:

$$\begin{aligned} \mathbf{X}'\mathbf{X}\mathbf{x} &= \frac{1}{2n^2} (\mathbf{1}'\mathbf{D}\mathbf{1})\mathbf{X}'\mathbf{1} + \mathbf{X}'(\mathbf{d}_{n+1} - \frac{1}{n}\mathbf{D}\mathbf{1}) + \left(\frac{1}{2n} \sum_{i=1}^n (d_{n+1,i}^2 - d_i^2) \right) \mathbf{X}'\mathbf{1} \\ &= \mathbf{X}'(\mathbf{d}_{n+1} - \frac{1}{n}\mathbf{D}\mathbf{1}) \end{aligned} \quad (4.10)$$

car $\mathbf{X}'\mathbf{1} = \mathbf{0}$. Et puisque l'on a supposé que la matrice \mathbf{X} est de plein rang $m = n - 1$ alors $\mathbf{X}'\mathbf{X}$ est inversible, l'équation 4.10 devient:

$$\mathbf{x} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{d}_{n+1} - \frac{1}{n}\mathbf{D}\mathbf{1})$$

Ceci nous donne les m premières coordonnées du point Q_{n+1} , La coordonnée de la dimension supplémentaire est trouvée facilement:

$$\begin{aligned} d_{n+1}^2 &= \sum_{k=1}^{m+1} x_{n+1,k}^2 \\ &= \sum_{k=1}^m x_{n+1,k}^2 + x_{n+1,m+1}^2 \\ &= \mathbf{x}'\mathbf{x} + x_{n+1,m+1}^2 \end{aligned}$$

d'où

$$x_{n+1,m+1}^2 = d_{n+1}^2 - \mathbf{x}'\mathbf{x}$$

A l'aide de la proposition 1, et de l'équation 4.8 on réécrit cette dernière équation:

$$x_{n+1,m+1}^2 = \frac{1}{n^2} \mathbf{1}' \mathbf{D} \mathbf{1} - \frac{2}{n} \mathbf{1}' \mathbf{d}_{n+1} - \mathbf{x}' \mathbf{x}$$

Ou encore

$$x_{n+1,m+1}^2 = \bar{\mathbf{D}} - \frac{2}{n} \mathbf{1}' \mathbf{d}_{n+1} - \mathbf{x}' \mathbf{x}$$

où $\bar{\mathbf{D}} = \frac{1}{n^2} \mathbf{1}' \mathbf{D} \mathbf{1}$ représente la moyenne sur les n^2 distances.

Corollaire 4.2.1 *Le résultat du théorème est applicable quelque soit la matrice \mathbf{X} dont les colonnes sont centrées à l'origine: matrice des coordonnées des points-individus par rapport à des axes orthogonaux. Si $\mathbf{X}'\mathbf{X} = \mathbf{\Lambda}$, où $\mathbf{\Lambda}$ est la matrice diagonale des valeurs propres (voir chapitre 3), alors les coordonnées du point Q_{n+1} seront:*

$$\mathbf{x}' = (x_1, \dots, x_m)' = \mathbf{\Lambda}^{-1} \mathbf{X}' (\mathbf{d}_{n+1} - \frac{1}{n} \mathbf{D} \mathbf{1}) \quad (4.11)$$

Remarque 4.2.2 La dernière forme du vecteur \mathbf{x} dans \mathcal{L} est simple à calculer.

Remarque 4.2.3 Remarquons que la $(m+1)$ ème coordonnée de $x_{n+1,m+1}^2$ est égale à $-\frac{2}{n} \mathbf{1}' \mathbf{d}_{n+1} - \mathbf{x}' \mathbf{x}$. Si cette expression est négative, alors on ne peut pas ajouter le point Q_{n+1} dans l'espace \mathcal{L}_{m+1} . C'est une autre façon de donner un condition nécessaire et suffisante pour qu'une distance soit euclidienne-implantable.

Remarque 4.2.4 La projection $\mathbf{x}_m \in \mathcal{L}_m$ du point \mathbf{x} est obtenue en remplaçant $\mathbf{\Lambda}$ par $\mathbf{\Lambda}_m$, \mathbf{X} par \mathbf{X}_m où \mathbf{X}_m est constitué des m premières colonnes de \mathbf{X} , et la diagonale de la matrice $\mathbf{\Lambda}_m$ est constituée des m plus grandes valeurs propres de $\mathbf{X}'\mathbf{X}$.

4.3 Axes non linéaires dans \mathcal{L}_{n-1}

Il est évidemment facile d'utiliser des axes orthogonaux pour représenter les individus, pour interpoler et pour prédire mais l'orthogonalité des axes n'est pas une nécessité. Nous avons vu que dans le cas du biplot linéaire, les axes étaient obliques. Nous allons maintenant tracer des axes qui sont des trajectoires dans \mathcal{L}_{n-1} et nous verrons que ces trajectoires partagent des propriétés importantes avec des axes conventionnels.

4.3.1 Trajectoires dans \mathcal{L}_{n-1}

Soit $\rho \mathbf{e}_r$ un vecteur observation d'un pseudo-individu pour lequel chaque variable a une mesure nulle sauf la r ième variable qui prend la valeur ρ . Soient les points $P \in R_p$ et $Q \in \mathcal{L}_m$ qui représente ce pseudo-individu dans chacun des sous-espaces. Le vecteur individu \mathbf{y}_i est représenté par le point $P_i \in R_p$ et $Q_i \in \mathcal{L}_m$. Nous allons trouver les coordonnées du point Q à l'aide de l'équation 4.11 et les distances entre le point P et les points P_i c'est à dire distances entre le point Q et les points Q_i). Le vecteur $\delta_{n+1}(\rho)$ des distances entre ce point P et les points P_i $i = 1, \dots, n$ est :

$$\begin{aligned} \delta_{n+1}(\rho) &= \begin{pmatrix} \delta^2(y_1, \rho \mathbf{e}_r) \\ \delta^2(y_2, \rho \mathbf{e}_r) \\ \vdots \\ \delta^2(y_n, \rho \mathbf{e}_r) \end{pmatrix} \\ &= \left\{ \delta^2(\mathbf{y}_i, \rho \mathbf{e}_r) \right\}_{i=1, \dots, n} \end{aligned} \quad (4.12)$$

où $\{\}_{1, \dots, n}$ représente le vecteur des n éléments.

$$\begin{aligned} \delta_{n+1}(\rho) &= \left\{ \sum_{k=1}^p f(y_{ik}, \rho \mathbf{e}_r) \right\}_{1, \dots, n} \\ &= \left\{ \sum_{k=1, k \neq r}^p f(y_{ik}, 0) + f(y_{ir}, \rho) \right\}_{1, \dots, n} \\ &= \left\{ \sum_{k=1}^p f(y_{ik}, 0) - f(y_{ir}, 0) + f(y_{ir}, \rho) \right\}_{i=1, \dots, n} \end{aligned} \quad (4.13)$$

Etant donné que la distance entre le point P et le point P_i est égale à la distance entre le point Q et le point Q_i on a alors $\mathbf{d}_{n+1}(\rho) = \delta_{n+1}(\rho)$ d'où:

$$\mathbf{d}_{n+1}(\rho) = \left\{ \sum_{k=1}^p f(y_{ik}, 0) - f(y_{ir}, 0) + f(y_{ir}, \rho) \right\}_{i=1, \dots, n} \quad (4.14)$$

Nous trouvons les coordonnées $\mathbf{x}' = (x_1, \dots, x_{m-1}, x_m)$ du point Q représentant le pseudo-individu $\rho \mathbf{e}_r$ en utilisant le théorème 1, en substituant 4.14 dans l'équation 4.11.

Pour chaque valeur de ρ , l'ajout du pseudo-individu $\rho \mathbf{e}_k$ nécessite une nouvelle dimension. Etant donnée que cette nouvelle dimension ne nous intéresse pas et que les coordonnées de \mathbf{x} dans \mathcal{L}_{n-1} ne dépendent pas de la n ième dimension, nous appellerons cette n ième dimension, une dimension résiduelle et tous les nouveaux points seront interpolés en utilisant cette même dimension résiduelle. La distance inter-point est exacte dans \mathcal{L}_{n-1} ; les distances inter-points dans la dimension résiduelle ne sont pas préservées et de toute façon, elles ne représentent aucun intérêt.

Une trajectoire représentant la k ième variable dans \mathcal{L}_{n-1} est définie par le lieu géométrique des points \mathbf{x} dans \mathcal{L}_{n-1} obtenus en interpolant $\rho \mathbf{e}_k$ pour ρ variant de $-\infty$ à $+\infty$. Cette trajectoire est généralement non linéaire et dépend de la distance d utilisée. On établit une échelle sur cette trajectoire en marquant les points de la trajectoire par $\rho = 1, \pm 2, \dots, etc.$

Remarque 4.3.1 En posant $\rho = 0$, on obtient le point de rencontre des p trajectoires: ce point O représente la moyenne des variables. Ce point O est différent du centre de gravité G du nuage de points de l'échantillon. Ce point O n'est donc pas à l'origine des axes \mathcal{L}_{n-1} .

Il est souhaitable que la méthode proposée pour tracer des trajectoires dans $\mathcal{L}_m \neq R_d$ donne des résultats compatibles avec la méthodologie des axes biplot dans $\mathcal{L} = R_d \subset R_p$. C'est ce que nous énonçons dans la proposition suivante:

Proposition 4.3.1 *Si la distance δ entre deux individus est une distance pythagorienne alors les trajectoires tracées en substituant $\rho \mathbf{e}_k$ dans l'équation 4.11, et en faisant varier ρ de $-\infty$ à $+\infty$, est une droite et celle-ci coïncide avec l'axe biplot classique.*

Démonstration. Soit $\mathbf{Y}_{n \times p}$ la matrice d'observations sur les n individus et soit $\mathbf{Y} = \mathbf{U}\mathbf{T}\mathbf{V}'$ sa décomposition en valeurs singulières. Les coordonnées des points $P_i, \in R_p, (i = 1, \dots, n)$ relatifs à la base \mathbf{V} sont les vecteur-lignes de $\mathbf{U}\mathbf{T}$.

Nous allons démontrer que l'axe qui représente chaque variable dans R_p peut être

obtenu de deux façons:

1. par la projection des pseudo-individus $\rho \mathbf{e}_k$ sur la base \mathbf{V}_p , où $-\infty < \rho < \infty$ tel décrit au chapitre 1
2. par l'interpolation des pseudo-individus $\rho \mathbf{e}_k$ dans R_p avec l'équation 4.11.

Nous démontrerons ainsi que les axes qui représentent les variables obtenus par chacune des deux méthodes sont identiques.

Soit ξ un nouveau point à interpoler; ses coordonnées relatif à la base \mathbf{V} sont $\tilde{\mathbf{x}} = \mathbf{V}'\xi$. Nous allons démontrer que les coordonnées du point ξ dans R_d calculées avec l'ajout d'un point ξ en se basant sur les distances entre ξ et \mathbf{y}_i sont égales à $\tilde{\mathbf{x}}$. On a:

$$\begin{aligned} \delta_{n+1,i}^2 &= (\xi - \mathbf{y}_i)'(\xi - \mathbf{y}_i) \\ &= \xi'\xi - 2\mathbf{y}_i'\xi + \mathbf{y}_i'\mathbf{y}_i \\ &= \xi'\xi - 2\mathbf{y}_i'\xi + \delta_i^2 \end{aligned}$$

donc

$$-\frac{1}{2}d_{n+1,i}^2 = -\frac{1}{2}d_i^2 + \mathbf{y}_i'\xi - \frac{1}{2}\xi'\xi \quad \text{car } \delta = d$$

Sous la forme matricielle

$$\mathbf{d}_{n+1} = -\frac{1}{2}\mathbf{d} + \mathbf{Y}\xi - \frac{1}{2}\xi'\xi\mathbf{1}$$

et par la proposition 1 on obtient:

$$\mathbf{d}_{n+1} = -\frac{1}{2n^2}(\mathbf{1}'\mathbf{D}\mathbf{1})\mathbf{1} - \frac{1}{n}\mathbf{D}\mathbf{1} + \mathbf{Y}\xi - \frac{1}{2}\xi'\xi\mathbf{1}$$

d'où

$$\mathbf{d}_{n+1} - \frac{1}{n}\mathbf{D}\mathbf{1} = -\frac{1}{2n^2}(\mathbf{1}'\mathbf{D}\mathbf{1})\mathbf{1} + \mathbf{Y}\xi - \frac{1}{2}\xi'\xi\mathbf{1} \quad (4.15)$$

En substituant 4.15 dans l'équation 4.11 on obtient:

$$\begin{aligned} \mathbf{x} &= \mathbf{\Lambda}^{-1}\mathbf{X}'\mathbf{Y}\xi \quad \text{car } \mathbf{X}'\mathbf{1} = \mathbf{0} \\ &= \mathbf{\Lambda}^{-1}\mathbf{X}'\mathbf{U}\mathbf{F}\mathbf{V}'\xi \end{aligned}$$

$$\begin{aligned}
&= \Lambda^{-1} \mathbf{X}' \mathbf{X} \mathbf{V}' \xi \\
&= \Lambda^{-1} \Lambda \mathbf{V}' \xi \\
&= \mathbf{V}' \xi \\
&= \tilde{\mathbf{x}}
\end{aligned}$$

Ce qui démontre que l'interpolation d'un point ξ à l'aide de l'équation 4.11 donne les mêmes résultats qu'en projetant orthogonalement ξ sur \mathbf{V} . Si on prend $\xi = \rho \mathbf{e}_k$ et on fait varier ρ de $-\infty$ à $+\infty$ alors \mathbf{x} va représenter une droite qui n'est autre que la projection de $\rho \mathbf{e}_k$ sur \mathbf{V} , c'est à dire l'axe biplot classique.

Nous allons voir à la sous-section suivante que ces trajectoires ont des propriétés qui nous permettent de considérer que ces trajectoires peuvent servir comme des axes conventionnels.

4.3.2 Prédiction conventionnelle

La distance entre le point Q_i et la r ième trajectoire est donnée par l'équation 4.14 le seul terme de cette équation qui dépend de ρ est $f(\rho, y_{ir}) : d_{n+1}^2(\rho)$ est minimale si $f(\rho, y_{ir})$ est minimale. Ce qui indique que le point sur la trajectoire qui est le plus près de Q_i est le point pour lequel $f(\rho, y_{ik})$ est minimal.: en effet si $\rho = y_{ik}$ on a $f(\rho, y_{ik}) = f(y_{ik}, y_{ik}) = 0$. Par construction de l'échelle sur les trajectoires, on a effectivement que ce point le plus près avait été marqué par $\rho = y_{ik}$. La valeur y_{ik} peut donc être retrouvée en érigeant une normale à la trajectoire. La prédiction des valeurs du point Q_i sur chacune des trajectoires est donc faite comme avec des axes conventionnels. Il peut y avoir plusieurs points normaux à la trajectoire qui est curviligne, nous choisissons le point plus près.

4.3.3 Interpolation conventionnelle

Avec des axes conventionnels, un point P_i de R_p qui représente l'individu $y_i = (y_{i1}, \dots, y_{ip})$ est déterminé uniquement par l'intersection de p sous-espaces; chacun étant

orthogonal à un des p axes. Il faut remarquer que la méthode conventionnelle appliquée à l'espace \mathcal{L}_m pour interpoler un point Q_i détermine une région et non pas un unique point. Voyons que cela ne pose aucun problème.

Le point Q_i peut être situé dans \mathcal{L}_{n-2} en construisant p sous-espaces de dimension $(n-2)$: le k ième sous-espace est orthogonal à la k ième trajectoire au point x_{ik} pour $k = 1, \dots, p$. Ces p sous-espaces ont une intersection dans au moins $n-p-1$ dimensions. On a donc qu'une région peut prédire les mêmes valeurs (x_{i1}, \dots, x_{ip}) . Toutefois, chaque point-individu d'une région prédit une unique valeur donnée par le point le plus près sur chaque trajectoire.

On peut donc conclure que les trajectoires se comportent suffisamment comme des axes conventionnels.

4.4 Axes biplot dans \mathcal{L}_m

L'unicité de l'axe biplot linéaire β_j pour fin d'interpolation et la multiplicité de l'axe biplot linéaire β_j pour fin de prédiction a été démontré au chapitre 2. L'axe biplot β_j pouvait être n'importe quel segment de droite dans R_d si on l'utilisait pour prédire. Il n'y avait qu'une seule façon de définir l'axe biplot β_j si on l'utilisait pour interpoler. Il était utile de choisir le même axe biplot β_j de référence pour faire ces deux opérations. Dans le cas non linéaire, nous verrons qu'il y a multiplicité de l'axe biplot β_j prédictif et unicité de l'axe biplot β_j pour l'interpolation. Nous verrons, de plus, que malheureusement, on ne peut choisir le même axe biplot β_j pour faire ces deux opérations dans le cas non linéaire.

4.4.1 Axes biplot d'interpolation

Dans le cas linéaire, l'axe biplot β_j , unique pour l'interpolation, est obtenu par la projection orthogonale de l'axe cartésien ρe_k de R_p sur la base \mathbf{V} dans R_d . Dans le cas non linéaire, nous projetons orthogonalement dans le sous-espace \mathcal{L}_m chacune des p trajectoires de \mathcal{L}_{n-1} décrites à la section 4.3.1. Rappelons que le sous-espace \mathcal{L}_m

est obtenu en projetant orthogonalement les n points $Q_i \in \mathcal{L}_{n-1}$ dans le sous-espace \mathcal{L}_m : les m premières coordonnées de Q_i fournissent les coordonnées de ces projections. En effet les coordonnées du point Q_i sont données par rapport à des axes de référence orthonormés. L'hypothèse que le sous-espace est trouvé à l'aide des coordonnées principales est ici utilisée. Il suffit d'utiliser les m premières colonnes de $\mathbf{\Lambda}$ et de \mathbf{X} de l'équation 4.11 pour trouver les m coordonnées d'un point \mathbf{Q} projeté dans \mathcal{L}_m . La j ème trajectoire est donc projetée de la même façon dans \mathcal{L}_m :

Définition 4.4.1 Les m premières coordonnées de la trajectoire tracée en interpolant $\rho \mathbf{e}_k$ pour tout $-\infty < \rho < +\infty$ est défini comme étant l'axe biplot β_j .

Le traçage des axes biplot est intéressant en soit mais ces axes biplot doivent avoir leur utilité.

Interpolation géométrique d'un point quelconque

Soit $\xi' = (\xi_1, \dots, \xi_p) = \sum_{j=1}^p \xi_j \mathbf{e}_j$ un nouveau vecteur que l'on veut interpoler dans \mathcal{L}_m ; celui-ci est représenté par le nouveau point Q_{n+1} . La distance $\delta_{n+1,i}^2$ de ξ à un autre vecteur-individu \mathbf{y}_i se décompose en p composantes car la distance est additive. Alors pour tout point $Q_i \in \mathcal{L}_m$ $1, \dots, n$ on a:

$$d_{n+1,i}^2 = \delta_{n+1,i}^2 = \sum_{r=1}^p f(y_{ij}, \xi_r) \quad (4.16)$$

La distance δ_{ir}^2 de $\xi_r \mathbf{e}_r$ à \mathbf{y}_i est:

$$d_{ir}^2 = \delta_{ir}^2 = f(y_{ir}, \xi_r) + \sum_{k=1}^p f(y_{ik}, 0) - f(y_{ir}, 0) \quad (4.17)$$

En substituant 4.17 dans 4.16 alors:

$$\begin{aligned} d_{n+1,i}^2 &= \sum_{r=1}^p d_{ir}^2 - p \sum_{k=1}^p f(y_{ik}, 0) + \sum_{r=1}^p f(y_{ir}, 0) \\ &= \sum_{r=1}^p d_{ir}^2 - (p-1) \sum_{r=1}^p f(y_{ir}, 0) \end{aligned} \quad (4.18)$$

La forme matricielle de l'équation 4.18 s'écrit:

$$\mathbf{d}_{n+1}(\xi) = \sum_{r=1}^p \mathbf{d}_{n+1}(\xi_r) - (p-1) \left\{ \sum_{r=1}^p f(y_{ir}, 0) \right\}_{1, \dots, n} \quad (4.19)$$

où $\mathbf{d}_{n+1}(\xi)$ et $\mathbf{d}_{n+1}(\xi_r)$ sont les vecteurs dont les composantes sont égales aux distances de ξ à P_i et de ξ_r à P_i respectivement

Le dernier terme de cette dernière équation est la distance du vecteur-individu Q_i à l'origine O qui est différent du centre de gravité G . En substituant 4.19 dans 4.11, on trouve les coordonnées \mathbf{x} du point Q_{n+1} :

$$\mathbf{x} = \mathbf{\Lambda}^{-1} \mathbf{X}' \left[\sum_{k=1}^p (\mathbf{d}_{n+1}(\xi_k) - \frac{1}{n} \mathbf{D} \mathbf{1}) - \left(\left\{ \sum_{k=1}^p f(y_{ik}, 0) \right\}_{1, \dots, n} - \frac{1}{n} \mathbf{D} \mathbf{1} \right) \right] \quad (4.20)$$

Le premier terme de 4.20 est une sommation de p composantes. Le k ième terme de cette sommation est $\mathbf{\Lambda}^{-1} \mathbf{X}' (\mathbf{d}_{n+1}(\xi_k) - \frac{1}{n} \mathbf{D} \mathbf{1})$. Ce terme représente un point Q_{ξ_k} sur la k ème trajectoire obtenu par l'interpolation du pseudo-point $\xi_k \mathbf{e}_k$. Ce point Q_{ξ_k} est exprimé dans les coordonnées du système d'axes cartésiens de \mathcal{L}_{n-1} ; le second terme de 4.20 donne l'interpolation de l'origine O qui est le point d'interrection des p trajectoires. Le point \mathbf{x} est donc une somme de p vecteurs $\overrightarrow{Q_{\xi_k} O}$ tracés dans l'espace euclidien \mathcal{L}_m . On peut donc faire l'interpolation par la méthode du parallélogramme pour situer le point Q . Plus simplement on trouve le centroïde des points $Q_{\xi_1}, Q_{\xi_2}, \dots, Q_{\xi_p}$ et en multipliant par p la longueur de ce centroïde. On a déjà utilisé cette méthode au chapitre 1 dans le cas du biplot linéaire.

Même si les trajectoires sont non linéaires, l'indépendance de la contribution de chaque variable nous permet d'interpoler en faisant la somme de p vecteurs d'une façon analogue au cas linéaire

Exemple

4.4.2 Axes biplot de prédiction

La prédiction, à partir d'un point Q dans l'espace \mathcal{L}_{n-1} , des valeurs que prendraient les variables est basée sur le principe du point le plus près: c'est donc le même principe qui était utilisé dans la théorie générale du biplot linéaire du chapitre 2. La différence étant que les axes biplot sont maintenant non linéaires. Nous allons mainte-

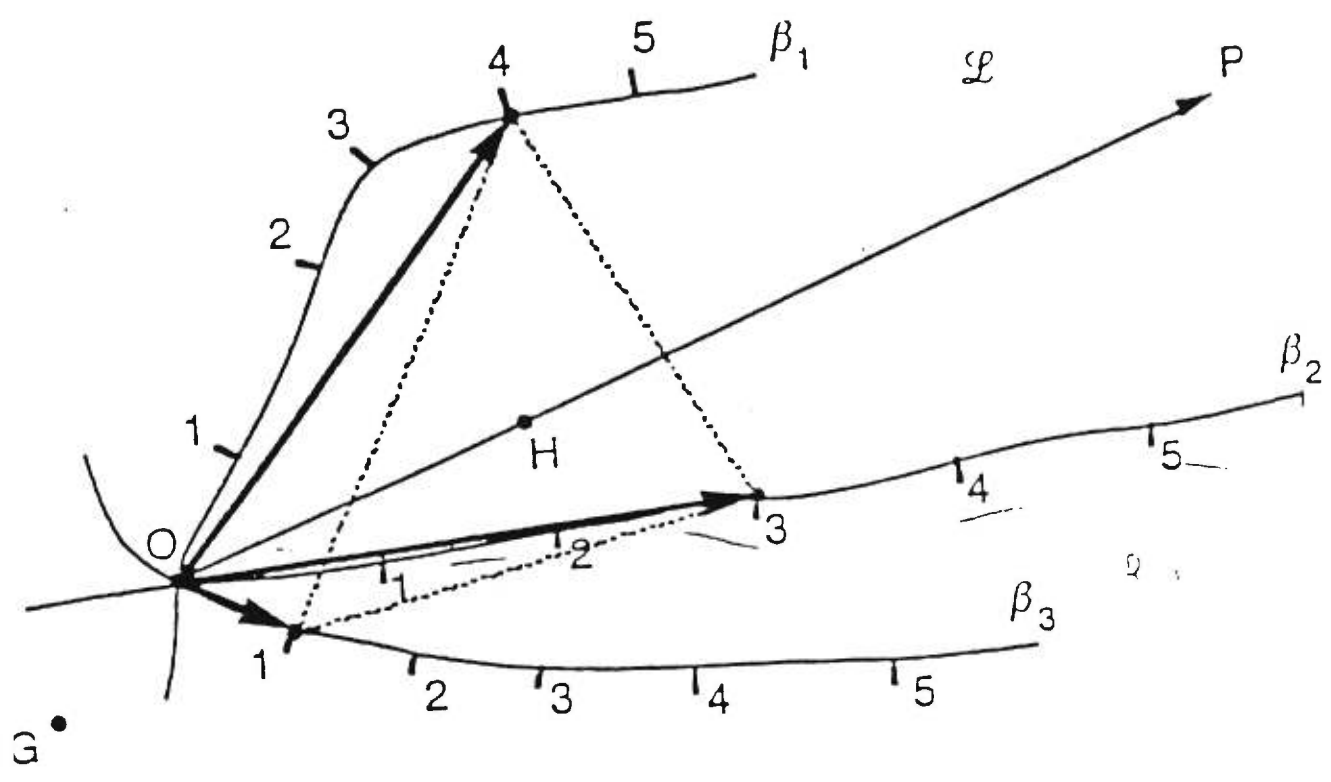


Figure 4.2: Interpolation d'un point $P(4, 3, 1)$ dans le cas non linéaire. « tiré de Gower et Hand 1996, page 109 ».

nant définir les axes biplot de prédiction. Avant tout, faisons la remarque suivante.

Soit ξ_1 un marqueur sur une trajectoire dans l'espace \mathcal{L}_{n-1}^+ où \mathcal{L}_{n-1}^+ est l'espace \mathcal{L}_{n-1} plus la dimension résiduelle: tous les points dans le sous-espace normale N_1 à cette trajectoire au point ξ_1 auront comme prédiction la valeur ξ_1 . En particulier tous les points de \mathcal{L}_m qui appartiennent à $N_1 \cap \mathcal{L}_m$ auront comme prédiction la valeur ξ_1 . Même si la trajectoire est non linéaire, le sous-espace $N_1 \cap \mathcal{L}_m$ est linéaire. Cette prédiction est illustrée à la figure 4.3. Celle-ci illustre aussi la prédiction d'une deuxième valeur ξ_2 pour une autre région de points dans \mathcal{L}_m . Sur cette figure apparait aussi le marqueur ξ_2 obtenu par l'intersection du sous-espace normal $N_2 \cap \mathcal{L}_m$ à la trajectoire ξ de l'espace \mathcal{L}_{m-1}^+ . On remarque que les sous-espace N_1 et N_2 ne sont pas parallèles; par conséquent un axe biplot linéaire qui est simultanément orthogonal à tous les sous-espaces $N \cap \mathcal{L}_m$ n'est plus possible. Il s'en suit que si par exemple $N_1 \cap \mathcal{L}_m$ et $N_2 \cap \mathcal{L}_m$ ne sont pas parallèles alors ils peuvent avoir une intersection dans \mathcal{L}_m et ainsi deux valeurs distinctes ξ_1 et ξ_2 seront prédites pour une même région de \mathcal{L}_m . Gower et Hand (1996, p110) font remarquer que si le sous-espace \mathcal{L}_m approxime bien le nuage de points alors de telles régions seraient au dehors de la région qui représente bien les données.

La prédiction d'une valeur ξ revient à choisir le sous-espace $N \cap \mathcal{L}_m$ qui prédira cette valeur ξ . Nous allons maintenant construire un axe biplot qui servira à la prédiction.

Définition de l'axe biplot de prédiction

Soit ξ un point sur une trajectoire dans \mathcal{L}_{n-1}^+ , soit N_ξ des sous-espaces orthogonaux à la trajectoire en ce point. L'axe biplot β_k de prédiction est le lieu géométrique des projections orthogonales du point O dans les sous espaces $N_\xi \cap \mathcal{L}_m$, où ξ est sur la k ième trajectoire tracée à la section 4.3.1. Nous allons maintenant illustrer que cet axe biplot β_k peut servir pour la prédiction.

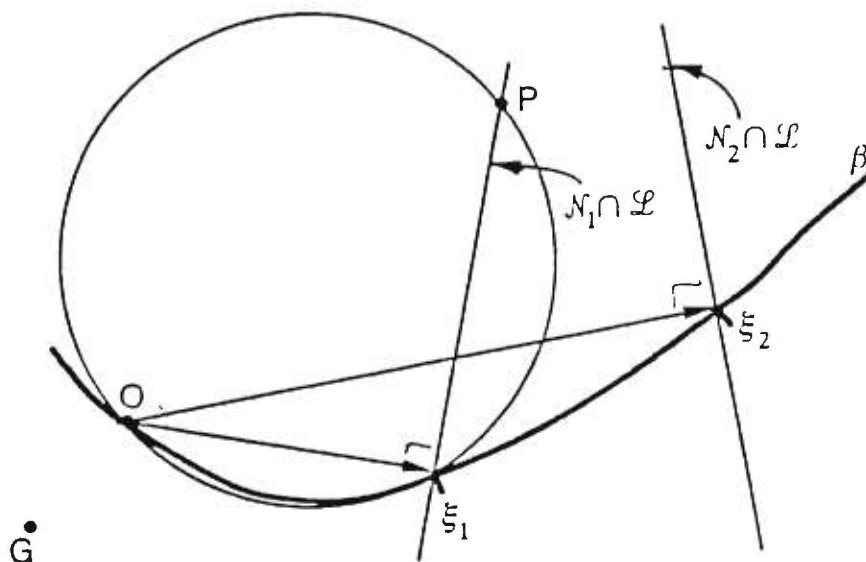


Figure 4.3: Utilisation de la trajectoire non linéaire β comme axe biplot. « tiré de Gower 1996, page 111 »

Propriétés

La figure 4.3 illustre la construction de l'axe biplot β et la façon de se servir de cet axe pour prédire la valeur ξ_1 que prend un point Q déjà dans \mathcal{L}_m . Cette propriété de l'axe biplot s'explique de la façon suivante.

Soit un cercle de diamètre OQ , ce diamètre sous tend des angles droits sur la circonférence: en effet, le diamètre du cercle fait un angle droit avec n'importe quel point de la circonférence. La valeur ξ_1 est un point d'intersection du cercle avec l'axe biplot qui vient d'être tracé. En effet, la projection orthogonale du point O sur l'axe β est marquée de la valeur ξ_1 . De plus la projection orthogonale du point Q sur cet axe biplot β prédit la valeur ξ_1 . La projection orthogonale lorsqu'il s'agit de distance pythagorienne est remplacée par une projection circulaire. Remarquez que si l'axe biplot est linéaire, les projections orthogonale et circulaire coïncident.

Propriétés Le choix du point O comme étant le point que l'on projette est un choix arbitraire. N'importe quel point aurait pu être choisi : les axes biplot auraient tous différé selon le choix de ce point mais la prédiction aurait été la même. Le point O a comme avantage qu'il a une signification: il représente le vecteur moyenne des variables. Ce point a le désavantage qu'il n'existe pas d'algorithme simple pour calculer les coordonnées de ce point.

Propriétés Cette méthode de tracer les axes biplot prédictifs n'est pas basée sur une équation paramétrique: la recherche d'une telle équation qui donnerait les coordonnées des trajectoires dans \mathcal{L}_{n-1} est un axe actuel de recherches. Le traçage des axes biplot serait ainsi facilité. En fait, on utilise l'équation 4.11 pour trouver ces coordonnées.

Gower et Hand (1996, p112) discute d'une approximation linéaire par morceau de cette trajectoire. On obtient ainsi localement des segments sur la trajectoire et l'on peut utiliser les résultats linéaires du chapitre 2.

CONCLUSION

Les récents développements de la théorie du biplot ont été étudiés dans ce mémoire. On a vu que les axes biplot classiques ont été généralisés pour des sous-espaces arbitraires dans lesquels on peut représenter les individus.

Si le sous-espace de représentation des individus est trouvé à partir des variables originales, et si la distance inter-individus est euclidienne, alors ces axes sont linéaires. Si le sous-espace de représentation des individus est trouvé à partir d'une matrice de distance euclidienne-implantable, alors ces axes sont généralement non linéaires.

Dans les deux cas, on peut utiliser ces axes biplot pour interpoler géométriquement un nouveau point, prédire géométriquement à partir d'un point et donner un sens à la configuration du nuage de points. On remarque que le biplot peut être vu comme de l'algèbre linéaire appliquée.

Il serait intéressant d'appliquer la technique du biplot linéaire général dans des sous-espaces où la représentation des individus est trouvée à partir de la méthodologie générale de la projection poursuite (« *projection poursuit* »).

RÉFÉRENCES

- Benzécri, J.-P. (1992). *Correspondance Analysis Handbook*. New York : Marcel Dekker.
- Cox, T. F., et M. A. A. Cox. (1994). *Multidimensional Scaling*. London : Chapman & Hall.
- Daigle, G., et L.-P. Rivest. (1992). « Robust biplot ». *The Canadian Journal of Statistics*, vol. 20, no 3, p. 241–255.
- Dempster, A. P. 1969. *Elements of Continuous Multivariate Analysis*. San Francisco : Adisson-Wesley Publishing Company.
- Digby, P. G. N., et J. C. Gower. 1986. *Ordination and classification*. Montréal : Les Presses de l'Université de Montréal.
- Eckart, C., et G. Young. 1936. « The approximation of one matrix by another of lower rank ». *Psychometrika*, vol. 1, p. 211–218.
- Gabriel, K. R. 1971. « The biplot graphical display of matrices with application to principal component analysis ». *Biometrika*, vol. 58, p. 53–67.
- Gabriel, K. R. 1981. « Biplot display of multivariate matrices for inspecting of data and diagnosis ». In *Interpreting Multivariate Data*, V. Barnett (comp.), New York : John Wiley & Sons.
- Gabriel, K. R. 1995. « Biplot display of multivariate categorical data, with couruents on multiple correspondance analysis. In *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski (comp.), Oxford : Oxford Science Publications.
- Gabriel, K. R., et C. L. Odoroff. 1986. « Some diagnoses of models by 3-D biplots ». In *Multidimensional Data Analysis*, J. De Leeuw, W. Heiser, J. Meulman et F. Critchley (comp.), Leinden : DSWO Press.
- Goodman. L. A. 1996. « A single general method for the analysis of cross-classified data: Reconsiliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondance analysis and association analysis ». *Journal of the American Statistical Association*, vol. 91, p. 408–428.
- Gower, J. C. 1966. « Some distance properties of latent root and vector methods used in multivariate analysis ». *Biometrika*, vol. 53, p. 325–338.

- Gower, J. C. 1968. « Adding a point to vector diagrams in multivariate analysis ». *Biometrika*, vol. 55, p. 582–585.
- Gower, J. C. 1982. « Euclidean distance geometry ». *The Mathematical Scientist*, vol. 7, p. 1–14.
- Gower, J. C. (1992). « Generalized biplots ». *Biometrika*, vol. 79, p. 475–493.
- Gower, J. C. 1995. « A general theory of biplots ». In *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski (comp.), Oxford : Oxford Science Publications.
- Gower, J. C., et P. G. N. Digby. 1981. « Expressing complex relationships in two dimensions ». In *Interpreting Multivariate Data*, V. Barnett (comp.), New York : John Wiley & Sons.
- Gower, J. C., et D. J. Hand. 1996. *Biplots*, London : Chapman & Hall.
- Gower, J. C., et S. Harding. 1988. « Nonlinear biplots ». *Biometrika*, vol. 75, p. 445–455.
- Gower, J. C., et P. Legendre. 1986. « Metric and euclidean properties of dissimilarity coefficients ». *Journal of Classification*, vol. 3, p. 5–48.
- Greenacer, M. J. 1981. « Practical Correspondance Analysis ». In *Interpreting Multivariate Data*, V. Barnett (comp.), New York : John Wiley & Sons.
- Greenacer, M. J. 1984. *Theory and Applications of Correspondance Analysis*. Londres : Academic Press.
- Greenacer, M. J. 1986. « Discussion on paper by Gabriel and Oxdoroff ». In *Multidimensional Data Analysis*, J. De Leeuw, W. Heiser, J. Meulman et F. Critchley (comp.), Leinden : DSWO Press.
- Krzanowski, W. J. 1988. *Principle of Multivariate analysis*, Oxford : Oxford University Press.
- Lebart, L., Moineau, A., et Warwick, K.M. (1984). « Multivariate Descriptive statistical Analysis ». Wiley, New York.
- Legendre, L., et P. Legendre. 1979. *Écologie numérique. Tome I : Le traitement multiple des données écologiques*. Montréal : Les Presses de l'Université du Québec.
- Mardia, K. V., J. T. Kent et J. M. Bibby. 1979. *Multivariate Analysis*. London : Academic Press.
- Ramsay, J. O. 1982. *Multiscale II Manual*, Département de psychologie, Université McGill, Montréal.
- Young, G., et Householder, A.S (1938). « Discusion of a set of points in terms of their mutual distances ». *Psychometrika*, vol. 3, p. 19–22.