Géométrie des distances et analyse en composantes principales*

Jean-Claude HAUSMANN

1 Introduction

Soit E un ensemble fini. On considère une application $d : E \times E \to \mathbf{R}$ satisfaisant

a) $d(x,y) \geq 0$

b) $d(x,y) = d(y,x)$

c) $d(x,x) = 0$.

EXEMPLE TYPIQUE: Si (X,δ) est un espace métrique, et si $\alpha : E \to X$ est une application, alors d_α définie par $d_\alpha(x,y) = \delta(\alpha(x),\alpha(y))^2$ satisfait aux conditions a), b) et c) ci-dessus.

On va s'intéresser au problèmes suivants :

Problème 1 (réalisation euclidienne de d) : Existe-t-il $\alpha : E \to \mathbf{R}^p$ telle que $d = d_\alpha$? (c'est-à-dire $d(x,y) = \|\alpha(x) - \alpha(y)\|^2$) ? Si "oui", quel est le p minimal ?

Motivations : on imagine que la fonction d est une donnée expérimentale sur l'ensemble E. On cherche, avec α, à se représenter E géométriquement. De tels problèmes apparaissent naturellement en stéréochimie [CH] et peuvent intervenir en mécanique céleste [AC]. Le besoin de compression des données ou de les représenter graphiquement, motive la recherche d'une représentation approximative dans un espace plus petit. Ceci permettra aussi d'extraire l'information principale des données et d'éliminer d'éventuels bruits de fond ou imprécisions de mesures. Pour la littérature à ce sujet (analyse en composantes principales), voir [Go], [DY], [Jo]. Le problème est donc le suivant :

*Partie du cours de Chapitre Choisis de mathématiques donné à l'Université de Genève aux printemps 1998 et 2000. Je remercie Martin Hairer, Stefan Schneider et Thierry Vust pour leur intérêt et leur aide.

Problème 2 : Etant donné un entier k, trouver une application $\alpha' : E \to \mathbf{R}^k$ telle que $d_{\alpha'}(x,y)$ présente le minimum possible de distorsion avec d. Comment mesurer cette distorsion ?

Pour simplifier l'exposition, on suppose que $E := \{0,1,\ldots,n\}$. L'application d est donnée par une matrice symétrique $D := (d(i,j)) \in \mathcal{M}_{(n+1)\times(n+1)}(\mathbf{R})$. Cette matrice représente une forme bilinéaire symétrique $\mathcal{D}$ sur $\mathbf{R}^{n+1}$:

$$\mathcal{D}(x,y) = \langle Dx,y\rangle = \langle x,Dy\rangle.$$

Soit $\mathcal{H}$ l'hyperplan de $\mathbf{R}^{n+1}$ défini par

$$\mathcal{H} := \{x = (x_0,\ldots,x_n) \in \mathbf{R}^{n+1} \mid \sum_{i=0}^{n} x_i = 0\}. \tag{1}$$

La réponse au problème 1 ci-dessus a été trouvée sous diverses formes dans les années 1930. Nous allons démontrer la version d'I. Schoenberg [Sch]:

Théorème 1.1 (Théorème de Schoenberg (1935)) *Il existe $\alpha : E \to \mathbf{R}^p$ telle que $d = d_\alpha$ si et seulement si la restriction $\mathcal{D}_{\mathcal{H}}$ de la forme bilinéaire symétrique $\mathcal{D}$ à $\mathcal{H}$ est semi-définie négative. L'entier p minimal pour un tel α est le rang de $\mathcal{D}_{\mathcal{H}}$.*

Exemple : Prenons $n = 3$ et

$$D := \begin{pmatrix} 0 & a^2 & c^2 \\ a^2 & 0 & b^2 \\ c^2 & b^2 & 0 \end{pmatrix} \tag{2}$$

qui représente une forme bilinéaire symétrique sur $\mathbf{R}^3$. Une base de l'hyperplan $\mathcal{H}$ est donnée par les vecteurs $v_1 := (-1,1,0)$ et $v_2 := (-1,0,1)$. Dans cette base, la matrice $D_{\mathcal{H}}$ de $\mathcal{D}_{\mathcal{H}}$ est

$$\begin{aligned} D_{\mathcal{H}} &= \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & a^2 & c^2 \\ a^2 & 0 & b^2 \\ c^2 & b^2 & 0 \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} -2\,a^2 & -a^2 - c^2 + b^2 \\ -a^2 - c^2 + b^2 & -2\,c^2 \end{pmatrix} \end{aligned}$$

La forme bilinéaire symétrique $\mathcal{D}_{\mathcal{H}}$ est semi-négative définie si et seulement si les valeurs propres de $D_{\mathcal{H}}$ sont ≤ 0. Comme $\operatorname{trace} \mathcal{D}_{\mathcal{H}} \leq 0$, cela équivaut à $\det \mathcal{D}_{\mathcal{H}} \geq 0$:

$$0 \leq \det \mathcal{D}_{\mathcal{H}} = 4a^2c^2 - (b^2 - c^2 - a^2)^2 = -a^4 - b^4 - c^4 + 2a^2b^2 + 2a^2c^2 + 2b^2c^2.$$

On a donc démontré

Proposition 1.2 *Trois nombres $a,b,c \geq 0$ sont les côtés d'un triangle si et seulement si*

$$-a^4 - b^4 - c^4 + 2a^2b^2 + 2a^2c^2 + 2b^2c^2 \geq 0. \tag{3}$$

Remarque : L'aire du triangle $\Delta_{a,b,c}$ satisfait

$$16\,\mathrm{Aire}\,(\Delta_{a,b,c})^2 = -a^4 - b^4 - c^4 + 2a^2b^2 + 2a^2c^2 + 2b^2c^2$$

(exercice) ce qui donne une interprétation géométrique de la proposition 1.2. On peut comparer avec la formule de Héron:

$$\mathrm{Aire}\,(\Delta_{a,b,c}) = \sqrt{p(p-a)(p-b)(p-c)}$$

où $p = (a+b+c)/2$. Cette formule est équivalente à

$$16\,\mathrm{Aire}\,(\Delta_{a,b,c})^2 = (a+b+c)(a+b-c)(a-b+c)(-a+b+c)$$

Sans restreindre la généralité, on peut supposer que $c \geq b$ et $c \geq a$. On voit ainsi que la condition (3) est équivalente à ce que l'inégalité triangulaire $c \geq a + b$ qui est bien la condition pour que a, b et c soient les côtés d'un triangle.

2 Matrices de Gram

Considérons un “nuage" $R := \{R_0, \ldots, R_n\}$ de $n+1$ points dans $\mathbf{R}^p$. On écrit R_j comme un vecteur colonne (r_{ij}) ce qui donne une matrice

$$R := \begin{pmatrix} r_{10} & \cdots & \cdots & r_{1n} \\ \vdots & & & \vdots \\ r_{p0} & \cdots & \cdots & r_{pn} \end{pmatrix} \in \mathcal{M}_{p\times(n+1)} \tag{4}$$

(on imagine que n est plus grand que p). La matrice R est la matrice de l'application linéaire $r : \mathbf{R}^{n+1} \to \mathbf{R}^p$ définie par $r(x_0, \ldots, x_n) = \sum_{i=0}^n x_i R_i$.

On considère les *matrices de Gram*

$$R^T R \in \mathcal{M}_{(n+1)\times(n+1)} \qquad \text{et} \qquad R\,R^T \in \mathcal{M}_{p\times p}.$$

Ce sont des matrices symétriques. Les coefficients de $R^T R$ sont les produits scalaires $\langle R_i, R_j\rangle$ des vecteurs colonne de R tandis que ceux de $R\,R^T$ sont les produits scalaires des vecteurs ligne.

Proposition 2.1 *a)* $\operatorname{rang} R^T R = \operatorname{rang} R R^T = \operatorname{rang} r$.

b) Si $n+1 \geq p$, les polynômes caractéristiques satisfont

$$P_{R^T R}(X) = X^{n+1-p} P_{R R^T}(X)$$

En particulier, $R^T R$ et $R R^T$ ont les mêmes valeurs propres non-nulles.

c) ces valeurs propres non-nulles sont toutes positives.

d) Si $v \in \mathbf{R}^{n+1}$ est un vecteur propre de $R^T R$ pour la valeur propre $\lambda \neq 0$, alors $r(v)$ est un vecteur propre de $R R^T$ pour la même valeur propre.

Nous allons prouver 2.1 dans le langage des applications linéaires. Rappelons que la transposée R^T de R est la matrice de l'application $r^\sharp$, adjointe à r pour le produit scalaire standard, définie par l'équation $\langle r(x),y \rangle = \langle x, r^\sharp(y) \rangle$. Le lemme suivant découle directement de cette équation.

Lemme 2.2 *i)* $\ker r^\sharp = (\operatorname{image} r)^\perp$.

ii) $\operatorname{image} r^\sharp = (\ker r)^\perp$. $\square$

Preuve de 2.1 Supposons que $r^\sharp \circ r(v) = \lambda v$ avec $v \neq 0$ et $\lambda \neq 0$. On a donc $r(v) \neq 0$ et

$$r(r^\sharp \circ r(v)) = \lambda r(v) = r \circ r^\sharp (r(v))).$$

Cela prouve l'assertion d). Cela prouve également que $r \circ r^\sharp$ et $r^\sharp \circ r$ ont les mêmes valeurs propres non-nulles avec même multiplicité (ce qui démontre le point a)). La formule b) s'en déduit en comptant la dimension des noyaux de $r \circ r^\sharp$ et $r^\sharp \circ r$ par le lemme 2.2. Enfin, la positivité des valeurs propres se voit par

$$\lambda \langle v,v \rangle = \langle \lambda v, v \rangle = \langle r^\sharp \circ r(v), v \rangle = \langle r(v), r(v) \rangle. \ \square$$

Proposition 2.3 *Soit $M \in \mathcal{M}_{n \times n}$ une matrice symétrique. Les deux conditions suivantes sont équivalentes :*

– *il existe $R \in \mathcal{M}_{p \times n}$ telle que $M = R^T R$.*

– $\operatorname{rang} M \leq p$ *et les valeurs propres de M sont ≥ 0.*

Preuve: Le fait que 2) implique 1) provient de la proposition 2.1. Réciproquement, soient $\lambda_1 \geq \cdots \geq \lambda_n$ les valeurs propres de M et $P = (P_1 \cdots P_n)$ une $(n \times n)$-matrice dont les vecteurs colonne forment une base orthonormale et $MP_i = \lambda_i P_i$. Définissons $R_i := \sqrt{\lambda_i} P_i$ et notons $R := (R_1 \cdots R_p) \in \mathcal{M}_{n \times p}$. Comme $P^T M P = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ et que $\lambda_i = 0$ pour $i \geq p+1 > \operatorname{rang} M$, on a

$$M = P \operatorname{diag}(\lambda_1, \ldots, \lambda_p, 0 \ldots, 0) P^T = R R^T. \ \square$$

3 Forme de Gram et forme de Schoenberg

Soit

$$R := \begin{pmatrix} r_{10} & \cdots & \cdots & r_{1n} \\ \vdots & & & \vdots \\ r_{p0} & \cdots & \cdots & r_{pn} \end{pmatrix} \in \mathcal{M}_{p\times(n+1)} \tag{5}$$

une matrice définissant, comme au § 2, un nuage de $n+1$ points dans $\mathbf{R}^p$. La matrice de Gram R^TR définit une forme bilinéaire symétrique $\mathcal{GR}$ sur $\mathbf{R}^{n+1}$. Observons que $\mathcal{GR}(x,y)$ est le produit scalaire dans $\mathbf{R}^p$ de $r(x)$ avec $r(y)$:

$$\mathcal{GR}(x,y) = x^TR^TRy = \langle Rx,Ry\rangle = \langle r(x),r(y)\rangle \tag{6}$$

La forme $\mathcal{GR}$ est donc semi-définie positive.

Nous allons comparer la forme $\mathcal{GR}$ avec la forme $\mathcal{D}$ donnée par la matrice des distances au carré

$$D = (d_{ij}) \qquad \text{avec} \qquad d_{ij} := \|r_i - r_j\|^2.$$

La *forme de Schoenberg* $\mathcal{S}$ du nuage R est la forme symétrique $\mathcal{S} := -\frac{1}{2}\mathcal{D}$

La proposition principale de ce paragraphe affirme que les restrictions $\mathcal{GR}_{\mathcal{H}}$ et $\mathcal{S}_{\mathcal{H}}$ de $\mathcal{GR}$ et $\mathcal{S}$ à l'hyperplan $\mathcal{H}$ (voir (1)) coïncident:

Proposition 3.1 $\mathcal{GR}_{\mathcal{H}} = \mathcal{S}_{\mathcal{H}}$.

Preuve: Les coefficients s_{ij} de la matrice S de la forme de Schoenberg sont

$$s_{ij} = -\frac{1}{2}\|R_i - R_j\|^2 = -\frac{1}{2}\|R_i\|^2 - \frac{1}{2}\|R_j\|^2 + \langle R_i,R_j\rangle. \tag{7}$$

Considérons les vecteurs colonne à $(n+1)$-lignes:

$$]R[:= \begin{pmatrix} \|R_0\|^2 \\ \vdots \\ \|R_n\|^2 \end{pmatrix} \quad \text{et} \quad \mathbf{1} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Les équations (7) s'écrivent matriciellement

$$S = R^TR - \frac{1}{2}]R[\,\mathbf{1}^T - \frac{1}{2}\mathbf{1}\,]R[^T. \tag{8}$$

Si $x,y \in \mathcal{H}$, on a $x^T\mathbf{1} = 0$ et $\mathbf{1}^Ty = 0$ et donc $x^TSy = x^TR^TRy$. $\square$

4 Preuve du théorème de Schoenberg

On vient de démontrer que la forme de Schoenberg $\mathcal{S}_{\mathcal{H}}$, restreinte à $\mathcal{H}$, est semi-définie positive (proposition 3.1). Donc $\mathcal{D}_{\mathcal{H}}$ est semi-définie négative si D est la matrice des distances au carré pour un nuage de points dans $\mathbf{R}^p$.

Réciproquement, supposons que la $\mathcal{S}_{\mathcal{H}}$ est semi-définie positive. Prenons, comme base de $\mathcal{H}$ les vecteurs

$$\mathcal{V} := \{v_1, \dots, v_n\}$$

où $v_i = e_i - e_0$ (e_i = i$^{\text{ème}}$ vecteur de la base standard de $\mathbf{R}^{n+1}$). Désignons par $S := (s_{ij}$ $(1 \leq i,j \leq n)$ la matrice de $\mathcal{S}_{\mathcal{H}}$ dans la base $\mathcal{V}$ ($s_{ij} = \mathcal{S}(v_i,v_j)$). On a

$$s_{ij} = \frac{1}{2}[d_{0j} + d_{i0} - d_{ij}] \tag{9}$$

Comme S est semi-définie positive, il existe, par la proposition 2.3, une matrice $R = (R_1 \cdots R_n) \in \mathcal{M}_{p\times n}$ ($p = \text{rang}\,(S)$) telle que $S = RR^T$. La matrice$R$ constitue donc un nuage de n points dans $\mathbf{R}^p$. Posons encore $R_0 := 0 \in \mathbf{R}^p$. Nous affirmons que *$D$ est la matrice des distances au carré des points $R_0, \dots, R_n$*. En effet :

$$\begin{aligned} \|R_j - R_i\|^2 &= \|R_j\|^2 + \|R_i\|^2 - 2\langle R_j,R_i\rangle = s_{ii} + s_{jj} - 2s_{ij} = \\ &= d_{0i} + d_{0j} - d_{0j} - d_{0i} + d_{ij} = d_{ij} \end{aligned}$$

pour $i,j \geq 1$ et

$$\|R_i - R_0\|^2 = \|R_i\|^2 = s_{ii} = d_{0i}.$$

Il reste à démontrer que $p_{\min} = \text{rang}\,\mathcal{D}_{\mathcal{H}}$, où $p_{\min}$ est le minimum des entiers p tels qu'il existe un nuage de $(n+1)$ points dans $\mathbf{R}^p$ dont la matrice des distances au carré est D. Comme on vient de construire un tel nuage pour lequel $p = \text{rang}\,\mathcal{S}_{\mathcal{H}} = \text{rang}\,\mathcal{D}_{\mathcal{H}}$, on a $p_{\min} \leq \text{rang}\,\mathcal{D}_{\mathcal{H}}$.

D'autre part, si R est un tel nuage dans $\mathbf{R}^p$, on a

$$\text{rang}\,\mathcal{D}_{\mathcal{H}} = \text{rang}\,\mathcal{G}\mathcal{R}_{\mathcal{H}} \leq \text{rang}\,\mathcal{G}\mathcal{R} = \text{rang}\,r^\sharp \circ r = \text{rang}\,r \leq p$$

d'où $p_{\min} \geq \text{rang}\,\mathcal{D}_{\mathcal{H}}$. $\square$

5 Compressions k-dimensionnelles

Soit $R := (R_0, \dots, R_n) \in \mathcal{M}_{p\times(n+1)}$ représentant un nuage de $n+1$ points dans $\mathbf{R}^p$. On aimerait trouver une matrice $R' := (R'_0, \dots, R'_n) \in \mathcal{M}_{k\times(n+1)}$

représentant un nuage de $n+1$ points dans $\mathbf{R}^k$ ($k \leq p$) qui soit, à isométrie affine près, le "plus proche possible" de R.

La notion de proximité entre deux nuages T et T' de $\mathbf{R}^p$ se définit par la fonction que nous appellerons *écart*:

$$\mathrm{Ec}\,(T,T') = \sum_{i=0}^{n} \|T_i - T'i\|^2. \tag{10}$$

Pour $k \leq p$, on définit $\mathrm{Ec}_{\min}(R,k)$ comme étant l'infimum (le point b) du théorème 5.1 ci-dessous prouvera que c'est un minimum) de $\mathrm{Ec}\,(R,R')$ pour R' un nuage dans un k-plan affine de $\mathbf{R}^p$ (on parlera de *nuage k-dimensionnel*).

Rappelons que R est la matrice d'une application linéaire $r : \mathbf{R}^{n+1} \to \mathbf{R}^p$. Notons $r_{\mathcal{H}} : \mathcal{H} \to \mathbf{R}^p$ la restriction de r à l'hyperplan $\mathcal{H}$. On introduit les notations suivantes:

• $b(R) := \dfrac{1}{n+1} \displaystyle\sum_{i=0}^{n+1} R_i$. C'est le *barycentre* de R.

• Si Π est un k-plan affine, on note R_Π le nuage obtenu par projection orthogonale de R sur Π.

D'autre part, soit V est un espace vectoriel de dimension d muni d'un produit scalaire et soit f est un endomorphisme autoadjoint de V. Soient $\lambda_1 \geq \cdots \geq \lambda_d$ les valeurs propres de d. Un *k-plan spectral* de f est un sous-espace vectoriel de dimension k de V engendré par les vecteurs propres de k des plus grandes valeurs propres de f. Si $\lambda_k > \lambda_{k+1}$, alors f admet un unique k-plan spectral. A l'autre extrême, tout k-plan est un k-plan spectral de id_V. On note encore $\mathrm{trace}_{\leq k} := \sum_{i=1}^{k} \lambda_i$.

Nous démontrerons le théorème suivant :

Théorème 5.1 (Théorème de compression) *a) Soit R' un nuage k-dimensionnel satisfaisant* $\mathrm{Ec}\,(R,R') = \mathrm{Ec}_{\min}(R,k)$. *Alors $R' = R_\Pi$ où Π est un k-plan affine passant par $b(R)$.*

b) Soit Π le k-plan affine passant par $b(R)$ et parallèle à un k-plan spectral de $r_{\mathcal{H}} \circ r_{\mathcal{H}}^\sharp$. Alors $\mathrm{Ec}\,(R,R_\Pi) = \mathrm{Ec}_{\min}(R,k)$.

c) $\mathrm{Ec}_{\min}(R,k) = \mathrm{trace}(r_{\mathcal{H}} \circ r_{\mathcal{H}}^\sharp) - \mathrm{trace}_{\leq k}(r_{\mathcal{H}} \circ r_{\mathcal{H}}^\sharp)$.

Remarques : 1) Soit $i : \mathcal{H} \to \mathbf{R}^{n+1}$ l'inclusion. Son adjoint $i^\sharp$ est la projection orthogonale π de $\mathbf{R}^{n+1}$ sur $\mathcal{H}$. On a $r_{\mathcal{H}} = r \circ i$ et donc

$$r_{\mathcal{H}} \circ r_{\mathcal{H}}^\sharp = r \circ i \circ \pi \circ r^\sharp.$$

On a $i\circ\pi(x) = x - \langle N,x\rangle N$ où $N = \frac{1}{\sqrt{n+1}}\mathbf{1}$. La matrice de $i\circ\pi$ est donc $I - NN^T = I - \frac{1}{n+1}\mathbf{1}\mathbf{1}^T$. La matrice de $r_{\mathcal{H}}\circ r^\sharp_{\mathcal{H}}$ est ainsi

$$R(I - \frac{1}{n+1}\mathbf{1}\mathbf{1}^T)R^T.$$

2) La situation $b(R) = 0$, que l'on peut obtenir par une translation, est favorable. En effet, cela veut dire que les vecteurs ligne de R sont dans $\mathcal{H}$, donc l'image de $r^\sharp$ est incluse dans $\mathcal{H}$. D'où $\pi\circ r^\sharp = r^\sharp$ et l'on peut, dans le théorème de compression, remplacer $r_{\mathcal{H}}\circ r^\sharp_{\mathcal{H}}$ par $r\circ r^\sharp$. Observons que le nuage $\overline{R}$ obtenu par translation de R et tel que $b(\overline{R}) = 0$ satisfait :

$$\overline{R} = R(I - \frac{1}{n+1}\mathbf{1}\mathbf{1}^T).$$

3) Observons que le plan Π du point b) du théorème de compression est unique si $\lambda_k > \lambda_{k+1}$.

Les parties b) et c) théorème 5.1 seront démontrées au § 8. Nous donnons ci-dessous la preuve d'une partie de a).

Preuve partielle de a) : Soit R'' un nuage dans un k-plan affine Π et soit R' la projection orthogonale de R sur Π. Il est clair que $\mathrm{Ec}\,(R,R') \leq \mathrm{Ec}\,(R,R'')$.

En effectuant au besoin une translation, on peut supposer que le barycentre de R est en 0. Soit

$$b' := b(R') = \frac{1}{n+1}\sum_{i=0}^{n} R'_i$$

le barycentre de R'. Définissons un nouveau nuage $\tilde{R}'$ par $\tilde{R}'_i = R'_i - b'$. Si R' est k-dimensionnel, $\tilde{R}'$ l'est aussi et on a $b(\tilde{R}') = 0$. Nous allons montrer que $\mathrm{Ec}\,(R,\tilde{R}') \leq \mathrm{Ec}\,(R,R')$. En effet :

$$\sum_{i=0}^{n} \|R_i - R'_i\|^2 = \sum_{i=0}^{n} \|R_i - \tilde{R}'_i + b'\|^2 = \sum_{i=0}^{n} \|R_i - \tilde{R}'_i\|^2 + (n+1)\|b'\|^2 + 2\sum_{i=0}^{n}\langle R_i - \tilde{R}'_i, b'\rangle.$$

Mais

$$\sum_{i=0}^{n}\langle R_i - \tilde{R}'_i, b'\rangle = \langle \underbrace{\sum_{i=0}^{n}(R_i - \tilde{R}'_i)}_{=0}, b'\rangle = 0.$$

D'où

$$\mathrm{Ec}\,(R,R') = \mathrm{Ec}\,(R,\tilde{R}') + (n+1)\|b'\|^2. \ \square$$

6 Sur les valeurs propres des endomorphismes auto-adjoints

Les énoncés de ce paragraphes sont classiques (voir [Fr, chapitre 6]).

Soit V un espace vectoriel muni d'un produit scalaire $\langle\,,\rangle$. On dénote par $S(V)$ la *sphère de rayon* 1 dans V:

$$S(V) := \{v \in V \mid \|v\| = 1\}.$$

Proposition 6.1 (Principe du maximum de Rayleigh) *Soit $f : V \to V$ un endomorphisme auto-adjoint. Soit $\lambda_{\max}$ (respectivement: $\lambda_{\min}$) le maximum ((respectivement: le minimum) des valeurs propres de f. Alors:*

$$\lambda_{\max} = \max_{v\in S(V)} \langle v,f(v)\rangle \qquad \textit{et} \qquad \lambda_{\min} = \min_{v\in S(V)} \langle v,f(v)\rangle$$

De plus, $\langle v,f(v)\rangle = \lambda_{\max}$ (respectivement: $\langle v,f(v)\rangle = \lambda_{\min}$) si et seulement si v est un vecteur propre pour $\lambda_{\max}$ (respectivement: $\lambda_{\min}$).

Preuve: Soit $\mathcal{P} = (p_1,\ldots,p_n)$ une base orthonormale de vecteurs propres pour f $(f(p_i) = \lambda_i p_i)$. Soit $v = \sum_{i=1}^n v_i p_i$ avec $v \in S(V)$ (comme $\mathcal{P}$ est orthonormale, on a $\sum_{i=1}^n v_i^2 = 1$). On a

$$\langle v,f(v)\rangle = \sum_{i=1}^n v_i^2\lambda_i \le \lambda_{\max}\sum_{i=1}^n v_i^2 = \lambda_{\max}. \tag{11}$$

L'égalité $\langle v,f(v)\rangle = \lambda_{\max}$ a lieu si et seulement si $v_i = 0$ pour $\lambda_i \neq \lambda_{\max}$. Cela est équivalent à dire que v est dans l'espace propre de $\lambda_{\max}$.

L'argument pour $\lambda_{\min}$ est complètement analogue. □

Soit $G(k,V)$ l'ensemble des k-plans de V (Grassmannienne). Si $\Pi \in G(k,V)$, on dénote $\Pi^\perp \in G(n-k,V)$ son complément orthogonal.

Proposition 6.2 (min-max principe de Courant) *Soient $\lambda_1 \ge \cdots \ge \lambda_n$ les valeurs propres d'un endomorphisme auto-adjoint $f : V \to V$. Alors, pour $k < n$,*

$$\lambda_{k+1} = \min_{\Pi\in G(k,V)} \left[\max_{u\in S(\Pi^\perp)} \langle u,f(u)\rangle\right].$$

Preuve: Définissons la fonction $\phi : G(k,V) \to \mathbf{R}$ par

$$\phi(\Pi) := \max_{u\in S(\Pi^\perp)} \langle u,f(u)\rangle.$$

Remarquons que ce maximum existe puisque $\langle u, f(u) \rangle$ est une fonction continue de u et que $S(\Pi^\perp)$ est compact. Par la proposition 6.1, on a,

$$\phi(\Pi_k) = \lambda_{k+1} \tag{12}$$

où Π_l est le l-plan engendré par les vecteurs propres p_j pour $j \leq l$.

Soit $\Pi \in G(k,V)$ et soit

$$w = \sum_{j=1}^{k+1} c_j p_j \in S(\Pi_{k+1} \cap \Pi^\perp). \tag{13}$$

Remarquons qu'un tel w existe puisque $\dim(\Pi_{k+1} \cap \Pi^\perp) \geq 1)$. On a

$$\phi(\Pi) \geq \langle w, f(w) \rangle = \sum_{i=1}^{k+1} c_i^2 \lambda_i \geq \lambda_{k+1} \sum_{i=1}^{k+1} c_i^2 = \lambda_{k+1}. \tag{14}$$

La proposition 6.2 découle de (12) et (14). $\square$

Soit W un sous espace de V. Si f est un endomorphisme de V, on définit $f_W : W \to W$ comme $f_W := p_W \circ f_{|W}$, où p_W est la projection orthogonale sur W. Si f est auto-adjoint, on vérifie que f_W l'est aussi.

Proposition 6.3 (Principe d'insertion.) *Soit $f : V \to V$ un endomorphisme auto-adjoint et W un sous espace de codimension d de V. Soient $\lambda_1 \geq \cdots \geq \lambda_n$ les valeurs propres de f et $\mu_1 \geq \cdots \geq \mu_{n-d}$ celles de f_W. Alors, pour $k = 1, \ldots, n-d$, on a*

$$\lambda_k \geq \mu_k \quad \text{et} \quad \mu_k \geq \lambda_{k+d}.$$

De plus, si $\lambda_k = \mu_k$ ou si $\mu_k = \mu_{k+d}$, alors W contient un vecteur propre de f pour la valeur λ_k.

PREUVE: Considérons les fonctions

$$\phi, \phi^W : \bigcup_{r=1}^{n} G(r,V) \to \mathbf{R}$$

définies par

$$\phi(\Pi) := \max\{\langle v, f(v) \rangle \mid v \in S(\Pi^\perp)\}$$

et

$$\phi^W(\Pi) := \max\{\langle v, f(v) \rangle \mid v \in S(\Pi^\perp) \cap W\}.$$

Si $\Pi \in G(k-1,V)$ on a

$$\phi(\Pi) \geq \phi^W(\Pi) = \phi^W((p_W(\Pi)).$$

Soit $\Pi_{k-1} \in G(k-1,V)$ engendré par les vecteurs propres pour $\lambda_1,\ldots,\lambda_{k-1}$; notons $l := \dim p_W(\Pi_{k-1})$. On a

$$\lambda_k = \phi(\Pi_{k-1}) \geq \phi^W(\Pi_{k-1}) = \phi^W(p_W(\Pi_{k-1})) \geq \min_{\Pi \in G(l,W)} \phi^W(\Pi) = \mu_l \geq \mu_k.$$

D'autre part, soit $\Pi_{k-1}^W \in G(k-1,W)$ engendré par les vecteurs propres pour les valeurs propres $\mu_1,\ldots,\mu_{k-1}$ de f_W. On a

$$\mu_k = \phi^W(\Pi_{k-1}^W) = \phi(\Pi_{k-1}^W \oplus W^\perp) \geq \min_{\Pi \in G(k-1+d,V)} \phi(\Pi) = \lambda_{k+d}.$$

Comme $\langle f(v),v\rangle = \langle f^W(v),v\rangle$ si $v \in W$, la dernière assertion du principe d'insertion découle du principe de Rayleigh. $\square$

Le principe d'insertion a plusieurs corollaires.

Corollaire 6.4 *Soit M une matrice symétrique et $\lambda_{\min}$ et $\lambda_{\max}$ ses valeurs propres extrêmes. Alors, pour tout i, le coefficient diagonal M_{ii} est dans l'intervalle* $[\lambda_{\min},\lambda_{\max}]$.

PREUVE: On applique le principe d'insertion à l'espace W de dimension 1 engendré par le ie vecteur de base e_i. La matrice de f_W est (M_{ii}). $\square$

Corollaire 6.5 (Principe des mineurs) *Soit A une $(n \times n)$-matrice symétrique représentant une forme bilinéaire symétrique $\mathcal{A}$ sur $\mathbf{R}^n$. On désigne par A_k le mineur de M formé des k premières lignes et colonnes. Alors, $\mathcal{A}$ est définie positive si et seulement si* $\det A_k > 0$ *pour* $k = 1,\ldots,n$.

PREUVE: $\mathcal{A}$ est définie positive si et seulement si les valeurs propres de A sont toutes positives. Comme A_k est la matrice de la restriction de $\mathcal{A}$ à $\mathbf{R}^k$, la condition $\det A_k > 0$ est évidemment nécessaire.

Réciproquement, supposons que $\det A_k > 0$ pour tout k et montrons que les valeurs propres de A_m sont toutes positives par récurrence sur m. La récurrence démarre banalement avec $m = 1$. Si toutes les valeurs propres de A_{m-1} sont positives, celles de A_m, par le principe d'insertion, le sont toutes sauf peut-être la plus petite. Comme $\det A_m > 0$, cette plus petite valeur propre sera aussi positive. $\square$

La conséquence du principe qui nous intéresse pour la démonstration du théorème 5.1 est la suivante.

Proposition 6.6 *Soit $f : V \to V$ un endomorphisme auto-adjoint. Soient $\lambda_1 \geq \ldots \geq \lambda_n$ les valeurs propres de f. Soit $\Pi \in G(k,V)$. Alors*

$$\operatorname{trace} f_\Pi \leq \sum_{i=1}^{k} \lambda_i = \operatorname{tr}_{\leq k}(f) \tag{15}$$

avec égalité dans (15) si et seulement si Π est un k-plan spectral de f. (voir § 5; un tel Π est unique si $\lambda_k > \lambda_{k+1}$). De même,

$$\operatorname{trace} f_\Pi \geq \sum_{i=n-k+1}^{n} \lambda_i = -\operatorname{tr}_{\leq k}(-f) \tag{16}$$

avec égalité si et seulement si Π est un k-plan spectral de $-f$. (un tel Π est unique si $\lambda_{n-k} > \lambda_{n-k+1}$).

PREUVE: Soient $\mu_1 \geq \cdots \geq \mu_k$ les valeurs propres de f_Π. Par le principe d'insertion

$$\operatorname{trace} f_\Pi = \sum_{i=1}^{k} \mu_i \leq \sum_{i=1}^{k} \lambda_i.$$

Si l'égalité se produit, on doit avoir $\mu_i = \lambda_i$ pour $i = 1,\ldots,k$. D'après la dernière assertion du principe d'insertion, on a que Π est un k-plan spectral de f.

L'inégalité (16) se démontre de la même manière. □

7 Compressions k-dimensionnelles vectorielles

Soit R un nuage de $n+1$ points dans $\mathbf{R}^p$. Comme premier pas pour la démonstration des parties b) et c) du théorème 5.1, on s'intéresse à des nuages R' approximant R qui sont dans des k-plans vectoriels (donc des éléments de la grassmannienne $G(k,\mathbf{R}^p)$. On définit pour ce faire $\mathrm{Ec}^{\mathrm{vec}}_{\min}(R,k)$ comme étant l'infimum (on verra que c'est un minimum) de $\mathrm{Ec}\,(R,R')$ pour R' un nuage dans un k-plan vectoriel de $\mathbf{R}^p$. Si $W \in G(k,\mathbf{R}^p)$, on note R_W l'image de R par la projection orthogonale sur W.

Le but de ce paragraphe est de démontrer le théorème suivant. On comparera son énoncé avec celui du théorème 5.1: la différence est qu'ici, $r_{\mathcal{H}}$ est remplacée par r.

Théorème 7.1 (Théorème de compression vectorielle) .

a) Soit W un k-plan spectral de $r \circ r^\sharp$. Alors $\mathrm{Ec}\,(R,R_W) = \mathrm{Ec}^{\mathrm{vec}}_{\min}(R,k)$.

b) Tout nuage R' dans un k-paln vectoriel satisfaisant $\mathrm{Ec}\,(R,R') = \mathrm{Ec}^{\mathrm{vec}}_{\min}(R,k)$ *est construit comme celui du point a). En particulier, si* $\lambda_k(r\circ r^\sharp) > \lambda_{k+1}(r\circ r^\sharp)$, *$R'$ est unique.*

c) $\mathrm{Ec}^{\mathrm{vec}}_{\min}(R,k) = \mathrm{trace}(r\circ r^\sharp) - \mathrm{trace}_{\leq k}(r\circ r^\sharp)$.

Preuve: Soit $\Pi \in G(k,\mathbf{R}^p)$. On décompose $\mathbf{R}^p = \Pi \oplus \Pi^\perp$ et on écrit $r = (\alpha,\beta)$. On a $r^\sharp\circ r = \alpha^\sharp\circ\alpha + \beta^\sharp\circ\beta$. Soit $r' = (\alpha,0)$; c'est l'application linéaire associée au nuage k-dimensionnel R_Π. On a

$$\mathrm{Ec}\,(R,R_\Pi) = \sum_{i=0}^{n} \|\beta(e_i)\|^2 = \mathrm{trace}(\beta^\sharp\circ\beta) = \mathrm{trace}\,(\beta\circ\beta^\sharp).$$

Mais $\beta\circ\beta^\sharp = (r\circ r^\sharp)_{\Pi^\perp}$. Par la proposition 6.6,

$$\mathrm{trace}\,(r\circ r^\sharp)_{\Pi^\perp} \geq \sum_{i=k+1}^{p} \lambda_i, \tag{17}$$

où $\lambda_1 \geq \cdots \geq \lambda_p$ sont les valeurs propres de $r\circ r^\sharp$. On en déduit que

$$\mathrm{Ec}^{\mathrm{vec}}_{\min}(R,k) \geq \sum_{i=k+1}^{p} \lambda_i.$$

Comme $\mathrm{Ec}\,(R,R_W) = \sum_{i=k+1}^{p} \lambda_i$ cela démontre a) et c). La proposition 6.6 dit que l'égalité dans (17) a lieu si et seulement si Π est obtenu de la même manière que W, ce qui démontre b). $\square$

Exemple dans le plan : On se donne un nuage

$$R := \begin{pmatrix} a_0 & a_1 & \cdots & a_n \\ b_0 & b_1 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

et on cherche une droite vectorielle W telle que $\mathrm{Ec}\,(R,R_W) = \mathrm{Ec}^{\mathrm{vec}}_{\min}(R,1)$. La droite W est d'équation $y = x\mathrm{tg}(\alpha)$.

On peut trouver α en résolvant un problème d'extremum de la manière suivante. Soit $\tilde{R}$ le nuage obtenu de R par rotation d'angle $-\alpha$:

$$\tilde{R} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a_0 & a_1 & \cdots & a_n \\ b_0 & b_1 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a\cos\alpha + b\sin\alpha \\ -a\sin\alpha + b\cos\alpha \end{pmatrix}.$$

On veut donc minimiser la fonction $A(\alpha) := \|-a\sin\alpha + b\cos\alpha\|^2$. On a

$$A(\alpha) = \|a\|^2\sin^2\alpha + \|b\|^2\cos^2\alpha - \langle a,b\rangle \sin 2\alpha$$

d'où
$$A'(\alpha) = \|a\|^2 \sin 2\alpha - \|b\|^2 \sin 2\alpha - 2\langle a,b\rangle \cos 2\alpha.$$
L'équation $A'(\alpha) = 0$ est donc équivalente à
$$\mathrm{tg}(2\alpha) = \frac{2\langle a,b\rangle}{\|a\|^2 - \|b\|^2}. \tag{18}$$
Comparons ce résultat avec celui donné par le théorème 7.1. On a
$$R^T R = \begin{pmatrix} \|a\|^2 & \langle a,b\rangle \\ \langle a,b\rangle & \|b\|^2 \end{pmatrix}$$
dont les valeurs propres sont
$$\frac{\|a\|^2 + \|b\|^2 \pm \sqrt{(\|a\|^2 - \|b\|^2)^2 + 4\langle a,b\rangle^2}}{2}.$$
Le vecteur propre $(1,\mathrm{tg}(\tilde{\alpha}))$ pour la grande valeur propre satisfait
$$\mathrm{tg}(\tilde{\alpha}) = \frac{2\langle a,b\rangle}{(\|a\|^2 - \|b\|^2)[1 + \sqrt{1 + \frac{4\langle a,b\rangle^2}{(\|a\|^2-\|b\|^2)^2}}} = \frac{\mathrm{tg}(2\alpha)}{1 + \sqrt{1 + \mathrm{tg}^2(2\alpha)}}. \tag{19}$$
On a donc bien $\tilde{\alpha} = \alpha$, ce que l'on peut vérifier dans (19) via la formule plus classique
$$\mathrm{tg}(2\alpha) = \frac{2\,\mathrm{tg}(\alpha)}{1 - \mathrm{tg}^2(\alpha)}.$$

8 Preuve du théorème de compression

On a déjà montré au paragraphe 7 que le nuage R' tel que $\mathrm{Ec}\,(R,R') = \mathrm{Ec}_{\min}(R,k)$ est, s'il existe, à chercher parmi les projections orthogonales sur un k-plan passant par le barycentre $b(R)$ de R. Via une translation, on peut supposer que $b(R) = 0$ (observons qu'une translation ne change pas $r_{\mathcal{H}}$).

Si $b(R) = 0$, on est dans la situation de chercher un k-plan vectoriel qui, par le théorème 7.1, est celui engendré par les vecteurs propres des k plus grandes valeurs propres de $r^\sharp \circ r$. Mais, puisque, $r(1,\ldots,1) = R_0 + \cdots + R_n = 0$, on a $r^\sharp_{\mathcal{H}} \circ r_{\mathcal{H}} = r^\sharp \circ r$. Avec le reste du théorème 7.1, ceci démontre le théorème 5.1. $\square$

Remarque : Le k-plan solution du problème de compression n'est pas, en général, parallèle au k-plan solution du problème de compression vectorielle. En effet, les vecteurs propres de $r^\sharp_{\mathcal{H}} \circ r_{\mathcal{H}}$ sont, en général, différents de ceux de $r^\sharp \circ r$. Pour un exemple, voir la fin du § 10.

9 Coordonnées de nuages approximants

Soit $R := \mathcal{M}_{p\times(n+1)}$ un nuage de $n+1$ points dans $\mathbf{R}^p$. Un *k-nuage approximant* pour R est un nuage R' dans $\mathbf{R}^k$ tel qu'il existe une isométrie $\beta : \mathbf{R}^k \to \mathbf{R}^p$ telle que $\mathrm{Ec}\,(R,\beta(R')) = \mathrm{Ec}_{\min}(R,k)$. Trouver un k-nuage approximant pour R constitue donc la réponse au problème 2 de l'introduction, lorsque la réponse au problème 1 est “oui".

D'après le théorème 5.1, on peut trouver un k-nuage approximant pour R de la façon suivante.

1. Translation R en $\bar{R}$ de manière que $b(\bar{R}) = 0$. Matriciellement : $\bar{R} = RQ$ où $Q := I - \frac{1}{n+1}\mathbf{1}\mathbf{1}^T$.
2. Chercher des vecteurs propres orthogonaux et unitaires $p_1, \dots p_k$ pour k des plus grandes valeurs propres de $\bar{R}\bar{R}^T$. On les regarde comme vecteurs colonne d'une matrice $P \in \mathcal{M}_{p\times k}$.
3. Un k nuage approximant R' sera donné par
$$R' = P^T\,\bar{R} \in \mathcal{M}_{k\times(n+1)}.$$

Exemple : Prenons $R := \frac{1}{\sqrt{2}} I \in \mathcal{M}_{(n+1)\times(n+1)}$. Cela correspond à $n+1$ points à distances mutuelles 1. On a donc

$$\bar{R} = \frac{1}{\sqrt{2}}\,Q = \frac{1}{(n+1)\sqrt{2}} \begin{pmatrix} n & -1 & -1 & \cdots & -1 \\ -1 & n & -1 & \cdots & -1 \\ & & \ddots & & \\ -1 & -1 & \cdots & & n \end{pmatrix}$$

On a $Q^TQ = Q^2 = Q$ puisque Q est la projection sur $\mathcal{H}$. Les valeurs propres de Q sont donc 1, avec espace propre $\mathcal{H}$, et 0. Si $P = (P_1, \dots, P_n) \in \mathcal{M}_{(n+1)\times n}$ est une base orthonormale de $\mathcal{H}$ alors $R' = \frac{1}{\sqrt{2}}P^TQ$ sera un nuage approximant. Comme $p_i \in \mathcal{H}$, on a $P^TQ = P^T$. On peut, par exemple, prendre pour P^T la matrice

$$P^T := \begin{pmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{-1}{\sqrt{1\cdot 2}} & 0 & \cdots & & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{-2}{\sqrt{2\cdot 3}} & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ \frac{1}{\sqrt{n\cdot(n+1)}} & \frac{1}{\sqrt{n\cdot(n+1)}} & \cdots & & & \frac{-(n+1)}{\sqrt{n\cdot(n+1)}} \end{pmatrix}.$$

10 Comparaison avec la méthode des moindres carrés

Soit

$$R := \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{M}_{2,n}$$

On suppose que les x_i sont tous différents ce qui donne une fonction $y_i = g(x_i)$. La **méthode des moindres carrés** permet de trouver la fonction affine $f(x) = ax + b$ qui approxime le mieux possible la fonction g.

Pour mesurer l'écart entre les y_i et $f(x_i)$, on utilise la fonction:

$$L(a,b) := \sum_{i=1}^{n} (y_i - ax_i - b)^2 \tag{20}$$

On a $L(a,b) = 0$ si et seulement si $y_i = ax_i + b$ pour tout i. Remarquons que $L(a,b)$ est la distance L^2 entre les fonctions f et g.

Les paramètres a,b minimisant $L(a,b)$ satisfont $\operatorname{grad} L(a,b) = 0$ ce qui équivaut aux deux équations:

$$\frac{\partial L}{\partial a} = 0 \quad \text{et} \quad \frac{\partial L}{\partial b} = 0. \tag{21}$$

Ces équations s'écrivent

$$\left\{ \begin{array}{lclcl} a \sum_{i=1}^{n} x_i^2 & + & b \sum_{i=1}^{n} x_i & = & \sum_{i=1}^{n} x_i y_i \\ a \sum_{i=1}^{n} x_i & + & bn & = & \sum_{i=1}^{n} y_i \end{array} \right. \tag{22}$$

On regarde ceci comme un système de deux équations en les inconnues a et b. Le déterminant du système est

$$n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2.$$

L'inégalité de Schwarz pour les vecteurs x et $(1,\ldots,1)$ implique que ce déterminant est non-nul. Le système est donc de Cramer et admet l'unique solution

$$\begin{aligned} a &= \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\ b &= \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \end{aligned} \tag{23}$$

Observons que $L(a,b) \to \infty$ quand $\|(a,b)\| \to \infty$. Comme il n'y a qu'un point tel que $\operatorname{grad} L = 0$, donc au plus un extremum, c'est forcément un minimum.

On peut toujours translater l'origine de manière que $\sum_{i=1}^n x_i = 0$ et $\sum_{i=1}^n y_i = 0$. On est alors dans la situation favorable où, en posant $x = (x_1,\ldots,x_n)$ et $y = (y_1,\ldots,y_n)$, on a

$$b = 0 \quad \text{et} \quad a = \frac{\langle x,y\rangle}{\|x\|^2}.$$

Exemples: Prenons pour R la matrice

$$R := \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & -1 & 5 & 2 \end{pmatrix}.$$

Alors

1) la droite vectorielle $D_{\min}^{\text{vec}}$ solution du problème de compression vectoriel, est engendrée par le vecteur propre

$$(1,\frac{1}{44} + \frac{1}{44}\sqrt{1937}) \approx (1,1.023)$$

pour la plus grande valeur propre de RR^T.

2) la droite affine $D_{\min}$ solution du problème de compression affine, passe par $b(R) = (5/2,7/4)$ et a pour direction le vecteur propre

$$(1,\frac{55}{36} + \frac{55}{36}\sqrt{4321}) \approx (1,3.35)$$

pour la plus grande valeur propre de RQR^T où $Q := I_4 - \frac{1}{2}\mathbf{1}\mathbf{1}^T$. On voit que $D_{\min}$ n'est pas parallèle à $D_{\min}^{\text{vec}}$.

3) la méthode des moindre carrés donne une pente de 9/10.

Bibliogrqaphie

[AC] A. Albouy & A. Chenciner. Le problème des n corps et les distances mutuelles. Inventionnes Math. 131 (1998), 151–184.

[CH] G.M. Crippen & T.F. Havel. Distance geometry and molecular conformation.

[Dy] K.I. Diamantaris & S.Y. Yung. Principal component, Neural Networks, theory and applications. John Wyley & sons, Inc. 1996.

[Fr] J. Franklin. Matrix theory. Prentice Hall Inc. 1968.

[Go] J.C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrica 3, (1966), 325–338.

[Go2] J.C. Gower. Euclidean Distance Geometry. Math. Scientist 7 (1982) 1–14.

[Jo] I.T. Jolliffe. Principal component analysis. Springer-Verlag 1986.

[Sch] I.J. Schoenberg. Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés...". Annals of Math. 36 (1935) 724–732.