

# LES TRANSFORMATIONS DE SCHOENBERG: PROPRIÉTÉS ET APPLICATIONS EN ANALYSE DES DONNÉES

François Bavaud

*Facultés des Lettres et des Géosciences, Université de Lausanne  
Bâtiment Anthropole, CH-1015 Lausanne*

**RESUME :** la classe de toutes les fonctions transformant, composante par composante, une distance euclidienne en une autre distance euclidienne a été établie par Schoenberg en 1938. Ce résultat, assez méconnu en Analyse des Données, élargit le champ d'application du *multidimensional scaling*, dans sa version ordinaire ou pondérée. Ainsi que l'illustrent quelques exemples, son potentiel semble prometteur pour les applications faisant intervenir des dissimilarités euclidiennes.

**ABSTRACT :** the class of all componentwise transformations of a Euclidean distance into another Euclidean distance has been determined by Schoenberg in 1938. This result, not so well-known in Data Analysis, broadens the scope and application of multidimensional scaling, in its ordinary and weighted versions. As illustrated by a few examples, its potential seems promising for applications involving Euclidean dissimilarities.

**MOTS-CLEF :** analyse discriminante, courbure de Menger, distances euclidiennes, données pondérées, espaces de Hilbert, fonctions complètement monotones, hélices, méthodes à noyaux, *multidimensional scaling*, théorème de Huygens

## Introduction

Soit  $D$  une matrice symétrique d'ordre  $n$  et de diagonale nulle, représentant des dissimilarités  $n$  entre individus. Il est bien connu (Young et Householder 1938; Schoenberg 1938b), que  $D$  est *euclidienne carrée* (i.e. représentable sous la forme de somme de carrés de différences) ssi la matrice des *produits scalaires*  $B_{ij}^a := -\frac{1}{2}(D_{ij} - D_{ia} - D_{ja})$  (où  $a$  est l'un des  $n$  objets, pris comme référence) est semi-définie positive (s.d.p.), avec  $p$  valeurs propres non nulles. De façon équivalente, dans le contexte pondéré où chaque objet  $i$  est muni du poids  $\pi_i > 0$  avec  $\sum_i \pi_i = 1$ ,  $D$  est  *$p$ -euclidienne* ssi la matrice des produits scalaires pondérés  $K := -\frac{1}{2}\sqrt{\Pi}H D H' \sqrt{\Pi}$  est s.d.p. avec  $p$  valeurs propres non nulles  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p > 0$ , où  $\Pi$  est la matrice diagonale contenant les poids, et  $H$  la matrice de projection de composantes  $h_{ij} = \delta_{ij} - \pi_j$ . Dans ce cas, la décomposition spectrale  $K = U\Lambda U'$  permet de reconstituer les coordonnées des objets comme  $x_{i\alpha} := \frac{\sqrt{\lambda_\alpha}}{\sqrt{\pi_i}} u_{i\alpha}$ , avec les distances euclidiennes carrées  $D_{ij} = \sum_{\alpha=1}^p (x_{i\alpha} - x_{j\alpha})^2$  et l'inertie  $\Delta := \frac{1}{2} \sum_{ij} \pi_i \pi_j D_{ij} = \sum_{\alpha=1}^p \lambda_\alpha$  (voir par exemple Cuadras et Fortiana (1996) ou Bavaud (2006)): ce résultat fonde le *multidimensional scaling pondéré* (MDS), permettant de reconstituer, à une isométrie près, les coordonnées des  $n$  objets à partir de

leurs distances euclidiennes, chaque sous-dimension  $\alpha$  exprimant une proportion  $\lambda_\alpha/\Delta$  de l'inertie totale.

Un autre résultat de Schoenberg, de la même année (1938a) est moins connu en Analyse de Données et en méthodes à noyaux (kernels), malgré son potentiel prometteur: il concerne l'identification de la classe la plus générale des fonctions  $\phi(D_{ij}) = \tilde{D}_{ij}$  transformant, composante par composante, une distance Euclidienne carrée  $D_{ij}$  en une autre distance Euclidienne carrée  $\tilde{D}_{ij}$ . L'ensemble de ces fonctions est donné par la classe

$$\varphi(D) = \int_0^\infty \frac{1 - \exp(-\lambda D)}{\lambda} g(\lambda) d\lambda \quad (1)$$

où  $g(\lambda) d\lambda$  est une mesure non-négative, pouvant contenir des parties singulières (distributions), telle que  $\int_\epsilon^\infty \frac{g(\lambda)}{\lambda} d\lambda$  est finie  $\forall \epsilon > 0$ .

Nous appellerons *transformation de Schoenberg* toute fonction  $\varphi(D)$  de la forme (1), dont les propriétés et applications constituent l'objet de cette communication. Une telle fonction transforme une configuration de  $n$  points (dont on peut calculer les distances euclidiennes carrées  $D$ ) en une nouvelle configuration (obtenue en appliquant le MDS aux nouvelles distances euclidiennes carrées  $\tilde{D} = \varphi(D)$ ), ouvrant ainsi de nouvelles perspectives en visualisation et en *clustering*.

## Quelques propriétés

Par construction,  $\phi(D) \geq 0$  avec  $\phi(0) = 0$ . Aussi,  $\phi'(D) = \int_0^\infty \exp(-\lambda D) g(\lambda) d\lambda$ , dont l'expression coïncide avec la classe des fonctions *complètement monotones* (Bernstein 1929); ainsi, les dérivées impaires  $\phi'(D)$ ,  $\phi'''(D)$ , etc. des transformations sont positives, leurs dérivées paires  $\phi''(D)$ ,  $\phi''''(D)$ , etc. étant négatives (Table 1). En particulier, la transformation identité  $\phi(D) = D$  s'obtient de  $g(\lambda) = \delta(\lambda)$ ; aussi,  $\sqrt{D}$  est euclidien si  $D$  l'est.

fonction $g(\lambda)$		transformation $\varphi(D)$	bornée	rectifiable
$g_1(\lambda) = \delta(\lambda - a)$	$a \geq 0$	$\varphi_1(D) = \frac{1 - \exp(-aD)}{a}$	✓	✓
$g_2(\lambda) = \theta(\lambda \leq \frac{\pi}{2}) \lambda \sin \lambda$		$\varphi_2(D) = \frac{D(D + \exp(-\frac{\pi}{2}D))}{1 + D^2}$	✓	✓
$g_3(\lambda) = \exp(-a\lambda)$	$a > 0$	$\varphi_3(D) = \ln(1 + \frac{D}{a})$	—	✓
$g_4(\lambda) = \lambda \exp(-a\lambda)$	$a > 0$	$\varphi_4(D) = \frac{D}{a(a+D)}$	✓	✓
$g_5(\lambda) = \frac{a}{\Gamma(1-a)} \lambda^{-a}$	$0 < a < 1$	$\varphi_5(D) = D^a$	—	—
		$\varphi_6(D) = \frac{D^a}{1+D^a} \quad 0 < a < 1$	✓	—

Table 1: quelques transformations de Schoenberg

Une transformation est *bornée* si  $\phi(\infty)$  est fini (i.e.  $\int_0^\infty \frac{g(\lambda)}{\lambda} d\lambda$  est fini). Une transformation est *rectifiable* si  $\phi'(0)$  est fini, (i.e.  $\int_0^\infty g(\lambda) d\lambda$  est fini). On peut montrer (Bavaud 2010):

- 1) que la transformation rectifiable d'un segment de droite de longueur  $L$  est une courbe se déployant dans un espace de dimension généralement infinie, de longueur  $\tilde{L} = \sqrt{\phi'(0)} L$
- 2) qu'un *angle droit se transforme en angle aigu*
- 3) que le carré de la courbure de Menger de la transformation rectifiable d'une ligne vaut

$$\kappa^2 = \lim_{\epsilon \rightarrow 0} \frac{4\varphi(\epsilon^2) - \varphi(4\epsilon^2)}{\varphi^2(\epsilon^2)} = -\frac{6\varphi''(0)}{(\varphi'(0))^2} \geq 0 .$$

## Illustrations et applications

### Grille

Les distances entre  $n = 100$  points disposés sur un grille bidimensionnelle (fig. 1, en haut à gauche) sont transformées selon  $D \rightarrow D^{0.4}$ . La nouvelle métrique étant euclidienne, *on peut donc y appliquer le MDS* (avec pondération uniforme); le nuage résultant est projeté selon les dimensions 1 et 2 (en haut à droite) et 3 et 4 (en bas à gauche). Le *scree plot* (en bas à gauche) montre qu'un nombre maximum de dimensions ( $p = 99$ ) sont impliquées dans la transformation - une conséquence de la nature continue de la mesure  $g(\lambda)$  correspondante (Schoenberg 1938a); la même transformation  $D \rightarrow D^{0.4}$  d'un rectangle (continu) entraînerait le déploiement de ce dernier dans un nombre infini de dimensions.

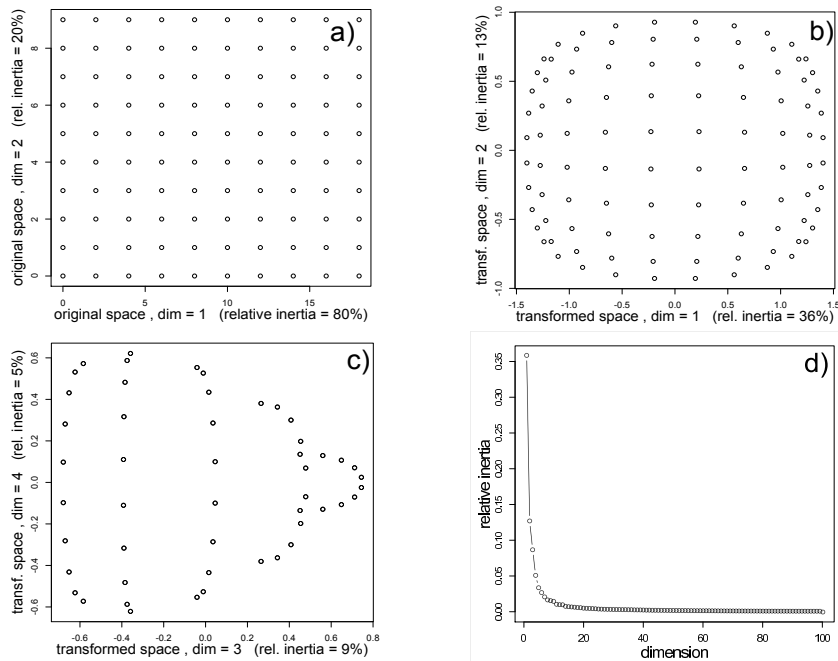


Figure 1: transformation  $D \rightarrow \tilde{D} := D^{0.4}$  d'une grille bidimensionnelle

## Barre

La figure 2 visualise la transformation de Schoenberg d'une barre quasi unidimensionnelle de  $n = 1'000$  points, générée aléatoirement par  $X_1 \sim U(0, 1000)$  et  $X_2 \sim U(0, 1)$  (a). L'image d'une droite par une transformation préservant la nature euclidienne des distances est appelée *screw line* par Schoenberg, et *hélice* par Kolmogorov (1940) (voir aussi Fuglede 2005) - une terminologie appropriée au vu de la figure.

Les deux premières dimensions du MDS expriment 61.0%, respectivement 15.1% de l'inertie relative. On peut montrer que les quantités exactes pour la transformation d'une ligne sont de  $\frac{6}{\pi^2} = 60.8\%$  et  $\frac{15}{2\pi^2} = 15.2\%$  respectivement.

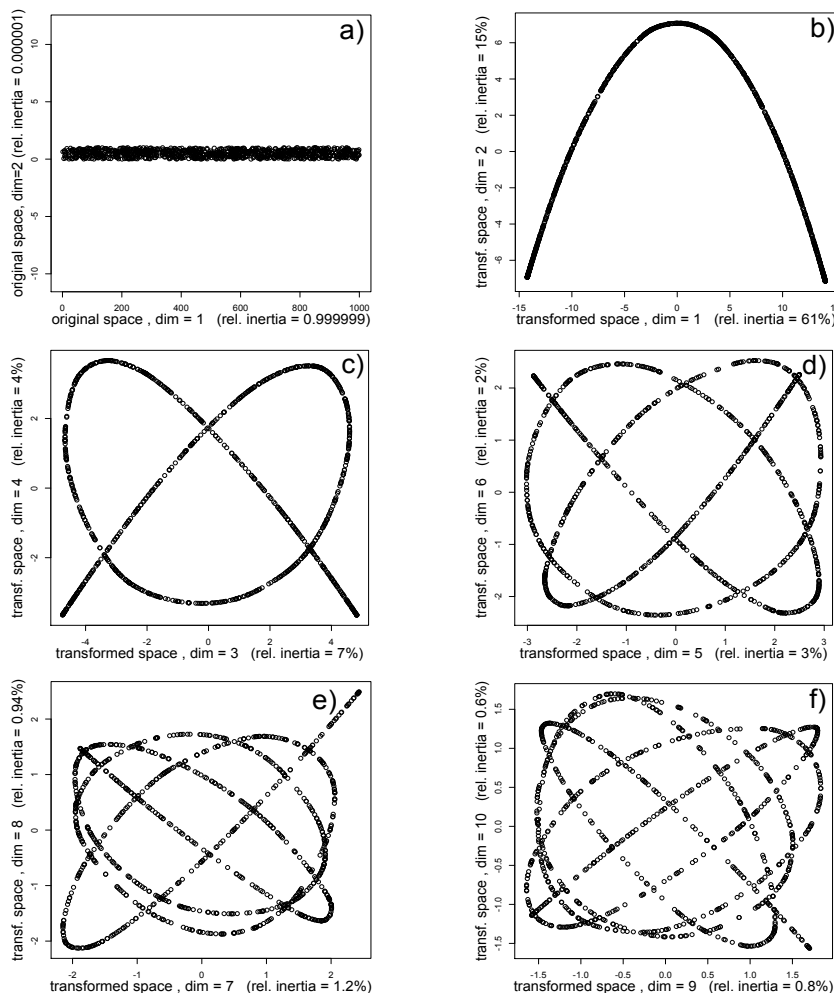


Figure 2: Premières projections de basse dimensionalité (b, c, d, e et f) de la transformation racine carrée  $\tilde{D} = \sqrt{D}$  à partir d'une barre (a)

## Analyse discriminante sur les distances

Soient  $n$  objets de caractéristiques  $p$ -dimensionnelles, définissant des distances euclidiennes carrées  $D_{ij}$  après standardisation et orthogonalisation des caractéristiques (distances de Mahalanobis). En supposant que ces objets appartiennent à des groupes  $g = 1, \dots, m$ , une stratégie élémentaire de discrimination serait d'attribuer chaque objet au groupe  $g$  dont le centroïde est le plus proche, à savoir à  $\arg \min_g D_{ig}$ , avec (théorème de Huygens)

$$\sum_j f_j^g D_{ij} = D_{ig} + \Delta_g \quad \Delta_g = \frac{1}{2} \sum_{jk} f_j^g f_k^g D_{jk} . \quad (2)$$

Ce qui fonctionne bien avec l'analyse discriminante linéaire de Fisher (1936) appliquée aux Iris ( $n = 150$ ,  $p = 4$ ,  $m = 3$ , pourcentage de bien classés de 97%), ne marche évidemment plus avec les données de la figure 3 ( $n = 150$ ,  $p = 2$ ,  $m = 3$ , pourcentage de bien classés de 35%, proche de la valeur de 33% sous attribution aléatoire). En transformant au préalable les distances euclidiennes, on obtient l'algorithme élémentaire:

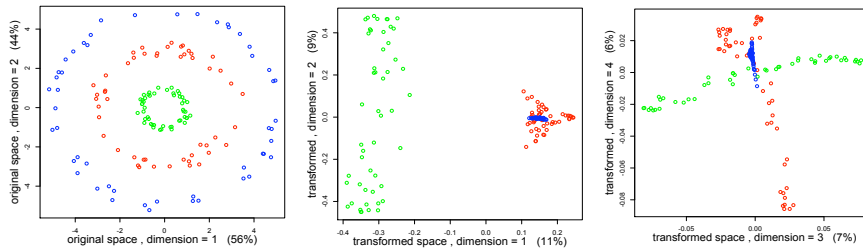


Figure 3: gauche: trois groupes de 50 individus, générés uniformément sur trois cercles de rayons 1, 3 et 5, avec un écart type respectif de 0.1, 0.3 et 0.2. Centre et droite: premières dimensions (MDS) de la configuration transformée par  $\varphi(D) = 1 - \exp(-0.65 D)$ .

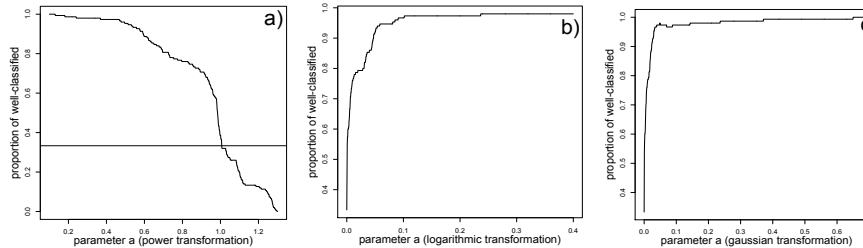


Figure 4: proportion d'individus bien classés, après transformation de Schoenberg des données de la figure 3, selon a) la transformation de puissance  $\varphi(D) = D^a$ ; à noter que les valeurs  $a > 1$  ne correspondent pas à une transformation valide au sens de Schoenberg. b) transformation logarithmique  $\varphi(D) = \ln(1 + aD)$ . c) transformation Gaussienne  $\varphi(D) = 1 - \exp(-aD)$ .

### Analyse discriminante sur les distances:

- 1) calculer  $\tilde{D}_{i\tilde{g}} = \sum_{j=1}^n f_j^g \tilde{D}_{ij} - \frac{1}{2} \sum_{j,k=1}^n f_j^g f_k^g \tilde{D}_{jk}$  où  $\tilde{D}_{ij} = \varphi(D_{ij})$   
et  $f_j^g = I(i \in g)/n_g$  ( $n_g = \sum_{j \in g} 1$ ) est la distribution dans le groupe  $g$
- 2) attribuer l'objet  $i$  au groupe  $\arg \min_{\tilde{g}} \tilde{D}_{i\tilde{g}}$ .

Les résultats de la figure 4 montrent que la proportion d'individus bien classés va jusqu'à 100% pour la transformation Gaussienne (pour  $a \geq 0.65$ ): une transformation de Schoenberg suffisamment vigoureuse parvient ainsi à transformer la configuration initiale de sorte que les trois groupes appartiennent à trois hyper-sphères *disjointes*.

D'une part, ce résultat est attendu: de nombreux travaux en *Machine Learning* ont démontré l'efficacité des plongements de haute dimensionalité des données, permettant de mieux les séparer (voir par exemple Chen et al. 2007). D'un autre côté, la simplicité conceptuelle, formelle et computationnelle de l'algorithme mérite d'être soulignée.

En guise de conclusion, un *travail de réécriture* des méthodes de *Machine Learning* (impliquant des noyaux isotropes définis positifs) en termes de *distances euclidiennes carrées* (plutôt qu'en termes des produits scalaires que sont les noyaux, justement) devrait favoriser une réappropriation stimulante et pédagogique de ces algorithmes dans les nombreux domaines de l'Analyse des Données impliquant ce type de dissimilarités.

## Bibliographie

- [1] Bavaud F. (2006) Spectral Clustering and Multidimensional Scaling : A Unified View. In Batagelj V. et al. (éditeurs) *Data Science and Classification*, 131-139, Springer.
- [2] Bavaud F. (2010) On the Schoenberg Transformations in Data Analysis: Theory and Illustrations. *Soumis pour publication*
- [3] Bernstein, S. (1929) Sur les fonctions absolument monotones. *Acta Math.*, **52**, 1-66.
- [4] Chen, D., He, Q., Wang, X. (2007) On linear separability of data sets in feature space, *Neurocomputing* **70**, 2441-2448.
- [5] Cuadras, C.M., Fortiana, J. (1996) Weighted continuous metric scaling. In: Gupta, A.K. et Girko, V.L. (éditeurs) *Multidimensional Statistical Analysis and Theory of Random Matrices*, 27-40, VSP, The Netherlands.
- [6] Fuglede, B. (2005) Spirals in Hilbert space: With an application in Information theory. *Expo. Math.* **23**, 23 - 45.
- [7] Kolmogorov, A.N. (1940) The Wiener helix and other interesting curves in the Hilbert space. *Dokl. Acad. Sci. USSR*, **26**, 115-118.
- [8] Schoenberg, I. J. (1938a) Metric Spaces and Completely Monotone Functions. *The Annals of Mathematics* **39**, 811-841.
- [9] Schoenberg, I. J. (1938b) Metric Spaces and Positive Definite Functions. *Transactions of the American Mathematical Society*, **44**, 522-536.
- [10] Young, G., Householder, A. S. (1938) Discussion of a set of points in terms of their mutual distances. *Psychometrika*, **3**, 19-22.