# Spectral clustering and multidimensional scaling: a unified view

François Bavaud

Section d'Informatique et de Méthodes Mathématiques
Faculté des Lettres, Université de Lausanne
CH-1015 Lausanne, Switzerland

(To appear in the proccedings of the *IFCS 2006 Conference*: "Data Science and Classification", Ljubljana, Slovenia, July 25 - 29, 2006)

**Abstract.** Spectral clustering is a procedure aimed at partitionning a weighted graph into minimally interacting components. The resulting eigen-structure is determined by a reversible Markov chain, or equivalently by a symmetric transition matrix $F$. On the other hand, multidimensional scaling procedures (and factorial correspondence analysis in particular) consist in the spectral decomposition of a kernel matrix $K$. This paper shows how $F$ and $K$ can be related to each other through a linear or even non-linear transformation leaving the eigen-vectors invariant. As illustrated by examples, this circumstance permits to define a transition matrix from a similarity matrix between $n$ objects, to define Euclidean distances between the vertices of a weighted graph, and to elucidate the "flow-induced" nature of spatial auto-covariances.

## 1 Introduction and main results

Scalar products between features define similarities between objects, and reversible Markov chains define weighted graphs describing a stationary flow. It is natural to expect flows and similarities to be related: somehow, the exchange of flows between objects should enhance their similarity, and transitions should preferentially occur between similar states.

This paper formalizes the above intuition by demonstrating in a general framework that the symmetric matrices $K$ and $F$ possess an identical eigenstructure, where $K$ (kernel, equation (2)) is a measure of similarity, and $F$ (symmetrized transition. equation (5)) is a measure of flows. Diagonalizing $K$ yields principal components analysis (PCA) as well as mutidimensional scaling (MDS), while diagonalizing $F$ yields spectral clustering. By theorems 1, 2 and 3 below, eigenvectors of $K$ and $F$ coincide and their eigenvalues are simply related in a linear or non-linear way.

Eigenstructure-based methods constitute the very foundation of classical multivariate analysis (PCA, MDS, and correspondence analysis). In the last decade, those methods have been very extensively studied in the machine learning community (see e.g. Shawe-Taylor and Cristianini 2004, and references therein), in relationship to manifold learning and spectral clustering

(Bengio et al. 2004). The general "$K - F$ connection" described here hence formalizes a theme whose various instances have already been encountered and addressed in the classical setup (see section 2.2) or in the kernel setup, at least implicitly. The relative generality of the present approach (weighted objects, weighted variables, weighted graphs) might provide some guidance for defining the appropriate objects (kernels, scalar products, similarities or affinities, etc.). Also, the same formalism permits to characterize a broad family of *separable auto-covariances*, relevant in spatial statistics.

**Multi-dimensional scaling (MDS) in a nutshell:** consider $n$ objects described by $p$ features. Data consist of $\Phi = (\varphi_{ij})$ where $\varphi_{ij}$ is the value of the $j$-th feature on the $i$-th object. Let $\rho_j > 0$ denote the weight of feature $j$, with $\sum_{j=1}^{p} \rho_j = 1$, and define the diagonal matrix $R := \mathrm{diag}(\rho)$. Also, let $\pi_i > 0$ denote the weight of object $i$, with $\sum_{i=1}^{n} \pi_i = 1$, and define $\Pi := \mathrm{diag}(\pi)$. Also, define

$$B_{ii'} := \sum_j \rho_j \varphi_{ij} \varphi_{i'j} \qquad D_{ii'} := B_{ii} + B_{i'i'} - 2B_{ii'} = \sum_j \rho_j (\varphi_{ij} - \varphi_{i'j})^2 \quad (1)$$

The scalar product $B_{ii'}$ constitutes a measure a *similarity* between objects $i$ and $i'$, while the squared Euclidean distance $D_{ii'}$ is a measure of their *dissimilarity*. Classical MDS consists in obtaining distance-reproducing coordinates such that the (total, weighted) *dispersion* $\Delta := \frac{1}{2} \sum_{ii'} \pi_i \pi_{i'} D_{ii'}$ is optimally represented in a low-dimensional space. To that effect, the coordinate $x_{i\alpha}$ of object $i$ on factor $\alpha$ is obtained from the spectral decomposition of the *kernel* $K = (K_{ii'})$ with $K_{ii'} := \sqrt{\pi_i \pi_{i'}} B_{ii'}$ as follows:

$$K := \sqrt{\Pi} B \sqrt{\Pi} = U \Gamma U' \qquad U = (u_{i\alpha}) \qquad \Gamma = \mathrm{diag}(\gamma) \qquad x_{i\alpha} := \frac{\sqrt{\gamma_\alpha}}{\sqrt{\pi_i}} u_{i\alpha} \quad (2)$$

where $U$ is orthogonal and contains the eigenvectors of $K$, and $\Gamma$ is diagonal and contains the eigenvalues $\{\gamma_\alpha\}$ of $K$. Features are *centred* if $\sum_i \pi_i \varphi_{ij} = 0$. In that case, the symmetric, positive semi-definite (p.s.d) matrices $B$ and $K$ obey $B\pi = 0$ and $K\sqrt{\pi} = 0$, and will be referred to as a *proper similarity matrix*, respectively *proper kernel matrix*. By construction

$$D_{ii'} = \sum_{\alpha \geq 2} (x_{i\alpha} - x_{i'\alpha})^2 \qquad \Delta = \sum_{\alpha \geq 2} \gamma_\alpha \quad (3)$$

where $\gamma_1 = 0$ is the trivial eigenvalue associated with $u_1 = \sqrt{\pi}$.

**Spectral clustering in a nutshell:** consider the $(n \times n)$ normalised, symmetric *exchange* matrix $E = (e_{ii'})$ where $e_{ii'} = e_{i'i} \geq 0$, $e_{i\bullet} := \sum_{i'} e_{ii'} > 0$, and $\sum_{ii'} e_{ii'} = 1$. By construction, $w_{ii'} := e_{ii'}/e_{i\bullet}$ is the transition matrix of a reversible Markov chain with stationary distribution $\pi_i := e_{i\bullet}$. In a weighted

graph framework, $e_{ii'}$ constitutes the weight of the undirected edge $(ii')$, measuring the proportion of units (people, goods, matter, news...) circulating in $(ii')$, and $\pi_i$ is the the weight of the object (vertex) $i$.

The *minimal normalized cut* problem consists in partitioning the vertices into two disjoints sets $A$ and $A^c$ as little interacting as possible, in the sense that

$$h := \min_A \frac{e(A, A^c)}{\min(\pi(A), \pi(A^c))} \quad (\text{with } e(A, A^c) := \sum_{i \in A, i' \in A^c} e_{ii'} \ , \ \pi(A) := \sum_{i \in A} \pi_i) \tag{4}$$

where the minimum value $h$ is the *Cheeger's constant* of the weighted graph.

The eigenvalues of $W = (w_{ii'})$ are real and satisfy $1 = \lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq -1$, with $\lambda_2 < 1$ iff the chain is irreducible and $\lambda_n > -1$ iff the chain is not of period two (bipartite graphs). The same eigenvalues appear in the spectral decomposition of the *symmetrized transition matrix* $F = (f_{ii'})$ defined as $f_{ii'} = e_{ii'}/\sqrt{\pi_i \pi_{i'}}$:

$$F := \Pi^{-\frac{1}{2}} E \Pi^{-\frac{1}{2}} = U \Lambda U' \qquad U = (u_{i\alpha}) \qquad \Lambda = \text{diag}(\lambda) \tag{5}$$

where $U$ is orthogonal and $\Lambda$ diagonal. By construction, $F\sqrt{\pi} = \sqrt{\pi}$. A symmetric, non-negative matrix $F$ with eigenvalues in $[-1, 1]$ and $F\sqrt{\pi} = \sqrt{\pi}$ will be refereed to as a *proper* symmetrized transition matrix.

In its simplest version, *spectral clustering* (see e.g. Ng et al. (2002); Verma and Meila (2003)) consists in partitioning the graph into two disjoints subsets $A(u) := \{i | u_{i2} \leq u\}$ and $A^c(u) := \{i | u_{i2} > u\}$, where $u_{i2}$ is the second eigenvector and $u$ a threshold, chosen as $u = 0$, or as the value $u$ making $\sum_{i \in A(u)} u_{i2}^2 \cong \sum_{i \in A^c(u)} u_{i2}^2$, or the value minimising $h(u) := e(A(u), A^c(u))/\min(\pi(A(u)), \pi(A^c(u)))$. Minimal normalized cut and spectral clustering are related by the Cheeger inequalities (see e.g. Diaconis and Strook (1991); Chung (1997))

$$2h \geq 1 - \lambda_2 \geq 1 - \sqrt{1 - h^2} \tag{6}$$

where the *spectral gap* $1 - \lambda_2$ controls the speed of the convergence of the Markov dynamics towards equilibrium.

**Theorem 1.** *($\mathbf{F} \to \mathbf{K}$). Let $E$ be an $(n \times n)$ exchange matrix with associated symmetrized transition matrix $F = U \Lambda U'$ and vertex weight $\pi$. Then any $(n \times n)$ matrix $K = (K_{ii'})$ of the form*

$$K := (a - b)F + (a + b)I - 2a\sqrt{\pi}\sqrt{\pi}' \tag{7}$$

*constitutes, for $a, b \geq 0$, a centred proper kernel with spectral decomposition $F = U \Gamma U'$ with eigenvalues $\gamma_\alpha = (a - b)\lambda_\alpha + (a + b) - 2a\, \delta_{\alpha 1}$.*

**Proof :** the eigenvectors $u_\alpha$ of $I$ and $\sqrt{\pi}\sqrt{\pi}'$ are identical to those of $F$, with associated eigenvalues $\mu_\alpha \equiv 1$ and $\mu_\alpha = \delta_{\alpha 1}$ respectively. In particular,

$K\sqrt{\pi} = [(a-b)+(a+b)-2a]\sqrt{\pi} = 0$, making $K$ centred. It remains to show the positive-definiteness of $K$, that is $\gamma_\alpha \geq 0$. Actually, $\gamma_1 = 0$ and, for $\alpha \geq 2$, $\gamma_\alpha = a(1+\lambda_\alpha) + b(1-\lambda_\alpha) \geq 0$ since $-1 < \lambda_\alpha < 1$. □

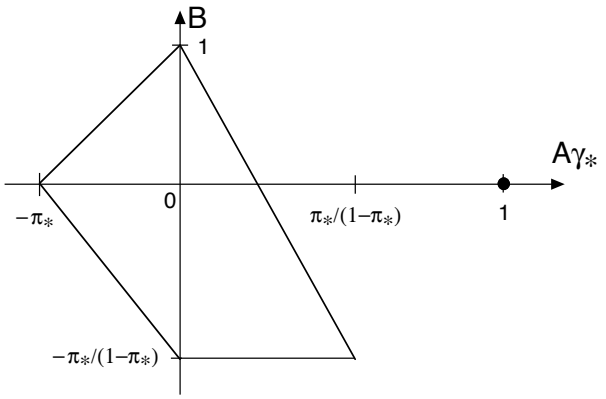


**Fig. 1.** domain of possible values $A\gamma_*$ and $B$ insuring the existence of a proper symmetrized transition $F$ from a kernel $K$ by (8). Although allowing for non-trivial values $A, B \neq 0$, the domain is not optimal, and degenerates into $A = 0$ and $B \in [0,1]$ for $n \to \infty$ in view of $\pi_* \to 0$. The point $(1,0)$ depicts the values corresponding to the FCA example of section (2.2).

**Theorem 2.** *(**K** → **F**). Let $K$ be an $(n \times n)$ centred kernel with trivial eigenvector $\sqrt{\pi}$. Then any $(n \times n)$ matrix $F = (f_{ii'})$ of the form*

$$F = AK + BI + (1-B)\sqrt{\pi}\sqrt{\pi}' \tag{8}$$

*constitutes, for $A \in [-\frac{\pi_*}{\gamma_*}, \frac{\pi_*}{(1-\pi_*)\gamma_*}]$ and $B \in [-\frac{\pi_* + \min(A,0)\gamma_*}{1-\pi_*}, \frac{\pi_* - |A|\gamma_*}{\pi_*}]$ (where $\gamma_* := \max_\alpha \gamma_\alpha$ and $\pi_* := \min_i \pi_i$), a non-negative symmetrized transition matrix with associated stationary distribution $\pi$* (see figure 1).

**Proof :** treating separately the cases $A \geq 0$ and $A \leq 0$, and using (in view of the positive-definite nature of $K$) $\max_i K_{ii} \leq \gamma_*$, $\min_i K_{ii} \geq 0$, $\max_{i \neq i'} K_{ii'} \leq \gamma_*$ and $\min_{i \neq i'} K_{ii'} \geq -\gamma_*$ as well as $\min_{i,i'} \sqrt{\pi_i \pi_{i'}} = \pi_*$ demonstrates that $F$ as defined in (8) obeys $\min_{i \neq i'} f_{ii'} \geq 0$ and $\min_i f_{ii} \geq 0$. Thus $e_{ii'} := \sqrt{\pi_i \pi_{i'}} f_{ii'}$ is symmetric, non-negative, and satisfies in addition $e_{i\bullet} = \pi_i$ in view of $K\sqrt{\pi} = 0$. □

The coefficients $(A, B)$ of theorem 2 are related to the coefficients $(a, b)$ of theorem 1 by $A = 1/(a-b)$ and $B = (b+a)/(b-a)$, respectively $a = (1-B)/2A$ and $b = -(1+B)/2A$. The maximum eigenvalue $\gamma_* := \max_\alpha \gamma_\alpha > 0$ of $K$ is $\gamma_* = a(1+\lambda_2) + b(1-\lambda_2) = (\lambda_2 - B)/A$ for $a > b$ (i.e. $A > 0$), and $\gamma_* = a(1+\lambda_n) + b(1-\lambda_n) = (\lambda_n - B)/A$ for $a < b$ (i.e. $A < 0$).

# 2 Examples

## 2.1 Spectral clustering: Swiss commuters

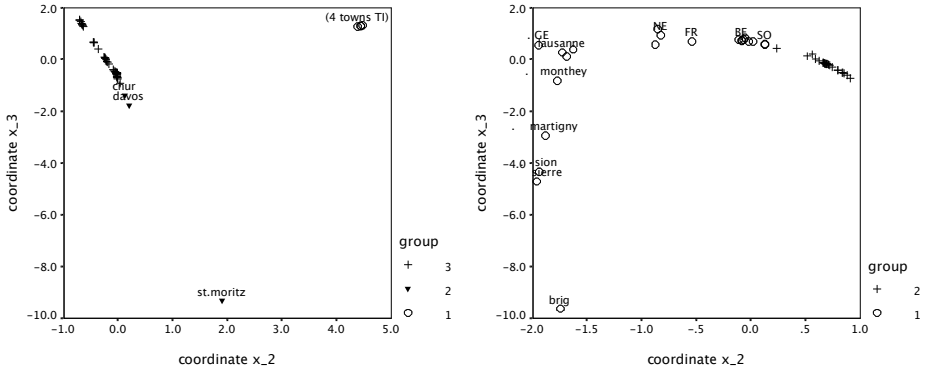


**Fig. 2.** Two-dimensional factorial towns configuration $x_{i\alpha}$ for $\alpha = 2, 3$ for the initial network ($n = 55$, left) and, for the largest sub-network obtained after four minimal normalized cuts ($n = 48$, right).

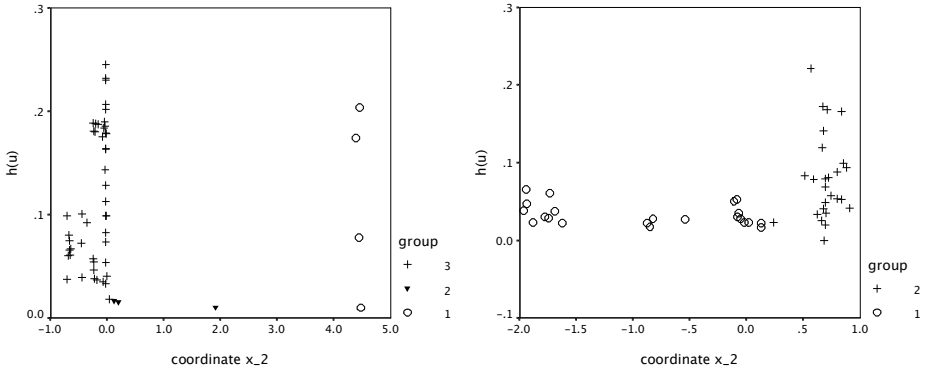


**Fig. 3.** Determining the minimal normalized cuts $\min_u h(u)$ along the "second-eigenvalue path" with discrete values $u_i = \sqrt{\pi_i}\, x_{i2}$. Left: 55 towns, from which Ticino (4 towns, 1st iteration) and Graubünden (3 towns, 2nd and 3rd iteration) are removed. Right: the resulting 48 towns, split into (VS-VD-GE) and (NE-FR-JU-BE-SO) for the first group, and the rest of the German-speaking towns for the second group.

The number of daily commuters $n_{ij}$ from place $i$ to place $j$ (between $n = 55$ extended Swiss towns) yields (after symmetrization) a weighted graph with associated transition matrix $F$.

Eigenvalues are $\lambda_1 = 1 > \lambda_2 = .9947 > \ldots > \lambda_{55} = .5116$. Factor coordinates $x_{i\alpha}$ (figure 2) define "flow-revealed" distances $D_{ii'}$. In view of theorem 1 (and in view of the arbitrariness of $\gamma(\lambda)$, and of the closeness between the eigenvalues $\lambda_\alpha$) the coordinates are simply defined (see equation (2)) as $x_{i\alpha} = u_{i\alpha}/\sqrt{\pi_i} = u_{i\alpha}/u_{i1}$. They are obviously reminiscent of the geographical map, but the precise mechanism producing the factor maps of figure 2 remains to be elucidated. The spectral clustering determination of the threshold $u$ minimizing $h(u)$ (section 1) is illustrated in figure 3.

### 2.2 Correspondence analysis: educational levels in the region of Lausanne

Let $N = (n_{ij})$ be a $(n \times m)$ contingency table counting the number of individuals belonging to category $i$ of $X$ and $j$ of $Y$. The "natural" kernel matrix $K = (K_{ii'})$ and transition matrix $W = (w_{ii'})$ associated with factorial correspondence analysis (FCA) are (Bavaud and Xanthos 2005)

$$K_{ii'} = \sqrt{\pi_i}\sqrt{\pi_{i'}} \sum_j \rho_j (q_{ij} - 1)(q_{i'j} - 1) \qquad w_{ii'} := \pi_{i'} \sum_j \rho_j q_{ij} q_{i'j} \quad (9)$$

where $\pi_i = n_{i\bullet}/n_{\bullet\bullet}$ are the row profiles, $\rho_j = n_{\bullet j}/n_{\bullet\bullet}$ the columns profiles, and $q_{ij} = (n_{ij}\, n_{\bullet\bullet})/(n_{i\bullet} n_{\bullet j})$ are the *independence quotients*, that is the ratio of the counts by their expected value under independence.

Coordinates $x_{i\alpha}$ (2) obtained from the spectral decomposition of $K$ are the usual objects' coordinates in FCA (for $\alpha \geq 2$), with associated $\chi$-square dissimilarities $D_{ii'}$ and $\chi$-square inertia $\Delta = \texttt{chi}^2/n_{\bullet\bullet}$ (Bavaud 2004). On the other hand, $w_{ii'}$ is the conditional probability of drawing an object of category $i'$ starting with an object of category $i$ and "transiting" over all possible modalities $j$ of $Y$. The resulting Markov chain on $n$ states is reversible with stationary distribution $\pi$, exchange matrix $e_{ii'} = e_{i'i} = \pi_i w_{ii'}$ and symmetrized transition matrix $f_{ii'} = \sqrt{\pi_i}\sqrt{\pi_{i'}} \sum_j \rho_j q_{ij} q_{i'j}$.

Here $K$ and $F$ are related as $K = F - \sqrt{\pi}\sqrt{\pi}'$, with values $A = 1$ and $B = 0$ (respectively $a = -b = 1/2$) and $\gamma_* = 1$ in theorems 2 and 1. The corresponding value lie outside the non-optimal domain of figure 1.

Data[1] give the number of achieved educational levels $i$ (8 categories) among $169'836$ inhabitants living in commune $j$ (among $p = 12$ communes around Lausanne, Switzerland). Eigenvalues are $\gamma_1 = 0$ and $1 > \gamma_2 = \lambda_2 = .023 > \ldots \lambda_8 = .000026$ with inertia $\Delta = .031$. While significantly non-zero ($n_{\bullet\bullet}\Delta >> \chi^2_{.99}[77]$), those low values are close to the perfect mobility case (section 4), that is regional educational disparities are small in relative terms.

[1] F.Micheloud, private communication

Figure 4 depicts the factor configuration ($\alpha = 2.3$) with coordinates (2) as well as dual regional coordinates. The biplot confirms the existence of the well-attested West-East educational gradient of the region.

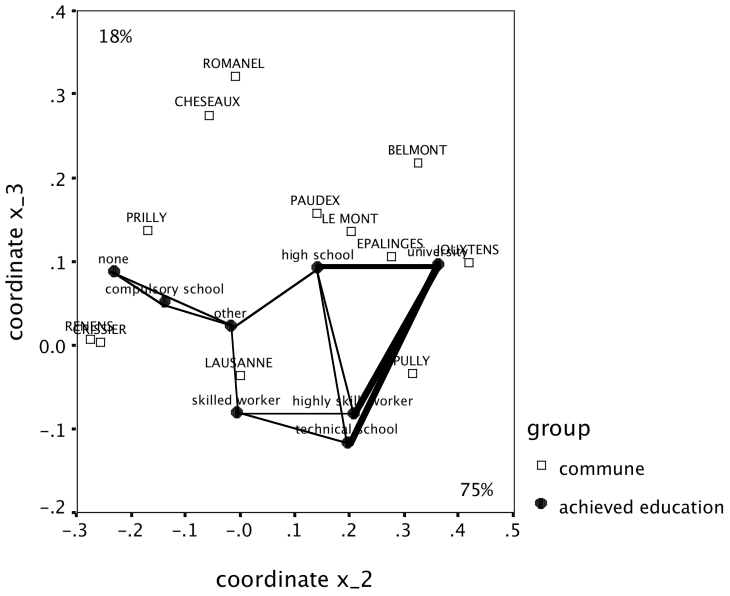


**Fig. 4.** Biplot: FCA rows and columns objects' coordinates. The symmetric quantity $s_{ii'} := w_{ii'}/\pi_{i'}$ is a size-independent measure of similarity with average 1 (Bavaud and Xanthos 2005), defining strong ($s \geq 1.05$), weak ($1.05 > s \geq 1$) or no ($s < 1$) links between distinct education levels.

## 3 Non-linear transformations

**Theorem 3.** *i) Let $K$ be a proper kernel. Then $K^r$ (for $r = 1, 2, \ldots$) and $h(K) := \sum_{r\geq 1} h_r K^r$ (where $h_r \geq 0$ and $\sum_{r\geq 1} h_r = 1$) are proper kernels.*

*ii) Let $F$ be a proper symmetrized transition. Then $F^r$ (for $r = 0, 1, 2, \ldots$), $f(F) := \sum_{r\geq 1} f_r F^r$ (where $f_r \geq 0$ and $\sum_{r\geq 1} f_r = 1$) and $cf(F) + (1-c)I$ (where $0 < c \leq 1$) are proper symmetrized transitions.*

*iii) $K$ and $F$ can be put in non-linear correspondence by*

$$h(K) = (a - \tilde{b})f(F) + (a + \tilde{b})I - 2a\sqrt{\pi}\sqrt{\pi}' \qquad a, \tilde{b} \geq 0 \qquad (10)$$

**Proof :** i) and ii) are immediate. Part iii) follows from theorem (1) and definition $\tilde{b} := (1-c)a + cb$. □

Since $h(U\Gamma U') = Uh(\Gamma)U'$ and $f(U\Lambda U') = Uf(\Lambda)U'$, theorem 3 exhibits a broad class of MDS - spectral clustering correspondences (see the examples of section 4), differing by their eigenvalues spectrum but sharing the same eigenvectors, in particular $u_1$ and hence the weights vector $\pi = u_1^2$.

## 4 Separable auto-covariances

The present formalism turns out to be relevant in spatial statistics, where spatial autocorrelation is defined by a covariance matrix between the objects (= regions).

To that extent, consider a spatial field $\{X_i\}_{i=1}^n$ measured on $n$ regions, with common expectation $E(X_i) = \mu$ and associated weights $\{\pi_i\}_{i=1}^n$. Let $\bar{X} := \sum_i \pi_i X_i$. The auto-covariance matrix $\Sigma = (\sigma_{ii'})$ is said to be *separable* if, for any $i$, the variables $X_i - \bar{X}$ and $\bar{X} - \mu$ are not correlated.

**Theorem 4.** *$\Sigma$ is separable iff $\Sigma\pi = \sigma^2\mathbf{1}$, where $\sigma^2 = E((\bar{X}-\mu)^2)$ and $\mathbf{1}$ is the unit vector. In this case, the $(n \times n)$ matrices*

$$K := \frac{1}{\sigma^2}\sqrt{\Pi}\Sigma\sqrt{\Pi} - \sqrt{\pi}\sqrt{\pi}' \qquad\qquad B = \frac{1}{\sigma^2}\Sigma - J \tag{11}$$

*(where $J := \mathbf{1}\mathbf{1}'$ is the unit matrix) constitute a proper kernel, respectively dissimilarity.*

**Proof :** $\Sigma\pi = \sigma^2\mathbf{1}$ iff $\sigma^2 = \sum_{i'} \pi_{i'}[E((X_i-\mu)(X_{i'}-\bar{X})) + E((X_i-\mu)(\bar{X}-\mu))] = E((X_i-\bar{X})(\bar{X}-\mu)) + E((\bar{X}-\mu)(\bar{X}-\mu))$ iff $E((X_i-\bar{X})(\bar{X}-\mu)) = 0$ and $E((\bar{X}-\mu)^2) = \sigma^2$. □

Under separability, equations (1) and (11) show the *variogram* of Geostatistics to constitute a squared Euclidean distance since $\mathrm{Var}(X_i - X_{i'}) = \sigma^2 D_{ii'}$. Observe that $\Sigma$ or $B$ as related by (11) yield (up to $\sigma^2$) the same distances. Together, theorem 3 (with $h(x) = x$) and theorem 4 imply the following

**Theorem 5.** *Let $f(F)$ the function defined in theorem 3 and $a, \tilde{b} \geq 0$. Then the $(n \times n)$ matrix*

$$\frac{1}{\sigma^2}\Sigma := (a-\tilde{b})\Pi^{-\frac{1}{2}} f(F)\Pi^{-\frac{1}{2}} + (a+\tilde{b})\Pi^{-1} + (1-2a)J \tag{12}$$

*constitutes a separable auto-covariance.*

Theorem 5 defines a broad class of "flow-induced" spatial models, among which (deriving the relations between parameters is elementary):

- the auto-regressive model $\Sigma = \sigma^2(1-\rho)(I-\rho W)^{-1}\Pi^{-1}$

- *equi-correlated* covariances $\sigma^{-2}\Sigma^2 = \tilde{a}\Pi^{-1} + \tilde{c}J$, with associated geostatistical distances $D_{ii'} = \tilde{a}(1/\pi_i + 1/\pi_{i'})$ for $i \neq i'$. This occurs under contrasted limit flows, namely (A) *perfect mobility flows* $w_{ii'} = \pi_{i'}$ (yielding $f(F) = F = \sqrt{\pi}\sqrt{\pi}'$) and (B) *frozen flows* $w_{ii'} = \delta_{ii'}$ (yielding $f(F) = F = I$).

Irrespectively of the function $f$, any auto-covariance $\Sigma$ defined in theorem 5 must be separable, a testable fact for a given empirical $\Sigma$. Also, the factorial configuration of the set of vertices in a weighted graph or of states in a reversible chain can be obtained by MDS on the associated geostatistical distances $D_{ii'}$. As demonstrated by theorem 3, all those configurations are identical up to dilatations of the factorial axes; in particular, the low-dimensional plot $\alpha = 2, 3$ is invariant up to dilatations, provided $f$ is increasing.

## References

BAVAUD, F. (2004): Generalized factor analyses for contingency tables. In: D.Banks et al. (Eds.): *Classification, Clustering and Data Mining Applications*. Springer, Berlin, 597-606.

BAVAUD, F. and XANTHOS, A. (2005): Markov associativities. *Journal of Quantitative Linguistics, 12, 123-137*.

BENGIO, Y., DELALLEAU, O., LE ROUX, N., PAIEMENT, J.-F. and OUIMET, M. (2004): Learning eigenfunctions links spectral embedding and kernel PCA. *Neural Computation, 16, 2197-2219*.

CHUNG, F. (1997): *Spectral graph theory*. CBMS Regional Conference Series in Mathematics 92. American Mathematical Society. Providence.

DIACONIS, P. and STROOK, D. (1991): Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab., 1, 36-61*.

NG, A., JORDAN, M. and WEISS, Y. (2002): On spectral clustering: Analysis and an algorithm. In T. G. Dietterich et al. (Eds.): *A*dvances in Neural Information Processing Systems 14. MIT Press, 2002.

SHAWE-TAYLOR, J. and CRISTIANINI, N. (2004): Kernel Methods for Pattern Analysis. Cambridge University Press.

VERMA, D. and MEILA, M. (2003): A comparison of spectral clustering algorithms. UW CSE Technical report 03-05-01.