QÜESTIIÓ, Vol 15, 3. pp. 267-278, 1991

LOCAL PRINCIPAL COMPONENTS ANALYSIS

TOMÀS ALUJA-BANET and RAMON NONELL-TORRENT

Facultat d'Informàtica de Barcelona (UPC)

Principal Components Analysis deals mainly with the analysis of large data sets with multivariate structure in an observational context for exploratory purposes. The factorial planes produced will show the main oppositions between variables and individuals. However, we may be interested in going further by controlling the effect of some latent or third variable which expresses some well-defined phenomenon. We go through this by means of a graph among individuals, following the same idea of instrumental variables as Rao, or partial correlation analysis. We call such analysis Local Principal Components Analysis, which consists of defining a semi-metric upon the variables space. Finally, we illustrate this with an example.

Keywords: principal components analysis, local analysis, partial analysis, semi-metric, singular value decomposition.

1. INTRODUCTION

Principal Components Analysis (PCA) looks for a few combinations which can be used to summarize the data, whilst losing in the process as little information as possible; the issued factorial plans reveals the most important features of the data. Very often, such patterns show the effect of some well-known phenomenon, for example a North-South effect in geographical data, a trend effect

-Departament d'Estadística i Investigació Operativa. Facultat d'Informàtica de Barcelona (UPC).

-Article rebut el gener de 1991.

in chronological data, or an income effect in sociological data. It could then be interesting to analyze the data controlling for these phenomena, following the same idea of instrumental variables as Rao [7], or partial correlation analysis. We perform this, avoiding the strong hypotheses implied by the former analyses, by defining a non-oriented graph which expresses a binary relation upon the individuals. This graph can be defined upon some "a priori" relationship on the individuals which we want to eliminate, such as a contiguity relation in geographical data, similar incomes of households in a socio-economic survey, or a trend effect in temporal data.

Local Principal Components Analysis (LPCA) consists of explaining the variation upon the edges of the graph. Thus, the issued patterns of LPCA display the oppositions between related individuals by the graph; for example, if the graph expresses a neighbouring relationship, then LPCA will show the oppositions between neighbouring individuals, that is, the oppositions *given* the location of individuals in the map.

Here we will present the rationale of LPCA starting from the generalized PCA, and following the Caillez-Pages [2] schema; then we introduce the LPCA as a transformation of the original raw data, and finally we will see that this transformation is equivalent to defining a semi-metric upon the variables space.

2. PRINCIPAL COMPONENTS ANALYSIS

Let X be a data matriz of n observartions of p variables with x_{ij} observation of jth variable, $j \in J = \{1, \ldots, p\}$, on the ith individual, $i \in I = \{1, \ldots, n\}$. Let M be the metric matrix in the euclidean individual-space $\mathcal{R}^p$ and D the diagonal metric matrix for $\mathcal{R}^n$ with weights $p_i (p_i > 0, \sum_{i \in I} p_i = 1)$ assigned to individuals.

The following dual schema gives us the relations between the spaces considered:

$$\begin{array}{ccc} \mathcal{R}^p & \xleftarrow{X^t} & \mathcal{R}^{n*} \\ \downarrow M & & \uparrow D \\ \mathcal{R}^{p*} & \xrightarrow{X} & \mathcal{R}^n \end{array}$$

The Principal Components Analysis (PCA) of the triplet (X, M, D) will show the main overall oppositions between variables and individuals. It can be presented as a method of finding the linear combination of variables with the greatest variance, or the method of finding the linear combination with the maximum correlation with the original variables, as a particular case of generalized canonical analysis. It can also be presented as a method of finding a subspace which maximizes the inter-points distances between individuals.

Let $\mathcal{L}_{n,p}$ denote the vector space of matrices with n rows and p columns. In $\mathcal{L}_{n,p}$, the Hilbert Schmidt norm of X is defined as:

$$\|X\|_{D,M} = \left[Tr(X^t DXM)\right]^{1/2},$$

where Tr is the trace function.

When $D = diag(p_1, i \in I)$, then $\|X\|^2_{D,M} = \sum_{i \in I} p_i \|x_i\|^2_M$ is interpreted as the inertia with respect to the origin (or to the barycentre if X is centered) of the set of weighted individuals $\{(x_i, p_i), i \in I\}$.

If X is *rank* r, the Singular Values Decomposition Theorem allows us to decompose X as:

$$X = V\Lambda U^t,$$

with $V^t DV = U^t MU = I$ where I is the identity matrix of order r, V and U are (n, r) and (p, r) matrices, and Λ is the diagonal matrix if the so-called singular values $\lambda_s (\lambda_s > 0)$ ordered by decreasing values.

This decomposition is referred to as the singular values decomposition (s. v. d.) of the triplet (X, M, D).

Let v_s and u_s be the sth columns of V and U respectively. The vectors $\{v_s, s = 1, \ldots, r\}$, left singular vectors of the above decomposition, form an orthonormal basis of the columns of X, and the vectors $\{u_s, s = 1, \ldots, r\}$ right singular vectors, form an orthonormal basis of the rows of X.

Let V_k and U_k be the submatrices built with the k first columns of V and U respectively, and let Λ_k be the diagonal matrix of the k first singular values $(k \leq r)$.

The projection operator onto the subspace $\langle V_k \rangle$ generated by the columns of V_k is denoted by Q_k, and we know that $Q_k = V_k V_k^t D$. In the variables-space, $P_k = U_k U_k^t M$ is the projection operator onto the subspace $\langle U_k \rangle$ generated by the columns of U_k.

Let Q_T be the projection operator onto a subspace T of the space $(\mathcal{R}^n, D)$; similarly, let P_W be the projection operator onto a subspace W of the subspace $(\mathcal{R}^p, M)$.

Then, $\hat{X}^k = V_k \Lambda_k U_k^t$ can be expressed as $\hat{X}^k = X P_k^t = Q_k X$, and we have:

$$\|Q_T X\|_{D.M}^2 \leq \|\hat{X}^k\|_{D.M}^2 \tag{1}$$

and

$$\|X P_W^t\|_{D.M}^2 \leq \|\hat{X}^k\|_{D.M}^2 \tag{2}$$

for all $k-$dimensional subspaces T and W of the variables space and of the individuals space respectively. The upper bound

$$\|\hat{X}^k\|_{D.M}^2 = \sum_{s=1}^{n} \lambda_s^2$$

is attained for $T = \langle V_k \rangle$ and $W = \langle U_k \rangle$ respectively.

Expression 2 can be written as

$$\sum_{s=1}^{n} \lambda_s^2 \geq \|X P_W^t\|_{D.M}^2 = \sum_{i \in I} p_i \|P_W(x_i)\|_M^2, \tag{3}$$

that is the inertia, with respect to the origin, of the projections of individuals onto the subspace W. Then, the subspace $\langle U_k \rangle$ is an optimal choice to keep the maximum inertia. Assuming that M is diagonal ($M = diag(m_j, j \in J)$, and that X is centered, expression 1 can be written as:

$$\|Q_T X\|_{D.M}^2 = \sum_{j \in J} m_j var\left(Q_T(x^j)\right) \leq \sum_{s=1}^{n} \lambda_s^2,$$

where x^j is the jth column of the matrix X. Now, the optimal choice for explaining the variance of the variables is $\langle V_k \rangle$.

On the individuals side, u_s is called the sth principal axis and we deduce that $P_k(x_i) = \sum_{s=1}^{k} \lambda_s v_{si} u_s$. The scalar $\lambda_s v_{si}$, co-ordinate of the ith individual on u_s, is called sth principal co-ordinate of individual i.

Analogously, on the variables side:

$$Q_k(x^j) = \sum_{s=1}^{k} \lambda_s u_{sj} v_s.$$

We know that there exist the following "transition formulae" with the left and right singular values:

$$\begin{aligned} U &= X^t D V \Lambda^{-1} \\ V &= X M U \Lambda^{-1}. \end{aligned}$$

Thus, the coordinate of individuals can be written as $\Psi = XMU$. Likewise the coordinates of variables are $\Phi = X^t DV$.

3. LOCAL PRINCIPAL COMPONENTS ANALYSIS

Let there be a binary relation between individuals with reflexive and symmetric properties. This relation can be represented by means of a non-oriented graph G, where the vertices are the individuals and the edges express the binary relation between individuals. Let Q be the (n, n) matrix associated to the graph and R the diagonal matrix of degrees of vertices.

Theorem 1

$R - Q$ is a positive semi-definite matrix, and it can be expressed as $R - Q = 1/2\, T^t T$, where T is a $(n \times n, n)$ matrix, that crosses the edges with the vertices. An edge joining vertices i and i' is coded by a sequence of zeros and a 1 and -1 in the i and i' columns. If an edge is not defined in the graph, it will be coded as a row of n zeros.

3.1 LPCA as a transformation of data matrix

Thus, T is an operator of differences of variables between related individuals by the graph; $\mathcal{R}^n \xrightarrow{T} \mathcal{R}^{n^2}$ and TX is a $(n \times n, n)$ matrix of differences.

Let us take $L = D \otimes D$ as a metric for $\mathcal{R}^{n^2}$. That is, we define as a weight of an edge the product of weights of the corresponding vertices of the edge.

The dual schema would now be:

$$\begin{array}{ccc} \mathcal{R}^p & \xleftarrow{(TX)^t} & \mathcal{R}^{n^{2*}} \\ \downarrow M & & \uparrow L \\ \mathcal{R}^{p*} & \xrightarrow{TX} & \mathcal{R}^{n^2} \end{array}$$

Local Principal Components Analysis (LPCA) consists of the singular value decomposition of the triplet (TX, M, L). Thus:

$$TX = V\Lambda U^t \quad \text{with} \quad V^tLV = U^tMU = I.$$

This leads to finding a subspace H which maximises, according to 3:

$$\|TXP_H^t\|_{L.M}^2 = \sum_{i,i' \in G} p_i p_{i'} d_H^2(i,i'),$$

that is, the interpoint distances related by the graph G, projected onto a subspace H. This implies the diagonalization of matrix $A = X^tT^tLTXM$ [1]. This matrix coincides with the local covariance matrix $1/mX^t(R-Q)X$ for the usual case of $p_i = 1/\sqrt{m}$ and $M = I$ (identity matrix), and with the contiguity matrix if $M = S^{-2}$ [6] (diagonal matrix of the inverse of variance of variables); in the latter case the diagonal of matrix A coincides with the Geary coefficients of contiguity of variables [4], whereas they coincide with the local covariance of variables in the former case.

Then the Hilbert Schmidt squared norm of matrix TX is $trace(A)$, which in the particular case of M diagonal, can be written as

$$tr(A) = \sum_j \sum_{i,i' \in G} p_i p_{i'} m_j (x_{ij} - x_{i'j})^2,$$

the measure of overall dispersion for the LPCA; which can be interpreted in the usual cases as the sum of local variances of variables or the sum of Geary coefficients of variables.

In LPCA we have the following transition formulae

$$\begin{aligned} U &= X^tT^tLV\Lambda^{-1} \\ V &= TXMU\Lambda^{-1}. \end{aligned}$$

The co-ordinates of edges (rows of TX) are $\Upsilon = TXMU = T\Psi$, where Ψ are the coordinates of individuals. The co-ordinates of variables are $\Phi = X^tT^tLV$.

3.2 LPCA as a semi-metric in $\mathcal{R}^n$

Local Principal Components Analysis can also be viewed as the s.v.d. of (X, M, T^tLT); notice that now T^tLT need not be a metric, but is always a semi-metric.

If we add to a triplet (X, M, D), the (n^2, n) matrix T and a metric L in $\mathcal{R}^{n^2}$, we have:

$$\begin{array}{ccccc} \mathcal{R}^p & \xleftarrow{X^t} & \mathcal{R}^{n*} & \xleftarrow{T^t} & \mathcal{R}^{n^{2*}} \\ \downarrow M & & \uparrow D & & \uparrow L \\ \mathcal{R}^{p*} & \xrightarrow{X} & \mathcal{R}^n & \xrightarrow{T} & \mathcal{R}^{n^2} \end{array}$$

Let us also consider that $rank(TX) = r$.

Then, the s.v.d. of (TX, M, L) gives us:

$$TX = V\Lambda U^t$$

whit $V^tLV = U^tMU = I$.

Let W be the subspace generated by all the rows of TX; the columns of the matrix U form an orthonormal basis of this subspace. Let also $W_k = \langle U_k \rangle$ be the subspace generated by the first k columns of U. As we have already seen, the subspace W_k is an optimal choice for keeping $Tr(P_S X^t T^t LTXM)$ maximum among all $k-$dimensional subspaces S of W. Now P_W, the projection operator onto W, can be written as UU^tM; this operator allows us to consider the following decomposition:

$$X = XP_W^t + X(I - P_W^t).$$

The projection onto W of any row of TX is the row itself; then, $TXP_W^t = TX$ and, therefore, $TX(I - P_W^t) = 0$.

Let us also consider

$$XP_W^t = XMUU^t = \tilde{V}\Lambda U^t,$$

where $\tilde{V} = XMU\Lambda^{-1}$ and $\tilde{V}^t(T^tLT)\tilde{V} = I$.

Summarizing,

$$X = \tilde{V}\Lambda U^t + X(I - {}^tP_W)$$

with $\tilde{V}^t\Delta\tilde{V} = U^tMU = I$, where Δ is the semi-metric T^tLT. Using the same notation as in PCA, this is to say that $\hat{X}_k = \tilde{V}_k\Lambda_k U_k^t$ can be considered as the closest matrix to X in the sense of the semi-metric Δ; notice that $\tilde{V}$ is $rank\ r$ (if, as usual, $r \leq n$).

Finally, we have $V = T\tilde{V}$; this illustrates the fact the variables have the same co-ordinates in the two approaches that we have developed: $(TX)^tLV =$

$X^t(T^tLT)\tilde{V}$. Obviously, the co-ordinates of the individuals are XMU while the co-ordinates of the edges are $TXMU$.

Then, the Local Principal Components Analysis consists of decomposing the triplet (X, M, Δ), where $\Delta = T^tLT$ is the semi-metric induced by the matrix T, obtained when crossing edges and vertices, and $L = D \otimes D$ is the metric induced by D on the edges.

Therefore, as we have seen, the two approaches that we have developed lead to the same projections using Local Principal Components Analysis.

4. RELATION BETWEEN LOCAL AND GLOBAL ANALYSES

Let B be a (n^2, n) matrix crossing the edges and vertices for a complete graph. It can be easily shown that the LPCA of BX is equivalent to the usual PCA of X. Moreover, we can decompose the columns of BX into two parts: $BX = TX + (B - T)X$, one in the local space and the other orthogonal to it. Since the number of edges of a contiguity or similarity graph is far lower than the complete graph, the local variable would be very much shorter than the global one. For this reason we weigh each individual by $p_i = 1/\sqrt{m}$ and not by the classical $1/n$; this involves expanding the local variable by a factor of n^2/m. The LPCA means analyzing only the variables projected into this local space.

Consequently, we can obtain the decomposition of the global (total) variability into two components: one, the local variability, expressing the oppositions between related individuals by the graph, and second, the outer variability to the graph.

We could be interested in evaluating the strength of relation of the global variables with their counterpart local ones. The projection of global variables upon the local space are the local ones. Thus the covariance matrix with both type of variables coincides, leaving aside a factor of expansion, with the local variance matrix.

$$V_{gl} = \frac{1}{n\sqrt{m}} X^tB^tTX = \frac{1}{n\sqrt{m}} X^t(R - Q)X.$$

Moreover, we can visualise the shift when moving from global variables to the local ones, by projecting both types of variables onto the same basis; the most natural choice is to take as a basis the factorial axis obtained from the global variables, and to project as supplementary the local ones.

5. EXAMPLE OF APPLICATION

In order to illustrate how *LPCA* works in practice we have taken a small data set concerning the 38 regions of Catalonia, giving for each of them the number of municipalities according to their altitude above the sea level. We choose $D = 1/nI$ and $M = S^{-2}$ as metrics for R^n and R^p respectively. In the following table we give the formed data matrix:

	Number of municipalities				
	[0m., 100m)	*[100m., 200m)*	*[200m., 600m)*	*[600m., 1000m)*	*≥ 1000m.*
Baix Llobregat	14	10	3	0	0
Barcelonès	5	2	0	0	0
Maresme	18	11	1	0	0
Vallès Occidental	4	8	11	0	0
Vallès Oriental	6	15	16	5	0
Alt Empordà	50	14	4	0	0
Baix Empordà	35	1	0	0	0
Garrotxa	0	2	19	0	0
Gironès	12	20	5	1	0
La Selva	10	9	6	1	0
Alt Camp	0	3	19	1	0
Alt Penedès	0	4	19	2	0
Baix Penedès	4	6	2	0	0
Garraf	4	2	2	0	0
Tarragonès	12	10	0	0	0
Baix Camp	2	6	15	4	0
Conca de Barberà	0	0	11	10	0
Priorat	0	1	20	3	0
Ribera d'Ebre	10	2	2	0	0
Baix Ebre	11	1	2	0	0
Montsià	5	4	2	0	0
Terra Alta	0	1	11	0	0
Cerdanya	0	0	0	0	16
Osona	0	0	23	23	1
Ripollès	0	0	2	12	10
Anoia	0	0	21	13	0
Bages	0	5	24	6	0
Berguedà	0	0	4	16	10
Solsonès	0	0	2	10	2
Les Garrigues	0	0	20	5	0
Noguera	0	0	33	2	0
Segarra	0	0	15	6	0
Segrià	2	12	28	0	0
Urgell	0	0	25	1	0
Alt Urgell	0	0	5	9	5
Pallars Jussà	0	0	4	10	3
Pallars Sobirà	0	0	1	6	8
Val d'Aran	0	0	0	7	2

Catalonia is a small country which extends from sea level to the high altitudes of the Pyrenees. Thus, we expect that a global analysis of such data will show these differences of altitude, defining the main opposition between regions. On the other hand, given that there is a contiguity relation upon the regions, it may be interesting to analyze such data matrix by eliminating the geographical location of regions.

Certainly, the overall (or classical) *PCA* of the defined triplet gives us a graphical display of the variables responsible for the oppositions between regions. In Figure 1 we can see (in capital letters) the first axis opposing municipalities below 200m. of altitude against municipalities above that altitude. Moreover, there appears a horse-shoe effect. We also give over the same display the projection of local variables as illustrative ones. Comparing these projections with their global counterpart, we can appreciate a shift to the origin; that is, local variables are "shorter" than the corresponding overall ones, indicating that variables are positive correlated upon the contiguity graph.

Finally, by performing an *LPCA* we can see in Figure 2 that the main opposition is now defined by the medium altitudes (between 200m. and 600m.) against higher and lower altitudes than the previous interval referred to. This reveals that one region normally has regions of similar altitude as neighbours; that is, it is very likely that one region at sea level will normally have as a neighbour another region with a medium level of altitude, and a region in the Pyrenees, will currently have as a neighbour another region with medium altitude; in both cases there will appear edges in which the main difference is defined by the medium altitudes (between 200m, and 600m.) or by the high or low altitudes (above 600m. or below 200m.). However, although they will be very unusual, it is possible to find edges defined by the difference between two regions, one at sea level and another in the high altitudes of the Pyrenees. This means that, keeping the geographical location constant, that is, for neighbour regions, the opposition is defined between opposition between 200m. and 600m. against below and above that interval. Finally, comparing the local oppositions with the overall analysis, we can see in this case that the first main local axis in fact corresponds to the second axis of global analysis.

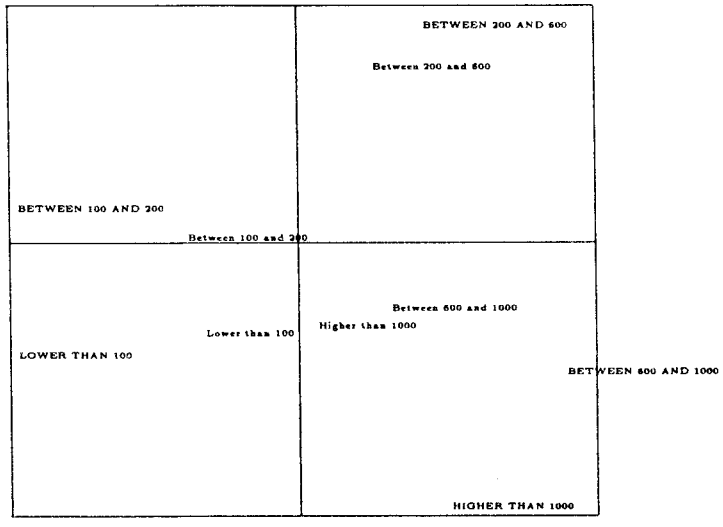

Figure 1. First graphical plan of overall analysis.

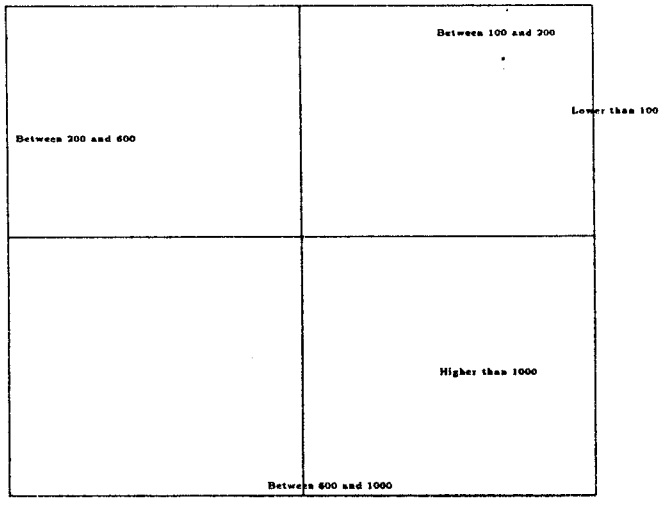

Figure 2. First graphical plan of local analysis.

6. REFERENCES

[1] **T. Aluja-Banet** (1988). "Local and Partial Correspondence Analysis. Application to the Analysis of Electoral Data". *Computational Statistics Quarterly,* **2**, 89–103.

[2] **F. Caillez** y **J.P. Pages** (1976). *Introduction à l'Analyse des Données.* SMASH. París.

[3] **A. Carlier** (1985). "Analyse des évolutions sur tables de contingence: Quelques aspects operationels". *4èmes. Journées Internationales sur Analyse des Données et Informatique.* INRIA, Versailles.

[4] **R.C. Geary** (1954). "The contiguity ratio and statistical mapping". *The Inc. Statistician,* 115–145.

[5] **M.J. Greenacre** (1984). *Theory and Applications of Correspondence Analysis.* Academic Press. London.

[6] **L. Lebart** (1969). "Analyse statistique de la contigüité". *Publ. ISUP,* **18**, 81–112.

[7] **C.R. Rao** (1964). "The use and interpretation of principal components analysis in applied research". *Sankhya,* **26**, 329–357.