

Some results on marginal nonlinear principal components

Componenti marginali non lineari: alcuni risultati

Aldo Goia and Ernesto Salinelli

Dip. di Scienze Economiche e Metodi Quantitativi - Università del Piemonte Orientale
e-mail: aldo.goia@eco.unipmn.it

Riassunto: Nel presente lavoro si studiano le componenti principali non-lineari marginali introdotte dal punto di vista teorico in Salinelli (1998 e 2001). L'obiettivo è quello di illustrarne il fondamento statistico (anche mediante il confronto con la definizione di componenti principali non-lineari proposta in De Leeuw *et al.*, 1981) e la valenza applicativa.

Keywords: Nonlinear Principal Component, Marginal Nonlinear Principal Components, B-spline Estimation, Entropy.

1. Introduction

The problem of nonlinear extensions of the classical *linear principal component analysis* has been treated in a large literature in order to face nonlinear dependencies between variables. We recall, for example, the works of De Leeuw *et al.* (1981), Koyak (1987), Besse and Ferraty (1995) which obtained nonlinear principal components by minimizing a loss function, and the paper of Donnell *et al.* (1994) which defined and studies the so-called *additive principal components*. Salinelli (1998 and 2001) proposed a more general definition of nonlinear principal components for continuous random variables with bounded densities studying its mathematical properties. In these works, the author sketched the definition of *marginal nonlinear principal components* (MNLPC) from a theoretical point of view for uniform and normal distributions.

The aim of the present paper is to analyze the definition of MNLPC, to illustrate its statistical foundation (through comparison with the definition of nonlinear principal component proposed in De Leeuw *et al.*, 1981) and its worthiness. We also propose an estimator based on B-spline approximation. We focus on MNLPC since the analysis represents a first step in the direction of the more general nonlinear principal components defined in Salinelli (1998) and for its analogies with the De Leeuw's *et al.* results.

2. Marginal nonlinear principal components

Let $\mathbf{X} = [X_1, X_2, \dots, X_q]^T$ be a continuous q -dimensional random vector (r.v.) with $\mathbb{E}[\mathbf{X}] = \mathbf{0}$, finite second moments $\Sigma_{\mathbf{X}} = \mathbb{E}[\mathbf{X}\mathbf{X}^T]$ and density $f_{\mathbf{X}}$ with support a domain (an open and connected set) $D \subseteq \mathbb{R}^q$. We denote with $\mathcal{L}^2 = \mathcal{L}^2(D, f_{\mathbf{X}})$ the space of square integrable centered real functions on D and with $\mathcal{L}_j^2 = \mathcal{L}_j^2(D_j, f_{X_j})$ the q spaces of square integrable centered real functions on the interval $D_j \subseteq \mathbb{R}$. $\mathcal{L}^2$ and $\mathcal{L}_j^2$ are Hilbert spaces with respect to the usual inner product $\langle \cdot, \cdot \rangle_0$ and $\langle \cdot, \cdot \rangle_{0,j}$ respectively. If $\mathcal{L}_{\text{loc}}^1(D_j)$ denotes the space of locally integrable functions defined on D_j , we introduce

the weighted Sobolev space

$$W_j^{1,2} = W^{1,2}(D_j, f_{X_j}) = \{u : u \in \mathcal{L}_j^2 \cap \mathcal{L}_{\text{loc}}^1(D_j), u' \in \mathcal{L}_j^2 \cap \mathcal{L}_{\text{loc}}^1(D_j)\} \quad (1)$$

where derivatives have to be intended in a weak sense. On each $W_j^{1,2}$ we consider the inner product $\langle u, v \rangle_{1,j} = \langle u', v' \rangle_{0,j}$ and the induced norm $\|\cdot\|_{1,j}$.

Definition 1. The j -th marginal nonlinear principal component is defined as the j -th linear principal component (LPC) of the r.v. $\Psi(\mathbf{X}) = [\psi_1(X_1), \psi_2(X_2), \dots, \psi_n(X_n)]^T$ where each $\psi_j \in W_j^{1,2}$ is the solution of the maximization problem

$$\begin{cases} \max_{\varphi} & \mathbb{E}[(\varphi(X_j))^2] \\ \text{sub} & \mathbb{E}[(\varphi'(X_j))^2] = 1 \end{cases} \quad (2)$$

If there exists, the solution of each maximization problem in definition 1 is the dominant eigenfunction of the differential operator $\mathcal{D}_j$ defined by

$$\mathcal{D}_j(\varphi) = -\varphi'' - \varphi' \frac{f'_{X_j}}{f_{X_j}}$$

associated to its first characteristic value λ_j (i.e. the inverse of the eigenvalue $1/\lambda_j$).

In our work we analyze some results on the existence and some properties (monotonicity, existence of zeroes, etc.) of these eigenfunctions.

In particular, for what concern the relative entropy (if there exists) H of each X_j , defined by $H(X_j) = -\mathbb{E}[\ln f_{X_j}]$, we prove the following result:

Lemma 2. Consider a continuous random variable X valued in the interval I with density f_X and entropy $H(X)$, and the transformation $h : I \rightarrow \mathbb{R}$ differentiable and strictly monotonic on the intervals I_s with $s = 1, \dots, m$, $\cup_{s=1}^m I_s = I$ and $I_s \cap I_j = \emptyset$ for $s \neq j$. If $h \in W^{1,2}(I, f_X)$ with $\mathbb{E}[(h')^2] = 1$, and $Y = h(X)$ then $H(Y) \leq H(X)$.

Furthermore, given a random variable X with zero mean and variance σ_X^2 , if $J(X)$ denote the so-called *negentropy*,

$$J(X) = H(X_{\text{gauss}}) - H(X)$$

where $X_{\text{gauss}} \sim \mathcal{N}(0, \sigma_X^2)$, we prove

Proposition 3. Consider a continuous random variable X with density f_X and entropy $H(X)$, defined on the interval I . If $\psi \in W^{1,2}$ is a solution of (2), then $J(\psi(X)) \geq J(X)$.

In the light of the latest result, it seems interesting to study the MNLPC as a method of de-gaussianization. Furthermore, since the optimal transformations are S-shaped, we explore the possibility of employment of the MNLPC when we are in presence of mixtures and we want to improve the performances of the classical LPCA for detecting clusters.

3. MNLPC and De Leeuw's et al. NLPC

In De Leeuw *et al.* (1981) the first p ($1 \leq p \leq q$) non linear principal components of a continuous r.v. $\mathbf{X}$ are introduced as the solution of the following minimization problem

$$\begin{cases} \min_{A, \mathbf{Y}, \Phi} \frac{1}{q} \sum_{j=1}^q \text{Var} \left[\phi_j(X_j) - \sum_{i=1}^p a_{ji} Y_i \right] \\ \text{sub } \mathbb{E}[\phi_j(X_j)] = 0, \quad \mathbb{E}[\phi_j^2(X_j)] = 1, \quad j = 1, \dots, q \end{cases} \quad (3)$$

where $\mathbf{Y} = [Y_1, \dots, Y_q]^T$ is a r.v. defined on the same probability space of $\mathbf{X}$, $\Phi = [\phi_1, \dots, \phi_q]^T$ is a q -dimensional vector of centered functions $\phi_i \in \mathcal{L}_j^2$ and A is a $q \times p$ matrix. To solve problem (3) is equivalent to look for the first p eigenvalues of the correlation matrix $R(\Phi(\mathbf{X}))$.

With respect a complete orthonormal basis $\{g_{js}\}_{s \in \mathbb{N}}$ of $\mathcal{L}_j^2$ it is possible to develop ϕ_j as

$$\phi_j = \sum_{s=0}^{+\infty} b_{js} g_{js} = \sum_{s=0}^{+\infty} \langle \phi_j, g_{js} \rangle g_{js}. \quad (4)$$

Denoting by C_{st} the covariance matrix (of order q) $C_{st} = [\mathbb{E}[g_{is}(X_i) g_{jt}(X_j)]]_{i,j=1, \dots, q}$ and $\mathbf{b}_k = [b_{rk}]_{r=1, \dots, q}$ since

$$\mathbb{E}[\phi_j(X_j), \phi_i(X_i)] = \sum_{s=0}^{+\infty} \sum_{t=0}^{+\infty} b_{js} b_{it} \mathbb{E}[g_{js}(X_j), g_{it}(X_i)]$$

it is possible to restate problem (3) as

$$\begin{cases} \max \sum_{s=0}^{+\infty} \sum_{t=0}^{+\infty} \mathbf{b}_s^T C_{st} \mathbf{b}_t \\ \text{sub } \sum_{s=0}^{+\infty} \mathbf{b}_s^T \mathbf{b}_s = 1 \end{cases}$$

We note that, if there exist MNLPC, each transformation ψ_j represents the first (with respect to the variance ordering) element of an orthonormal basis of $W_j^{1,2}$. By a density argument, $\lambda_{1,j}^{-1/2} \psi_j$ represents the “first element” of the corresponding orthogonal basis of $\mathcal{L}_j^2$.

This result is interesting since in general it is not guaranteed that the elements g_{js} have variances decreasing with s except that when r.v. is normal (see De Leeuw et al., 1981, p. 385). Thus, in practice to use the MNLPC, by introducing a standardization constraint on the transformations, is equivalent to solve an eigenvalue problem on the matrix C_{11} (see De Leeuw et al., 1981, p. 381) but with a choice of basis elements optimal in a variance sense.

4. B-spline Estimation

The estimation problem involves a two step procedure: first we obtain an estimation $\hat{\psi}_j$ of the transformations ψ_j and therefore we apply the classical LPC algorithm to the vector

$\widehat{\Psi}(\mathbf{X})$.

About the first step, let $\{[X_{1,i}, X_{2,i}, \dots, X_{q,i}]^T, i = 1, \dots, n\}$ be a sample of size n of i.i.d. r.v. drawn from $\mathbf{X}$, the sampling problem of (2) may be written as q independent maximization problems:

$$\begin{cases} \max_{\varphi} & \frac{1}{n} \sum_{i=1}^n \varphi^2(X_{ji}) \\ \text{sub} & \frac{1}{n} \sum_{i=1}^n (\varphi')^2(X_{ji}) = 1 \end{cases} \quad (5)$$

We consider now the set $S_{j,k,d}$ of B-spline functions $B_{j,s}$ defined on a compact set $\tilde{D}_j \subset D_j$ with degree d and k knots. The spline approximation of ψ_j is

$$\tilde{\psi}_j = \sum_{s=1}^{k+d} \theta_{js} B_{js} = \theta_j^T B_j$$

where $\theta_j = (\theta_{j1}, \dots, \theta_{j,k+d})^T \in \mathbb{R}^{k+d}$. Thus we may write an approximated spline version of (2) as

$$\begin{cases} \max_{\theta_j} & \theta_j^T M_j \theta_j \\ \text{sub} & \theta_j^T M_j^* \theta_j = 1 \end{cases} \quad (6)$$

where M_j and M_j^* are $(k+d) \times (k+d)$ matrices defined as

$$M_j = \frac{1}{n} \sum_{i=1}^n B_j(X_{ji}) B_j(X_{ji})^T \quad M_j^* = \frac{1}{n} \sum_{i=1}^n B_j'(X_{ji}) B_j'(X_{ji})^T.$$

Maximization problem (6) leads to solve the generalized eigenvalue problem of the pencil $(M_j - \lambda M_j^*)$ whose solution can be found using the QZ algorithm (Moler and Stewart, 1973). We also prove some asymptotic results for the above estimator.

References

- Besse P.C., Ferraty F. (1995) A fixed effect curvilinear model, *Computational Statistics*, 10, 339-351.
- De Leeuw J., Van Rijekevorsel H., Van der Wouden H. (1981) Nonlinear Principal Components Analysis with B-splines, *Methods of Operations Research*, 33, 379-393.
- Donnell D.J., Buja A., Stuetzle W. (1994) Analysis of additive dependencies and concutivities using smallest additive principal components, *The Annals of Statistics*, 22 (4), 1635-1668.
- Koyak R.A. (1987) On measuring internal dependence in a set of random variables, *The Annals of Statistics*, 15, 1215-1228.
- Moler C.B., Stewart G.W. (1973). Am algorithm for generalized matrix eigenvalue problems. *SIAM J. Numer. Anal.*, 10, 241-256.
- Salinelli E. (1998) Nonlinear principal components I. Absolutely continuous random variables with positive bounded densities, *The Annals of Statistics*, 26 (2), 596-616.
- Salinelli E. (2001) Nonlinear Principal Components II: The Normal Distribution, W.P. n.7 del Dipartimento di Scienze Economiche e Metodi Quantitativi, Università del Piemonte Orientale "A. Avogadro".