The Annals of Statistics
1998, Vol. 26, No. 2, 596–616

NONLINEAR PRINCIPAL COMPONENTS I. ABSOLUTELY CONTINUOUS RANDOM VARIABLES WITH POSITIVE BOUNDED DENSITIES[1]

BY ERNESTO SALINELLI

University of Torino

Nonlinear principal components for an absolutely continuous random vector $\mathbf{X}$ with positive bounded density are defined as the solution of a variational problem in a suitable function space. In this way transformations depending on all the components of $\mathbf{X}$ are obtained. Some properties of nonlinear principal components are proved: in particular, it is shown that the set of nonlinear principal transformations of $\mathbf{X}$ is an orthonormal basis for the function space associated with the optimal problem. The spectral decomposition of $\mathbf{X}$ and its covariance matrix with respect to this basis are given. A notion of marginal nonlinear principal components is sketched and the relations with nonlinear principal components are shown. Finally, treating the case of random vectors distributed on unbounded domains, the existence problem is shown to be related to the global existence of the moment generating function of $\mathbf{X}$. Since it is not restrictive, definitions and results are stated in terms of a uniformly distributed random vector.

1. Introduction. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ be a real random vector (r.v.) with zero expectation $\mathbf{E}(\mathbf{X})$ and positive definite covariance matrix $\mathbf{\Sigma}$ having distinct eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$.

A linear operator T: $\mathbf{R}^n \to \mathbf{R}^n$ is said to be a *linear principal component transformation* if each component T_j: $\mathbf{R}^n \to \mathbf{R}$, $j = 1, 2, \ldots, n$, is a solution of the maximum problem

$$\begin{aligned} &\text{maximize} && \mathbf{E}\big((A'\mathbf{X})^2\big), \\ &\text{subject to} && \|A\|^2 = 1, \qquad A \in L(n), \\ &&& \mathbf{E}\big((A'\mathbf{X})(T_s'\mathbf{X})\big) = 0, \qquad s = 1, 2, \ldots, j-1,\ j > 1, \end{aligned} \tag{1}$$

where $\|\cdot\|$ denotes the usual sup-norm on $L(n)$, the linear n-dimensional space of linear functionals on $\mathbf{R}^n$.

The set of n finite-dimensional variational problems (1) always admits as a solution an orthogonal operator T whose components are a set of orthonormal eigenvectors T_j, $j = 1, 2, \ldots, n$, of $\mathbf{\Sigma}$ corresponding to the eigenvalues λ_j. Since we assume that eigenvalues λ_j are distinct, the solution is unique modulo n-dimensional transformations of the form $S = \operatorname{diag}\{\pm 1, \pm 1, \ldots, \pm 1\}$. The n components Y_j of the r.v. $\mathbf{Y} = T\mathbf{X}$ are called *principal components* [in the sequel, linear principal components (LPCs)]. From (1) it follows that $\lambda_j =$

Received November 1995; revised October 1997.
[1]Research supported in part by M.U.R.S.T.
AMS 1991 *subject classifications*. Primary 62H25; secondary 35J20.
Key words and phrases. Nonlinear principal components, moment generating function, Sobolev spaces, Laplacian.

$\mathrm{Var}(Y_j)$ and so $\mathrm{Var}(Y_1) > \mathrm{Var}(Y_2) > \cdots > \mathrm{Var}(Y_n) > 0$. If the components X_j of $\mathbf{X}$ are largely correlated, then by definition the first few LPCs will account for most of the variation in the original variables. Conversely, the last few LPCs identify directions with smaller variation, that is, approximately linear relationships among the original variables.

The definition introduced above is essentially due to Hotelling (1933), but several other equivalent characterizations of LPCs are known [see, e.g., Jolliffe (1986)]. Probably the most important of them is owing to Pearson (1901): LPCs are introduced as solutions of the geometric optimization problem that consists of finding lines and planes which best fit a set of points in an n-dimensional space. The best-fitting line identifies the first LPC, and the direction of the last LPC is orthogonal to the best-fitting plane.

Several authors in the past pointed out the importance of considering transformations that are sensitive to the covariability of the components X_j through their higher moments in order to detect nonlinear relationships between them. This fact motivated many attempts to define "generalized" or "nonlinear" principal components: we recall some of the most important among them.

Following the Hotelling approach, in Gnanadesikan [(1977), page 53], LPC analysis was performed on a r.v. whose components are obtained from $\mathbf{X}$ by "completely specified, but otherwise arbitrary, functions of the original variables." In particular, for $n = 2$, *quadratic principal components* are defined as the LPCs of the r.v. $\tilde{\mathbf{X}} = (X_1, X_2, X_1^2, X_1X_2, X_2^2)'$. In the same way *polynomial principal components* can be defined to detect approximate polynomial relationships in multivariate data, but "practical experience suggests that generally the method is not very successful because it is heavily affected by small errors in the data" [Flury (1994)]. Furthermore, the normalizing Euclidean constraint on the coefficients of the nonlinear terms appears to be theoretically not well founded; moreover, higher-order nonlinear components are not orthogonally invariant, as pointed out in Mizuta (1984) for the quadratic case.

More recently, Donnell, Buja and Stuezle (1994) defined, for a standardized r.v. $\mathbf{X}$, the kth smallest *additive principal component* (APC) as an additive function $\sum_{i=1}^n \phi_i^{(k)}(X_i)$ of the components X_i such that the r.v. $\mathbf{\Phi}^{(k)} = (\phi_i^{(k)})_i$ solves

$$
\begin{aligned}
&\text{minimize} \quad \mathbf{E}\left(\left(\sum_{i=1}^n \phi_i(X_i)\right)^2\right) \\
&\text{subject to} \quad \sum_{i=1}^n \mathbf{E}\big((\phi_i(X_i))^2\big) = 1, \qquad \phi_i \in H_i, \\
&\qquad\qquad \sum_{i=1}^n \mathbf{E}\big(\phi_i(X_i), \phi_i^{(s)}(X_i)\big) = 0, \qquad s = 1, 2, \ldots, k-1,
\end{aligned} \tag{2}
$$

where the H_i's are closed subspaces of centered, square integrable functions. In this way the authors were able to study nonlinear additive relations among predictor variables in additive regression models and to investigate dependencies of low codimension considering all the variables symmetrically. The

APCs can be considered a genuine generalization of the LPCs. Indeed, when $H_1 \times \cdots \times H_n = L(n)$, problem (2) admits as solutions the usual LPCs; moreover, the APCs are solutions of an opportune eigenproblem. However, whereas LPCs are defined for both covariance and correlation matrices, APC analysis works only in the second case: if $\mathbf{X}$ is not standardized and $\phi_i(X_i) = a_i X_i$, the normalization constraint $\sum_{i=1}^{n} \mathbf{E}((\phi_i(X_i))^2) = 1$ does not coincide with the classical $\|a\|^2 = 1$. Furthermore, as pointed out by the authors, "APCs can only capture structure in pairwise marginals.... Structure that depends on higher-order marginals is elusive for this class of techniques...."

We recall also the works of De Leeuw, Van Rijekevorsel and Van der Wouden (1981), Koyak (1987) and Gifi (1990), where nonlinear principal components of a standardized r.v. were defined by minimizing a loss function, but each transformed variable depended only on the corresponding component of the initial r.v.

Generalizing Pearson's work, Hastie and Stuetzle (1989) defined a *principal curve* (PC) for the density $f_{\mathbf{X}}$ of an absolutely continuous r.v. $\mathbf{X}$ with $\mathbf{E}(\mathbf{X}) = \mathbf{0}$ and finite second moments as any smooth and compact curve Γ such that $\mathbf{E}(\mathbf{X} \mid \pi_\Gamma(\mathbf{X}) = x) = x$ for almost every $x \in \Gamma$. The map π_Γ: $\mathbf{R}^n \to \Gamma$ assigns to each $x \in \mathbf{R}^n$ a point $\pi_\Gamma(x) \in \Gamma$ realizing the distance from x to Γ. Roughly speaking, a principal curve passes through the middle of a data set, in the sense that every point on the curve is the average of the observations projecting onto it. The idea is to introduce a theoretically well-founded method to fit one-dimensional nonlinear manifolds to multivariate data. The authors show that if a PC is a straight line, then it is a LPC and moreover that PC are critical points of the distance function $d^2(\mathbf{X}, \Gamma) = \mathbf{E}(\|\mathbf{X} - \pi_\Gamma(\mathbf{X})\|^2)$: if Γ_t is a smooth family of curves such that $\Gamma_0 = \Gamma$, then

$$\frac{d}{dt} d^2(\mathbf{X}, \Gamma_t)|_{t=0} = 0.$$

Although the authors give an algorithm to find the PCs, they are unable to prove existence results, to estimate how many different PCs can be found for a given distribution or to investigate their properties.

More recently, Duchamp and Stuetzle (1993, 1995a, b), studying PCs in the plane, showed that, under appropriate assumptions on Γ and $f_{\mathbf{X}}$, PCs are solutions of a second-order ordinary differential equation. In this way it is possible to compute PCs: in particular, the cases of a uniform density on rectangles and annuli were analyzed. Unfortunately the main result of the papers mentioned above is negative: whereas the largest and the smallest LPCs are the maximum and minimum of the distance $d^2(\mathbf{X}, \Gamma)$, respectively, all the PCs are only saddlepoints. This result in particular implies that cross-validation is not a viable method for choosing model complexity in the estimation of PCs.

In this paper we introduce a notion of *nonlinear principal components* (NLPCs) by extending problem (1) to a function space more general than $L(n)$ obtaining, from a mathematical point of view, a variational problem on an infinite-dimensional Hilbert space. In general, this problem seems very difficult to deal with; for this reason, we have restricted our attention to absolutely

continuous r.v.s with positive bounded densities. In this way we have been able to answer several questions and obtain a basic model that we hope to extend in our future research to more general settings. Since it is not restrictive, for the sake of simplicity, definitions and results are stated in terms of a uniformly distributed r.v.

The notion of NLPC makes it possible to overcome some of the problems that limit the above-mentioned approaches. First of all, since each NLPC is a function of all the components of $\mathbf{X}$, it captures structure that depends on marginals of every order. Second, NLPCs can be defined even for nonstandardized r.v.s as done in the sequel. Finally, NLPCs can be considered to be a theoretically well-founded generalization of LPCs for several reasons: they solve a maximum variance problem, are invariant under orthogonal transformations, their variances are the eigenvalues of an opportune integral operator and the corresponding eigenfunctions form an orthonormal basis for the function space where the maximum variance problem is solved.

In what follows, we "translate" (see Section 2) problem (1) to a suitable space of square integrable functions and introduce the main definition of NLPCs. Moreover, using some classical results in partial-differential equations (PDEs) literature, we study the existence problem. In Section 3 some properties of NLPCs similar to the ones of the linear case are proved. Subsequently, Section 4 deals with the particular cases of a uniform r.v. on rectangles and annuli: NLPCs are given and it is shown that the nonlinear eigenstructure of $\mathbf{X}$ contains information on size and length of the domain considered. For the rectangle, a definition of nonlinear principal components based on the marginals (MNLPCs) is sketched. It is proved that in general the first n NLPC variances are greater than the corresponding MNLPCs, whereas equality holds only under particular assumptions on the shape of the rectangle. Finally, in Section 5, we treat the case of a r.v. distributed on an unbounded domain. In this way the statistical meaning of the conditions which guarantee the existence of NLPCs is clarified in terms of the moment generating function.

2. Definition and existence of NLPCs. Let $\mathbf{X}$ be an absolutely continuous real r.v. with positive bounded density $f_{\mathbf{X}}$ on a domain (an open, connected set) $\Omega \subset \mathbf{R}^n$, that is, there exist positive constants c_1 and c_2 such that

$$0 < c_1 \le f_{\mathbf{X}}(x) \le c_2 \qquad \forall x \in \Omega. \tag{3}$$

We also suppose, as usual, that $\mathbf{X}$ has finite second moments and $\mathbf{E}(\mathbf{X}) = \mathbf{0}$; furthermore, for the sake of simplicity, we assume $m\Omega = 1$.

We begin with observing that problem (1) can be restated as

$$\begin{aligned} &\text{maximize } \mathbf{E}\big((A'\mathbf{X})^2\big), \\ &\text{subject to } \|\nabla A\|^2 = 1, \qquad A \in L(n), \\ &\quad \mathbf{E}\big((A'\mathbf{X})(T_s'\mathbf{X})\big) = 0, \qquad s = 1, 2, \ldots, j-1,\ j > 1, \end{aligned} \tag{4}$$

where $\nabla = (\partial/\partial x_i)_{i=1,2,\ldots,n}$ stands for the gradient operator. It is clear that if we want to extend problem (4) to nonlinear functions $\psi\colon \Omega \to \mathbf{R}$, we have to

require them to belong to $\mathscr{L}^2_{f_{\mathbf{X}}} = \mathscr{L}^2_{f_{\mathbf{X}}}(\Omega) = \{u\colon \int_\Omega |u|^2 f_{\mathbf{X}}\, dx < +\infty\}$ equipped, as usual, with the inner product $\langle u, v\rangle_0 = \int_\Omega uv f_{\mathbf{X}}\, dx$ and the induced norm $\|\cdot\|_0 = \sqrt{\langle u, v\rangle_0}$. Furthermore, to generalize the second condition in (4), it seems reasonable to look for nonlinear transformations ψ with some smoothness properties: in our context it is natural to assume each ψ to be differentiable in the distributional sense [see Adams (1975)] with derivatives in $\mathscr{L}^2_{f_{\mathbf{X}}}$. With this choice we can express the normalization constraint as $\mathbf{E}(\|\nabla\psi\|^2) = 1$. Under this condition, if $\overline{\psi}$ maximizes $\mathrm{Var}(\psi(\mathbf{X}))$, then every $\overline{\psi} + k$, $k \in \mathbf{R}$, does it; thus we limit our attention (as in the linear case) to centered functions ψ, that is, $\mathbf{E}(\psi(\mathbf{X})) = 0$. In conclusion, we ask that each ψ belongs to the Sobolev space of zero mean functions $\dot{H}^1_{f_{\mathbf{X}}} = \dot{H}^1_{f_{\mathbf{X}}}(\Omega) = \{\psi \in \mathscr{L}^2_{f_{\mathbf{X}}}(\Omega)\colon \mathbf{E}(\psi(\mathbf{X})) = 0, \nabla\psi \in \mathscr{L}^2_{f_{\mathbf{X}}}(\Omega)\}$. This is a Hilbert space [see Treves (1975)] with respect to the inner product $\langle u, v\rangle_1 = \langle \nabla u, \nabla v\rangle_0$ and induced norm $\|\cdot\|_1$. In this way we are able to express the normalization constraint in terms of the norm of the space where the maximum problem is given. Finally, note that the set $L_\Omega(n)$ of the restrictions to Ω of elements of $L(n)$ is included in $\dot{H}^1_{f_{\mathbf{X}}}$.

Now we have all the data to introduce our main definition.

DEFINITION 1. Let $\mathbf{X}$ be an absolutely continuous r.v. with positive bounded density on a domain $\Omega \subset \mathbf{R}^n$. The random variable $Y_j^* = \varphi_j(\mathbf{X})$ is said to be the jth NLPC of $\mathbf{X}$ if φ_j is a solution of the problem

$$
\begin{aligned}
&\text{maximize} && \mathbf{E}\big(\psi(\mathbf{X})^2\big) \\
&\text{subject to} && \mathbf{E}\big(\|\nabla\psi(\mathbf{X})\|^2\big) = 1, \qquad \psi \in \dot{H}^1_{f_{\mathbf{X}}}, \\
&&& \mathbf{E}\big(\psi(\mathbf{X}), \varphi_s(\mathbf{X})\big) = 0, \qquad s = 1, 2, \ldots, j-1,\ j > 1.
\end{aligned} \tag{5}
$$

It is easy to prove that if (3) holds, the spaces $\dot{H}^1_{f_{\mathbf{X}}}$ and $\mathscr{L}^2_{f_{\mathbf{X}}}$ coincide, respectively, with $\dot{H}^1$ and $\mathscr{L}^2$, obtained with the particular choice $f_{\mathbf{X}} = 1$. Thus, definitions and results stated for uniform densities hold in general for strictly positive and bounded densities: for this reason, in the sequel, for the sake of simplicity, we will refer only to a uniform density.

It is well known [see, e.g., Mikhailov (1978)] that $\varphi_j \in \dot{H}^1$ is a solution of the problem (5) if and only if there exists a constant λ_j^* such that

$$
\langle \varphi_j, h\rangle_0 - \lambda_j^* \langle \varphi_j, h\rangle_1 = 0 \qquad \forall h \in \dot{H}^1 \tag{6}
$$

with the natural condition of zero normal exterior derivative $\partial\varphi_j/\partial\eta$ on the boundary $\partial\Omega$. This means that φ_j is a weak solution [see, e.g., Treves (1975)] corresponding to λ_j^* of the Neumann problem

$$
\begin{aligned}
-\lambda^* \Delta\psi &= \psi, \qquad x \in \Omega, \\
\frac{\partial\psi}{\partial\eta} &= 0, \qquad x \in \partial\Omega,
\end{aligned} \tag{7}
$$

where $\Delta = \sum_{i=1}^{n} \partial^2/\partial x_i^2$ denotes the Laplace operator. Equivalently, φ_j is a solution of problem (5) if and only if it is a normalized eigenfunction of the differential operator $-\Delta$ corresponding to the jth characteristic value λ_j^*, where $\lambda_1^* \geq \lambda_2^* \geq \cdots \geq \lambda_j^*$.

From (6) we can conclude that, as in the linear case, the constant λ_j^* represents the variance of the jth NLPC. Furthermore, from (7) it is obvious that NLPCs for a uniformly distributed r.v. cannot be linear since $L_\Omega(n)$ belongs to the kernel of Δ: this, in particular, implies $\lambda_j^* > \lambda_j$ for $j = 1, 2, \ldots, n$.

Having introduced the notion of NLPCs, we now study the existence problem. To do this, it is crucial to distinguish whether Ω is bounded or not. In the first case, problem (7) is extensively treated in the PDEs literature and several results are available [see, e.g., Mikhailov (1978) and Treves (1975)]. The following theorem summarizes the most important of them.

THEOREM 2. *Let Ω be a bounded domain in $\mathbf{R}^n$ with continuous boundary $\partial\Omega$. Then $-\Delta$ admits countably many real, positive characteristic values $\lambda_1^* > \lambda_2^* \geq \lambda_3^* \geq \cdots$, corresponding to the Neumann condition; this set has the zero value as unique finite limit point. Every characteristic value corresponds to a finite number of eigenfunctions mutually orthogonal in $\dot{H}^1$.*

Theorem 2 tells us that under sufficiently general assumptions, each r.v. $\mathbf{X}$, uniformly distributed on a bounded domain $\Omega \subseteq \mathbf{R}^n$, admits countably many NLPCs with positive variance; the maximum variance λ_1^* is associated to the unique (modulo a change of sign) nonlinear component Y_1^*, whereas, for $j > 1$, it is possible to obtain a finite set $\{Y_j, Y_{j+1}, \ldots, Y_{j+k}\}$ of uncorrelated NLPCs which have the same variance. Similarly to the linear case, the sequence $\{\varphi_s\}_{s\in\mathbf{N}}$ of NLPC transformations represents a (orthonormal) basis for the space where the maximum variance problem is considered. Furthermore, $\dot{H}^1$ being dense in $\mathscr{L}^2$, the sequence $\{\lambda_s^{*-1}\varphi_s\}_{s\in\mathbf{N}}$ is an orthonormal basis for this space. Roughly speaking, the set of NLPCs represents the "largest" family of uncorrelated random variables that are $\mathscr{L}^2$ transformations of a (uniformly distributed) r.v. $\mathbf{X}$.

The proof of Theorem 2 depends on two results:

1. $\dot{H}^1$ is a Hilbert space;
2. the imbedding i_Ω: $\dot{H}^1 \to \mathscr{L}^2$, $i_\Omega(u) = u$, is compact (Rellich's theorem).

Using these facts, it is possible to prove the existence of a unique bounded linear (integral) operator G: $\mathscr{L}^2 \to \dot{H}^1$, the so-called *Green operator*, such that $\langle \psi, h\rangle_0 = \langle G\psi, h\rangle_1$ for all $h \in \dot{H}^1$. Its restriction to $\dot{H}^1$ is positive, self-adjoint, compact and thus (6) can be written as $G\psi = \lambda^*\psi$ and the conclusions of the previous theorem follow by standard results on spectral theory for compact operators on Hilbert spaces. The Green operator is the equivalent of the covariance operator $\mathbf{\Sigma}$ in the sense that if $L_\Omega(n)$ is considered as subspace of $\dot{H}^1$, the problem $\mathbf{\Sigma}T_j = \lambda_j T_j$ can be written as

$$\langle \mathbf{\Sigma}T_j, A\rangle_1 = \lambda_j \langle T_j, A\rangle_1 \qquad \forall A \in L_\Omega(n).$$

In any case, $\mathbf{\Sigma}$ is not the restriction of G to $L_\Omega(n)$. Indeed, if $u \in L_\Omega(n)$, then $v = Gu$ if and only if $-\Delta v = u$ and thus v cannot be linear.

The solutions of problem (5) are elements of $\dot{H}^1$, but under some regularity assumptions on $\partial\Omega$ several results on their regularity can be proved. We limit ourselves to recall that the eigenfunctions φ_j are always elements of $C^\omega(\Omega)$ (that is, they are analytic in Ω); furthermore the φ_j's are smooth on $\overline{\Omega}$, the closure of Ω, when $\partial\Omega$ is smooth; finally if $\partial\Omega$ is analytic, then $\varphi_j \in C^\omega(\overline{\Omega})$ [see, e.g., Egorov and Shubin (1992) or Nečas (1967)]. In this last case, since by construction $0 \in \Omega$, we can identify each φ_j with its Taylor series at the origin and this suggests that our approach can be considered as an infinite-dimensional generalization of the approach proposed in Gnanadesikan [(1977), page 53]. Note that if Ω is bounded, the analyticity of NLPC transformations implies the existence of all the moments of the r.v. $\mathbf{X}$. Later we will show that this result holds also for unbounded domains.

To conclude this section we may say that NLPCs can be considered a "natural" generalization of LPCs for several reasons: $L_\Omega(n)$ is a subset of $\dot{H}^1$ and if this set in (5) is substituted with $L_\Omega(n)$, we obtain problem (1); furthermore this conclusion also remains true if $\mathbf{X}$ is standardized. In the following section we will show that NLPCs have many properties typical of LPCs.

3. Some properties of NLPCs. It is well known that LPC analysis is invariant under orthogonal transformations [see, e.g., Mizuta (1984)] in the sense that if T and $\tilde{T}$ are two LPC operators corresponding to the r.v.'s $\mathbf{X}$ and $\widetilde{\mathbf{X}} = \Theta\mathbf{X}$, respectively, where Θ: $\mathbf{R}^n \to \mathbf{R}^n$ is any orthogonal transformation, then the equality $T\mathbf{X} = S\widetilde{T}\widetilde{\mathbf{X}}$ holds. NLPC analysis is orthogonally invariant, too. Indeed, if Θ is any such transformation and f is a sufficiently regular function, we have $(\Delta \circ f)\Theta = \Delta(f \circ \Theta)$ and the conclusion follows from (7).

It is also well known [see, e.g., Jolliffe (1986)] that a r.v. $\mathbf{X}$ and its covariance matrix $\mathbf{\Sigma}$ always can be decomposed with respect to the orthonormal basis of $\mathbf{R}^n$ consisting of the eigenvectors of $\mathbf{\Sigma}$:

$$\mathbf{X} = \sum_{j=1}^{n} Y_j T_j, \qquad \mathbf{\Sigma} = \sum_{j=1}^{n} \lambda_j T_j T_j'. \tag{8}$$

A trivial computation shows that (8) holds even if we consider $L_\Omega(n)$ as a subspace of $\dot{H}^1$ and put $\mathbf{X} = (\pi_1(\mathbf{X}), \pi_2(\mathbf{X}), \ldots, \pi_n(\mathbf{X}))'$, where $\pi_j \in L_\Omega(n)$ is the projection function $\pi_j(x) = x_j$, $j = 1, 2, \ldots, n$.

Following the same idea we are able to obtain the decomposition of $\mathbf{X}$ with respect to the orthonormal basis of $\dot{H}^1$ consisting of the eigenfunctions of $-\Delta$. Since

$$X_j = \sum_{s=1}^{+\infty} Y_s^* \langle \varphi_s, \pi_j \rangle_1 = \sum_{s=1}^{+\infty} Y_s^* \int_\Omega \frac{\partial \varphi_s}{\partial x_j}(x)\, dx,$$

we get

$$\mathbf{X} = \sum_{s=1}^{+\infty} Y_s^* \int_\Omega \nabla\varphi_s(x)\,dx.$$

The series defining X_j converges also in the $\mathscr{L}^2$ sense and the φ_s are orthogonal in $\mathscr{L}^2$. Hence we easily obtain the *spectral decomposition* of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \sum_{s=1}^{+\infty} \lambda_s^* \left(\int_\Omega \nabla\varphi_s(x)\,dx\right)\left(\int_\Omega \nabla\varphi_s(x)\,dx\right)', \tag{9}$$

where, as usual, λ_s^* are the variances of NLPCs.

In the linear case the operator T preserves the total variability of $\mathbf{X}$, that is, $\operatorname{tr}\mathbf{\Sigma} = \operatorname{tr}\mathbf{\Sigma}_\mathbf{Y}$, where $\mathbf{Y} = T\mathbf{X}$ and tr denotes the trace operator. This is a direct consequence of the fact that the scaling factor given by the absolute value of the Jacobian determinant $|\det JT| = |\det T|$ is unitary at every point $x \in \Omega$. In particular, this implies $m\Omega = mT\Omega$. In the nonlinear context the previous conclusions are no longer true.

THEOREM 3. *Let $\varphi^{(n)}\colon \Omega \to \mathbf{R}^n$ be the transformation whose components $\varphi_j^{(n)}$, $j = 1, 2, \ldots, n$, are the first n NLPC transformations of a uniform r.v. $\mathbf{X}$ on the bounded domain Ω with C^2 boundary. Then $\int_\Omega |\det J\varphi^{(n)}|\,dx \le 1$.*

PROOF. By the assumption on $\partial\Omega$, the functions $\varphi_j^{(n)} \in B(\dot{H}^1)$ are continuously differentiable. Since

$$\left|\det J\varphi^{(n)}\right| \le \prod_{j=1}^{n} \left\|\nabla\varphi_j^{(n)}\right\|$$

by the Jensen and the arithmetic–geometric means inequalities, we obtain

$$\begin{aligned}\left(\int_\Omega \left|\det J\varphi^{(n)}\right| dx\right)^{2/n} &\le \int_\Omega \left|\det J\varphi^{(n)}\right|^{2/n} dx \\ &\le \int_\Omega \sum_{j=1}^{n} \frac{\left\|\nabla\varphi_j^{(n)}\right\|^2}{n}\,dx = 1,\end{aligned}$$

which implies the thesis. □

It is necessary, mainly for applications, to introduce some criteria to decide how many NLPCs must be considered. We know that each nonlinear variance is strictly greater than the variance of LPCs. By the previous remark we also know that nonlinear variances take into account the covariability of all the moments of the r.v. $\mathbf{X}$. We are unable to give a complete answer to the problem. We limit ourselves to investigate the relations between linear and nonlinear variances.

Observe that spectral decomposition (9) permits us to obtain a relation between the total variability of $\mathbf{X}$ and the variances of NLPCs:

$$\operatorname{tr}\boldsymbol{\Sigma} = \sum_{s=1}^{+\infty} \lambda_s^* \sum_{i=1}^{n} \left(\int_\Omega \frac{\partial \varphi_s}{\partial x_i}(x)\, dx \right)^2. \tag{10}$$

Now we consider another relation between the linear and nonlinear variances.

THEOREM 4. *Let $\mathbf{X}$ be a r.v. uniformly distributed on the bounded domain Ω. Then for every $k \in \mathbf{N}\backslash\{0\}$ there exists a constant C_k such that*

$$\operatorname{tr}\boldsymbol{\Sigma} \geq C_k \frac{k}{\sum_{s=1}^{k} \lambda_s^{*-1}}.$$

PROOF. Consider the spaces $\mathscr{L}^2$ and $\dot{H}^1$ with their orthonormal bases $\{\varphi_s \lambda_s^{*-1}\}$ and $\{\varphi_s\}$, respectively. Since the projection functions π_j, $j = 1, 2, \ldots, n$, are elements of both spaces, we consider their Fourier series expansions

$$\pi_j = \begin{cases} \displaystyle\sum_{s=1}^{+\infty} \langle \pi_j, \varphi_s \lambda_s^{*-1/2} \rangle_0 \varphi_s \lambda_s^{*-1/2}, & \text{in } \mathscr{L}^2, \\ \displaystyle\sum_{s=1}^{+\infty} \langle \pi_j, \varphi_s \rangle_1 \varphi_s, & \text{in } \dot{H}^1. \end{cases}$$

The second series converges also in the $\mathscr{L}^2$-norm and we obtain the equality

$$\langle \pi_j, \varphi_s \rangle_0 \varphi_s \lambda_s^{*-1} = \langle \pi_j, \varphi_s \rangle_1.$$

Now consider the projection $\widetilde{\pi}_j$ of π_j on the k-dimensional, $k \geq 1$, subspace of $\dot{H}^1$ generated by the first k eigenfunctions of $-\Delta$. From Hölder's inequality it follows that

$$\begin{aligned} \|\widetilde{\pi}_j\|_1^2 &= \sum_{s=1}^{k} \left| \langle \pi_j, \varphi_s \rangle_1 \right|^2 = \sum_{s=1}^{k} \lambda_s^{*-1} \left| \langle \pi_j, \varphi_s \lambda_s^{*-1/2} \rangle_0 \right|^2 \\ &\leq \sum_{s=1}^{k} \lambda_s^{*-1} \|\pi_j\|_0^2 \|\varphi_s \lambda_s^{*-1/2}\|_0^2 \\ &= \|\pi_j\|_0^2 \sum_{s=1}^{k} \lambda_s^{*-1} \end{aligned}$$

and we finally have

$$\operatorname{tr}\boldsymbol{\Sigma} = \sum_{j=1}^{n} \|\pi_j\|_0^2 \geq \frac{k}{\sum_{s=1}^{k} \lambda_s^{*-1}} \frac{\sum_{j=1}^{n} \|\widetilde{\pi}_j\|_1^2}{k}. \qquad \square$$

Several characterizations of LPCs are given in terms of optimality properties [see, e.g., Darroch (1965), Okamoto and Kanazawa (1968) and Rao (1973)]. Some of them are immediately true also in the nonlinear case. Indeed, by construction, if k is any integer greater than or equal to 1 and $\psi^{(k)} = (\psi_1, \psi_2, \ldots, \psi_k)$ with $\psi_j \in \dot{H}^1$, the trace and the determinant of $\mathbf{\Sigma}_{\psi(\mathbf{X})}$ are maximized by taking $\psi^{(k)} = \varphi^{(k)}$. The possibility to introduce NLPCs by a loss function seems (to be) more difficult. We limit ourselves to recalling that if $R^{-1}[\psi] = \int_\Omega \psi^2 / \int_\Omega |\nabla\psi|^2$ denotes the reciprocal of the Rayleigh quotient [see Bandle (1980)], then the so-called Poincaré principle holds: if E_k is an arbitrary k-dimensional linear subspace of $\dot{H}^1$, then

$$\lambda_k^* = \max_{E_k} \min_{\psi \in E_k} R^{-1}[\psi].$$

This result is the infinite-dimensional generalization of the one that in the linear framework allows us to show the equivalence between the variational and loss-function approaches to LPC analysis.

To conclude our remarks about NLPCs we prove a simple result concerning the moments of a generic NLPC.

THEOREM 5. *Let $\Omega \subset \mathbf{R}^n$ be a bounded domain with smooth boundary. Then each NLPC Y_j^* of a r.v. $\mathbf{X}$ uniformly distributed on Ω admits moments of any order and*

$$\mathbf{E}\big(Y_j^{*n}\big) = (n-1)\mathbf{E}\big(Y_j^{*2}\big) \int_\Omega \varphi_j^{n-2} \|\nabla \varphi_j\|^2, \qquad j = 1, 2, \ldots, \ n = 2, \ldots.$$

PROOF. Since $\partial\Omega$ is smooth it follows from Remark 1 that $\varphi_j \in C^p(\overline{\Omega})$ for all p and so $Y_j^* = \varphi_j(\mathbf{X})$ admits moments of any order. Furthermore, applying the divergence theorem we obtain

$$\begin{aligned}\mathbf{E}\big(Y_j^{*n}\big) &= \int_\Omega \varphi_j^n(x)\, dx = -\lambda_j^* \int_\Omega \varphi_j^{n-1}(x) \Delta\varphi_j(x)\, dx \\ &= (n-1)\mathbf{E}\big(Y_j^{*2}\big) \int_\Omega \varphi_j^{n-2}(x) \|\nabla\varphi_j(x)\|^2\, dx. \qquad \square\end{aligned}$$

4. Two examples: uniform distributions on rectangles and annuli. We consider two particular examples which allow us to explore in detail the meaning and the properties of NLPCs, especially their relationships with LPCs.

4.1. *Uniform distributions on the rectangle.* Let $\mathbf{X}$ be a r.v. uniformly distributed on the open rectangle $\Omega = \prod_{i=1}^n (-a_i/2, a_i/2)$ with $a_1 \geq a_2 \geq \cdots \geq a_n > 0$. Since $m\Omega = \prod_{i=1}^n a_i$, $\mathbf{E}(\mathbf{X}) = \mathbf{0}$ and $\mathbf{\Sigma} = \operatorname{diag}\{a_1^2/12, a_2^2/12, \ldots, a_n^2/12\}$, the LPC transformation is the identity function.

The characteristic values of $-\Delta$ on Ω are [see Bandle (1980)]

$$\lambda_{m_1, \ldots, m_n}^* = \left(\sum_{i=1}^n \left(\frac{m_i \pi}{a_i} \right)^2 \right)^{-1}, \qquad m_i \in \mathbb{N}, \tag{11}$$

where $i = 1, 2, \ldots, n$ and $(m_1, m_2, \ldots, m_n)$ differs from the zero vector. The corresponding normalized eigenfunctions are

$$\varphi_{m_1,\ldots,m_n}(x) = \left[m\Omega \left(\prod_{i=1}^{n} \xi_i(m_i) \right)^{-1} \lambda^*_{m_1,\ldots,m_n} \right]^{1/2} \times \prod_{i=1}^{n} \cos\left(\frac{m_i \pi}{a_i} \left(x_i + \frac{a_i}{2} \right) \right), \tag{12}$$

where $\xi_i(m_i) = a_i/2$ if $m_i \neq 0$ and $\xi_i(0) = 1$. It is interesting to observe that if $a_1 > a_2$, the first NLPC Y_1^* of $\mathbf{X}$ is a random variable with the celebrated *arcsine distribution* on the open interval $(-\sqrt{2a_1 \cdot m\Omega}/\pi, \sqrt{2a_1 \cdot m\Omega}/\pi)$. Indeed the first characteristic value of $-\Delta$ is $\lambda_1^* = (a_1/\pi)^2$ with corresponding eigenfunction

$$\varphi_1(x) = g(x_1) = \frac{\sqrt{2a_1 m\Omega}}{\pi} \sin\left(\frac{\pi}{a_1} x_1 \right).$$

Eigenfunctions (12) are obtained with the so-called *method of separation of the variables*: they are products of the solutions γ_{ij} of the variational problems

$$\begin{aligned} &\text{maximize} \quad \mathbf{E}\big(\psi(X_i)^2\big) \\ &\text{subject to} \quad \mathbf{E}\big(|\psi'(X_i)|^2\big) = 1, \qquad \psi \in \dot{H}^1(\Omega_i), \\ &\qquad\qquad \mathbf{E}\big(\psi(X_i), \gamma_{is}(X_i)\big) = 0, \qquad s = 1, 2, \ldots, j-1,\ j > 1, \end{aligned}$$

where $\Omega_i = (-a_i/2, a_i/2)$, $i = 1, 2, \ldots, n$, and $\dot{H}^1(\Omega_i)$ is defined as in the multidimensional case. Putting $\gamma_i = \gamma_{i1}$, we call *marginal nonlinear principal components* (MNLPCs) the random variables

$$Z_i = \gamma_i(X_i) = \frac{\sqrt{2}a_i}{\pi} \sin\left(\frac{\pi}{a_i} X_i \right), \qquad i = 1, 2, \ldots, n.$$

The r.v. $\mathbf{Z} = \Upsilon(\mathbf{X})$, $\Upsilon = (\gamma_1, \ldots, \gamma_n)$, has independent components Z_i, but Υ does not necessarily coincide with the identity operator since each γ_i is sensitive to all the moments of X_i. In any case, the variances of the MNLPCs are proportional to the corresponding linear ones:

$$\frac{\mathbf{E}(X_j^2)}{\mathbf{E}(Z_j^2)} = \frac{\pi^2}{12}, \qquad j = 1, 2, \ldots, n.$$

It is interesting to note that the zero set of $\widetilde{\Upsilon}_i = (\gamma_1, \ldots, \gamma_{i-1}, \gamma_{i+1}, \ldots, \gamma_n)$ coincides with the ith LPC.

The following theorem collects some results about the relations between NLPCs, MNLPCs and LPCs. Let $r = \min\{k \in \mathbf{N}\backslash\{0\}\colon\ \operatorname{tr}\boldsymbol{\Sigma} \leq \operatorname{tr}\boldsymbol{\Sigma}_{\mathbf{Y}_k^*}\}$, where $\mathbf{Y}_k^* = \varphi^{(k)}(\mathbf{X})$. Obviously $1 \leq r \leq n$.

THEOREM 6. *Let* $\Omega = \prod_{i=1}^{n}(-a_i/2, a_i/2)$ *with* $a_1 \geq a_2 \geq \cdots \geq a_n > 0$. *The following conclusions hold*:

(i) $r = 1$ *if and only if* $a_1 \geq \sqrt{\sum_{i=2}^{n} a_i^2 \pi^2/(12 - \pi^2)}$;

(ii) $\operatorname{Var}(Z_j) \leq \operatorname{Var}(Y_j^*)$ *for all* $j = 1, 2, \ldots, n$, *and* $\operatorname{Var}(Z_j) = \operatorname{Var}(Y_j^*)$ *if and only if* $a_n > a_1/2$. *Under this assumption, the difference* $\operatorname{tr} \boldsymbol{\Sigma}_{\mathbf{Y}_r^*} - \operatorname{tr} \boldsymbol{\Sigma}$ *is minimized with* $m\Omega =$*constant if and only if* $r = n$ *and* $a_1 = a_2 = \cdots = a_n$;

(iii) *under the assumption of the previous point,* $r = n$ *if and only if* $a_n > \sqrt{\sum_{i=1}^{n-1} a_i^2(12 - \pi^2)/\pi^2}$. *In particular, if* $a_1 = a_2 = \cdots = a_n$, *then* $r = n$ *if and only if* $n \leq 5$.

PROOF. (i) Since $\lambda_1^* = (a_1/\pi)^2$ it is sufficient to note that the request $r = 1$ is equivalent to $(a_1/\pi)^2 \geq (1/12)\sum_{i=1}^{n} a_i^2$.

(ii) The first conclusion follows immediately from (11), the inequality on the a_j's and the double implication $a_n > a_1/2 \Leftrightarrow \lambda^*_{0,0,\ldots,n} > \lambda^*_{2,\ldots,0,0}$.

Now, consider the minimum problem

$$\begin{aligned} &\text{minimize} \quad \frac{1}{\pi^2}\sum_{i=1}^{r} a_i^2 - \frac{1}{12}\sum_{j=1}^{n} a_j^2, \qquad 1 \leq r \leq n, \\ &\text{subject to} \quad \prod_{i=1}^{n} a_i = V, \qquad V \in \mathbf{R}^+, a \in \mathbf{R}^n. \end{aligned}$$

Introducing the Lagrangian function and imposing the necessary condition for a minimum, we obtain the system

$$\begin{aligned} 2\left(\frac{1}{\pi^2} - \frac{1}{12}\right)a_s - \mu\frac{V}{a_s} = 0, \qquad & s = 1, 2, \ldots, r, \\ -\frac{1}{6}a_s - \mu\frac{V}{a_s} = 0, \qquad & s = r+1, \ldots, n, \\ \prod_{i=1}^{n} a_i = V. & \end{aligned} \tag{13}$$

Clearly problem (13) admits a unique solution if and only if $r = n$. In this case we obtain a convex objective function and the result follows by solving system (13).

(iii) Under the assumption $a_n > a_1/2$ the request $r = n$ is equivalent to

$$\frac{1}{\pi^2}\sum_{i=1}^{n-1} a_i^2 < \frac{1}{12}\sum_{i=1}^{n} a_i^2$$

and the thesis immediately follows. We omit the trivial proof of the last statement. □

Statement (ii) tells us that MNLPCs give the same information as the first n NLPCs if and only if the "weight" of each component X_i is not too relevant with respect to the others. Under another point of view, we can say that eigenvalues λ_j^* are sensitive to the shape of Ω. Indeed it is an easy task by (11) to verify

that the eigenfunctions corresponding to the first k eigenvalues depend only on x_1 if and only if $a_1 > ka_2$. In particular, in the plane, this means that if we consider a long, thin rectangular region (along the x_1 axis), the two-dimensional nonlinear eigenstructure of $\mathbf{X} = (X_1, X_2)$ is "similar" to the one of X_1. To remain in the bidimensional case, Sleeman and Zayed (1984) proved that for the so-called *trace function*

$$\Theta(t) = \operatorname{tr}\left(e^{-t\Delta}\right) = \sum_{j=1}^{+\infty} \exp(-\mu_j t),$$

where $\mu_j = \lambda_j^{*-1}$, the following asymptotic expansion on rectangular region $(0, a) \times (0, b)$ holds:

$$\Theta(t) \sim \frac{ab}{4\pi t} + \frac{2(a+b)}{4(4\pi t)^{1/2}} + \frac{1}{4} \quad \text{as } t \to 0. \tag{14}$$

Thus, we may say that the trace function contains information on the volume and the length of the perimeter of Ω.

A final remark: since a trivial computation shows that the zero set of $\cos[m_i\pi(x_i + a_i/2)/a_i]$ is given by

$$x_i = \frac{a_i}{m_i}\left(k - \frac{m_i + 1}{2}\right), \qquad m_i \neq 0,\ k \in \mathbf{Z},$$

we conclude by (11) that in the bidimensional case the zero set (known as *nodal lines*) of eigenfunctions corresponding to simple eigenvalues is a union of equally spaced lines, parallel to the coordinate axis, that is, the intersection of principal curves with the rectangle [see Duchamp and Stuetzle (1993), page 9]. In the other cases, zero sets can be very complicated [see Courant and Hilbert (1989)]: they probably contain principal curves.

4.2. *Uniform distributions on the annulus.* Consider now a uniform distribution on the annulus (in polar coordinates) $\Omega = \{(\rho, \theta)\colon 0 < R_1 < \rho < R_2,\ \theta \in (0, 2\pi]\}$. Obviously $\mathbf{E}(\mathbf{X}) = \mathbf{0}$ and $\mathbf{X}$ has uncorrelated but not independent components. By the separation of variables method it can be shown that the characteristic values $\lambda^*_{n,m}$ of $-\Delta$ are of the form

$$\lambda^*_{n,m} = \frac{R_1^2}{x^2},$$

where x is the nth root of the cross-product derivative Bessel equation

$$J'_m(x)N'_m(\delta x) - J'_m(\delta x)N'_m(x) = 0, \qquad m \in \mathbb{N}.$$

J_m and N_m are Bessel functions of order m of the first and second kind, respectively, and $\delta = R_2/R_1$ [see, e.g., Arfken (1966)]. The corresponding eigenfunctions are

$$\begin{aligned}\varphi_{n,m}(\rho, \theta) = {} & J_m\left(\sqrt{\mu^*_{n,m}}\rho\right)\left(\alpha_{n,m}\cos m\theta + \beta_{n,m}\sin m\theta\right) \\ & + N_m\left(\sqrt{\mu^*_{n,m}}\rho\right)\left(\widetilde{\alpha}_{n,m}\cos m\theta + \widetilde{\beta}_{n,m}\sin m\theta\right)\end{aligned}$$

with $\mu^*_{n,m} = \lambda^{*-1}_{n,m}$. Since the constants $\alpha_{n,m}$, $\beta_{n,m}$, $\widetilde{\alpha}_{n,m}$ and $\widetilde{\beta}_{n,m}$ are arbitrary when $m \neq 0$, all the eigenvalues are at least double since, if $\widetilde{\alpha}_{n,m} = \widetilde{\beta}_{n,m} = 0$, the associated eigenfunctions $J_m\big(\sqrt{\lambda^*_{n,m}}\rho\big)\cos m\theta$ and $J_m\big(\sqrt{\lambda^*_{n,m}}\rho\big)\sin m\theta$ are linearly independent.

For the annulus the definition of MNLPC is not immediate: the marginal distributions are not uniform and so definitions and results previously obtained are not available. The notion of weighted Sobolev space is necessary, but we do not develop this subject here.

Sleeman and Zayed (1984) showed that

$$\Theta(t) \sim \frac{\pi(R_2^2 - R_1^2)}{4\pi t} + \frac{2\pi(R_1 + R_2)}{4(4\pi t)^{1/2}} + O\big(t^{1/2}\big) \quad \text{as } t \to 0,$$

that is, the spectral decomposition of $\mathbf{X}$ into NLPCs determines the area of Ω and the length of its boundary $\partial\Omega$ as happens for rectangular regions. Furthermore, the constant term in (14) is now zero because Ω presents one hole.

5. Unbounded domains. In Section 2 we showed that a r.v. uniformly distributed on a bounded domain Ω with sufficiently regular boundary $\partial\Omega$ always admits NLPCs. It is evident that such a r.v. has finite moments of every order and its moment generating function (m.g.f.) $\psi_{\mathbf{X}}(t) = \mathbf{E}(e^{\langle t, \mathbf{X}\rangle})$ is finite for any $t \in \mathbf{R}^n$.

Now, if $\mathbf{X}$ is distributed on an unbounded domain Ω, the existence of its moments is not guaranteed. The main result of this section is that if $\partial\Omega$ is sufficiently regular, the existence of NLPCs can be considered equivalent to the global existence of $\psi_{\mathbf{X}}$. This result appears (to be) reasonable since NLPCs are "objects" sensitive to the covariability between the moments of every order of $\mathbf{X}$.

So, let $\mathbf{X}$ be a uniformly distributed r.v. on the unbounded domain Ω in $\mathbf{R}^n$ with $m\Omega = 1$. We suppose that $\mathbf{X}$ has expectation and finite second-order moments. To apply the previous variational approach to our eigenvalue problem we need to prove the compactness of the imbedding $i_\Omega\colon \dot{H}^1 \to \mathscr{L}^2$. This result is more difficult to obtain with respect to the bounded case since both the regularity of $\partial\Omega$ and its behavior at infinity are relevant. We start by recalling a simple necessary condition of compactness.

THEOREM 7 (Adams and Fournier (1971a). *Let Ω be a domain in $\mathbf{R}^n$. If i_Ω is compact the following conclusions hold*:

(i) $m\Omega < +\infty$;
(ii) $\lim_{r\to+\infty} e^{kr} m\Omega_r = 0\ \forall k$, *where* $\Omega_r = \{x \in \Omega\colon \|x\| > r\}$.

The first statement is obviously equivalent to the possibility of defining a r.v. uniformly distributed on Ω. The second statement allows us to prove the following theorem.

THEOREM 8. *If i_Ω is compact, then $\mathbf{X}$ admits moments of every order.*

PROOF. From Theorem 7 the compactness assumption implies conclusion (ii). This implies the existence of an index $\overline{n}$ such that for every $n > \overline{n}$ and for any positive integer k we have $m\,\Omega_n < e^{-nk}$. Furthermore,

$$\begin{aligned}\int_{\{n<\|x\|<n+1\}\cap\Omega} \|x\|^k\,dx &\le (n+1)^k m\{n < \|x\| < n+1\}\cap\Omega \\ &\le (n+1)^k m\Omega_n \\ &\le (n+1)^k e^{-nk}.\end{aligned}$$

So the series $\sum_{n=0}^{+\infty}\int_{\{n<\|x\|<n+1\}\cap\Omega}\|x\|^k\,dx$ converges and this proves the thesis. □

The statement of the last theorem cannot be reversed: later we will show that if a r.v. $\mathbf{X}$ is uniformly distributed on $\Omega = \{(x, y) \in \mathbf{R}^2\colon x > 0,\ 0 < y < e^{-x}\}$, then it admits moments of every order but the imbedding i_Ω is not compact.

The following theorem tells us that even if we consider more restrictive assumptions on the moments of $\mathbf{X}$, the compactness of i_Ω is not guaranteed.

THEOREM 9. *Condition* (ii) *in Theorem* 7 *is equivalent to the existence on* $\mathbf{R}^n$ *of the m.g.f. of* $\mathbf{X}$.

PROOF. The existence on $\mathbf{R}^n$ of $\psi_{\mathbf{X}}$ is equivalent to

$$\lim_{r\to+\infty}\int_{\Omega_r} e^{\langle t,\,x\rangle}\,dx = 0 \qquad \forall t \in \mathbf{R}^n \tag{15}$$

and so we prove the theorem with respect to this condition. Since

$$e^{|\langle t,x\rangle|} \le e^{\langle t,x\rangle} + e^{-\langle t,x\rangle},$$

if (15) holds, then $\lim_{r\to+\infty}\int_{\Omega_r} e^{|\langle t,x\rangle|}\,dx = 0\ \forall t$. So if we put $t = (\alpha, 0, \ldots, 0)$, we obtain

$$\int_{\Omega_r} e^{|\langle t,\,x\rangle|}\,dx = \int_{\Omega_r}\exp(\|\alpha\|\,\|x_1\|)\,dx \ge e^{\|\alpha\|r} m\Omega_r$$

and the thesis follows immediately from the choice $|\alpha| = k$.

Conversely, let $t \in \mathbf{R}^n$ and choose $k > 0$ such that $|t| < k$. Then if $n \in \mathbf{N}$, we have

$$\begin{aligned}\int_{\Omega\cap\{n<\|x\|<n+1\}} e^{\|t\|\,\|x\|}\,dx &< e^{\|t\|(n+1)} m\{n < \|x\| < n+1\}\cap\Omega \\ &\le e^{\|t\|(n+1)} m\Omega_n \\ &\le C e^{\|t\|(n+1)} e^{-kn} \\ &= C e^{-n(k-\|t\|)+\|t\|},\end{aligned}$$

which implies the convergence of $\int_\Omega e^{\|t\|\,\|x\|}\,dx$. □

The result stated in Theorem 9 tells us that it is not sufficient that the tail of the distribution of $\mathbf{X}$ converges quickly to zero to obtain the compactness of i_Ω: as in the bounded case, a sufficient regularity of $\partial\Omega$ is also crucial.

Necessary and sufficient or just sufficient conditions for the compactness of i_Ω are known [see, e.g., Adams and Fournier (1971a) or Berger (1991)], but they are difficult to apply. In any case, particular domains allow us to determine the compactness of i_Ω directly. In Adams and Fournier (1971b) it is shown that if $\Omega = \{(x, y) \in \mathbf{R}^2 \colon x > 0, 0 < y < f(x)\}$, where f is a positive, decreasing, continuously differentiable function on $[0, +\infty)$ with bounded derivative f' (we call Ω Adams domain), then i_Ω is compact if and only if

$$\lim_{x \to +\infty} \frac{f(x+\varepsilon)}{f(x)} = 0 \qquad \forall \varepsilon > 0. \tag{16}$$

Now we show that condition (16) is equivalent to the global existence of the m.g.f. of $\mathbf{X}$.

THEOREM 10. *Let Ω be an Adams domain. Then condition* (16) *is equivalent to the existence for all $(t, \tau) \in \mathbf{R}^2$ of the m.g.f. $\psi_{\mathbf{X}}$ of a r.v. $\mathbf{X}$ uniformly distributed on Ω.*

PROOF. First we observe that

$$\psi_{\mathbf{X}}(t, \tau) = \begin{cases} \displaystyle\int_0^{+\infty} f(x) e^{tx}\, dx, & \text{if } \tau = 0, \\ \displaystyle\tau^{-1} \int_0^{+\infty} [e^{\tau f(x)} - 1] e^{tx}\, dx, & \text{if } \tau \neq 0. \end{cases}$$

We conclude that the existence for all $(t, \tau) \in \mathbf{R}^2$ of ψ_X is equivalent to the convergence for all $t \in \mathbf{R}$ of the integral $\int_0^{+\infty} f(x) e^{tx}\, dx$. Note that the only relevant case is $t > 0$.

Now suppose that (16) is true and consider the series with positive terms

$$\sum_{n=0}^{+\infty} a_n = \sum_{n=0}^{+\infty} \int_n^{n+1} f(x) e^{tx}\, dx. \tag{17}$$

Having fixed $\varepsilon \in (0, 1)$, we obtain

$$\begin{aligned} \frac{a_{n+1}}{a_n} = \frac{\displaystyle\int_{n+1}^{n+2} f(x) e^{tx}\, dx}{\displaystyle\int_n^{n+1} f(x) e^{tx}\, dx} &\leq \frac{\displaystyle\int_{n+1}^{n+2} f(x) e^{tx}\, dx}{\displaystyle\int_n^{n+1-\varepsilon} f(x) e^{tx}\, dx} \\ &\leq \frac{e^{t(n+2)} f(n+1)}{(1-\varepsilon) e^{tn} f(n+1-\varepsilon)} = \frac{e^{2t}}{1-\varepsilon} \frac{f[(n+1-\varepsilon)+\varepsilon]}{f(n+1-\varepsilon)}. \end{aligned}$$

So, for all t, $a_{n+1}/a_n \to 0$ as $n \to +\infty$ and $\int_0^{+\infty} f(x) e^{tx}\, dx$ converges.

Conversely, let (17) converge when $t > 0$. Observe that since f is decreasing, $\lim_{x \to +\infty} f(x+\varepsilon)/f(x) \leq 1$. By contradiction, suppose there exists an $\overline{\varepsilon} > 0$

such that (16) is not true. Then there exists a $\overline{\delta} > 0$ such that for all real numbers $M > 0$ we can find $x(M) > M$:

$$\frac{f(x(M)+\overline{\varepsilon})}{f(x(M))} \geq \overline{\delta} > 0. \tag{18}$$

Since f is monotone, if (18) holds for $\overline{\varepsilon} > 0$ then it holds for every $\varepsilon < \overline{\varepsilon}$. Consider now a sequence $\{x_n\}_{n\geq 0}$ whose elements are (note that a such sequence is not uniquely determined)

$$x_0 = x(1), \qquad x_n = x(x_{n-1} + \overline{\varepsilon}) \qquad \forall n \geq 1.$$

The sequence $\{x_n\}_{n\geq 0}$ is increasing, it diverges to $+\infty$ and $x_{n+1} - x_n > \overline{\varepsilon}$ for all n. Now we introduce the sequence $\{y_n\}_{n\geq 0}$ defined as

$$y_{3n} = x_n, \qquad y_{3n+1} = x_n + \frac{\overline{\varepsilon}}{2} \qquad y_{3n+2} = x_n + \overline{\varepsilon}, \qquad n \in \mathbf{N},$$

and study the convergence of the series

$$\sum_{n=0}^{+\infty} b_n = \sum_{n=0}^{+\infty} \int_{y_n}^{y_{n+1}} f(x)e^{tx}\,dx.$$

For $\gamma \in (0, \overline{\varepsilon}/2)$ we obtain

$$\begin{aligned}\frac{b_{3n+1}}{b_{3n}} &= \frac{\int_{x_n+\overline{\varepsilon}/2}^{x_n+\overline{\varepsilon}} f(x)e^{tx}\,dx}{\int_{x_n}^{x_n+\overline{\varepsilon}/2} f(x)e^{tx}\,dx} \geq \frac{\int_{x_n+\overline{\varepsilon}/2+\gamma}^{x_n+\overline{\varepsilon}} f(x)e^{tx}\,dx}{\int_{x_n}^{x_n+\overline{\varepsilon}/2} f(x)e^{tx}\,dx} \\ &\geq \frac{\overline{\varepsilon}/2-\gamma}{\overline{\varepsilon}/2} e^{t\gamma} \frac{f(x_n+\overline{\varepsilon})}{f(x_n)} \geq \frac{\overline{\varepsilon}/2-\gamma}{\overline{\varepsilon}/2} e^{t\gamma}\overline{\delta}\end{aligned}$$

for n sufficiently large. From the arbitrariness of t it follows that the existence of $\overline{t} > 0$ is such that

$$\liminf_{n\to+\infty} \frac{b_{n+1}}{b_n} > 1$$

and this implies that the integral $\int_0^{+\infty} f(x)e^{tx}\,dx$ cannot globally converge. □

A more general class of domains in $\mathbf{R}^n$ for which i_Ω is compact was introduced in Berger (1991). Let

$$\Omega = \left\{ x \in \mathbf{R}^n \colon 0 < x_n < +\infty, \left(\sum_{i=1}^{n-1} x_i^2\right)^{1/2} < g(x_n) \right\},$$

where $g(s) = \exp\{-\phi(s)\}$ is continuous on $[0, +\infty)$. Furthermore suppose there exists $s_0 > 0$ such that:

1. $\phi \in C^2([s_0, +\infty))$;
2. $\phi' \to +\infty$ as $s \to +\infty$;
3. ϕ' is strictly monotone if $s \geq s_0$.

Since for such domains i_Ω is compact, from Theorem 7 it follows that $\mathbf{X}$ admits m.g.f. on $\mathbf{R}^n$. We are able to give a direct proof of this fact.

THEOREM 11. *A r.v.* $\mathbf{X}$ *uniformly distributed on a "Berger domain"* Ω *admits m.g.f. on* $\mathbf{R}^n$.

PROOF. Put $x' = (x_1, x_2, \dots, x_{n-1})$ and $t' = (t_1, t_2, \dots, t_{n-1})$. If $t \neq 0$, we have

$$\begin{aligned}\int_\Omega e^{\langle t,x\rangle}\,dx &= \int_0^{+\infty} \exp(t_n x_n)\,dx_n \int_{\|x'\|<g(x_n)} \exp(\langle t', x'\rangle)\,dx' \\ &\le \int_0^{+\infty} \exp(t_n x_n)\,dx_n \int_{\|x'\|<g(x_n)} \exp(\|t'\|\,\|x'\|)\,dx' \\ &= \omega_n n \int_0^{+\infty} \exp(t_n x_n) \\ &\qquad \times \left[\frac{(-1)^{n-1}(n-1)!}{\|t'\|^n}\big(\exp(\|t'\|g(x_n)) - 1\big) + R(x_n)\right] dx_n,\end{aligned}$$

where ω_n denotes the measure of the unit sphere in $\mathbf{R}^n$ and

$$R(x_n) = \exp\big(\|t'\|g(x_n)\big) \sum_{j=1}^{n-1} \frac{(-1)^{j-1}(n-1)!}{\|t'\|^j (n-j)!}\big(g(x_n)\big)^{n-j}.$$

By the assumptions on g, the term $R(x_n)$ is negligible and thus the convergence of the last integral is equivalent to the convergence of

$$\int_0^{+\infty} \exp(t_n x_n - \phi(x_n))\,dx_n. \tag{19}$$

From the assumptions on ϕ we can write

$$\phi(s) = \phi(s_0) + \int_{s_n}^{s} \phi'(u)\,du.$$

Furthermore, by assumption 2, for any t_n there exists $\bar{s} > s_0$ such that for all $s > \bar{s}$ we have $\phi'(s) > 2|t_n|$. So, for every fixed t_n we obtain

$$\phi(s) = \phi(s_0) + \int_{s_0}^{\bar{s}} \phi'(u)\,du + \int_{\bar{s}}^{s} \phi'(u)\,du > C + 2(s - \bar{s})|t_n|.$$

We can conclude that

$$\begin{aligned}\exp\big\{t_n x_n - \phi(x_n)\big\} &< \exp\big\{-C - 2(x_n - \bar{x}_n)|t_n| + t_n x_n\big\} \\ &= C_1 \exp\{-2x_n|t_n| + t_n x_n\} \\ &< C_1 \exp\{-x_n|t_n|\}\end{aligned}$$

and so the integral (19) converges. □

6. Discussion: extensions and open problems. As seen above, the notion of NLPC generalizes the classical linear notion and maintains several of their characteristics such as invariance under orthogonal transformations, orthogonality or dependence on all the components of the initial r.v. The results of the previous sections have shown that each NLPC is sensitive to all the moments of every component of $\mathbf{X}$ and on its marginal of every order. In particular, as proved in Section 5, this requires $\mathbf{X}$ to have higher regularity than the finiteness of its second moments.

Our approach to nonlinear principal components relates them to the spectral theory of differential operators. This relation between the statistical problem and its mathematical formulation seems to be promising. For example, in Section 4, we have succeeded in exhibiting all the NLPCs for rectangular and annular domains; moreover, several other domains allow a direct calculation. As far as we know this does not generally happen in other approaches to nonlinear principal components. Another advantage is that we obtain, when possible, a way to compute an orthonormal basis of $\mathscr{L}^2(\Omega)$, whereas the other approaches for the calculation of the corresponding nonlinear principal components require us to know a priori the "right" basis to compute their transformations [see Donnell, Buja and Steutzle (1994) and De Leeuw, Van Rijekevorsel and Van der Wouden (1981)]. Roughly speaking, knowledge of the differential operator whose spectrum is studied yields the basis.

Even from a theoretical point of view, a lot of results and techniques like the finite elements method for the computation of the eigenfunctions and the corresponding eigenvalues are available. We have not developed here a sample theory of NLPCs, but it seems to be straightforward; indeed, it is not a posteriori surprising that our problem for a uniform distribution is associated with the spectral analysis of the Laplacian operator: remember that uniform density is an eigenfunction corresponding to the zero eigenvalue of $-\Delta$. This remark might allow us to obtain a "discrete theory" starting from the discrete version of the Laplacian.

The assumption on the distribution of $\mathbf{X}$ is obviously too restrictive and for this reason this paper must be viewed as introductory. To explore more general situations, it is necessary to refer to the more sophisticated theory of weighted Sobolev spaces. Thus it also should be possible to introduce the definition of MNLPCs sketched in Section 4 and to prove some extensions of Theorem 6 such as the reasonable conjecture that in the normal case the MNLPC transformation is the identity operator. Furthermore, we hope to clarify the relations between NLPCs or MNLPCs and some of the nonlinear generalizations of the LPCs existing in the literature, perhaps obtaining a unifying approach to generalized principal components. Our research is continuing and further results will be announced.

In any case, even in the simple case considered here several questions are open. As an example, it would be of interest to study the stochastic processes associated by NLPCs to the r.v. $\mathbf{X}$ or to investigate of the unbounded domains class for which the existence of NLPCs is guaranteed. In particular, it should be relevant to construct an example of a r.v. $\mathbf{X}$ with finite second moments for

which the spectrum of the associated Green operator is not discrete: if such an example exists, the presence of a continuous spectrum needs to be clarified. Moreover, our analysis performed on rectangles and annuli can be extended to other domains such as the equilateral triangle, where the components of $\mathbf{X}$ are correlated.

Finally, we observe also that from a mathematical point of view, the study of NLPCs seems to be relevant: on one hand it can justify more research in some classical areas as, for example, the study of the Laplacian spectrum; on the other hand, we cannot exclude that the statistical–probabilistic approach could produce some technique to solve some mathematical problems. For example, conditions that assure the entirety of the characteristic function of a r.v. might be used to study the compactness of the imbedding i_Ω.

Acknowledgments. The author wishes to thank Sandro Salsa for his many helpful comments and suggestions. Also, thanks are due Donato Michele Cifarelli, Marco Scarsini, an Associate Editor and two anonymous referees for their comments. In any case, the responsibility for any mistake must be ascribed completely to the author.

REFERENCES

ADAMS, R. A. (1975). *Sobolev Spaces.* Academic Press, New York.

ADAMS, R. A. and FOURNIER, J. (1971a). Some imbedding theorem for Sobolev spaces. *Canada J. Math.* **23** 517–530.

ADAMS, R. A. and FOURNIER, J. (1971b). A compact imbedding theorem for functions without compact support. *Canad. Math. Bull.* **14** 305–309.

ARFKEN, G. (1966). *Mathematical Methods for Physicists*. Academic Press, New York.

BANDLE, C. (1980). *Isoperimetric Inequalities and Applications*. Pitman, Boston.

BERGER, G. (1991). Different eigenvalue asymptotics for Dirichlet and Neumann boundary conditions of elliptic operators of second order in some irregular domains. *Math. Nachr.* **152** 159–171.

COURANT, R. and HILBERT, D. (1989). *Methods of Mathematical Physics* **1**. Wiley, New York.

DARROCH, J. N. (1965). An optimal property of principal components. *Ann. Math. Statist.* **36** 1579–1582.

DE LEEUW, J. and VAN RIJEKEVORSEL, H. and VAN DER WOUDEN, H. (1981). Nonlinear principal components analysis with B-splines. *Methods Oper. Res.* **33**, 379–393.

DONNELL, D. J., BUJA, A. and STUETZLE, W. (1994). Analysis of additive dependencies and concurtivities using smallest additive principal components, *Ann. Statist.* **22** 1635–1668.

DUCHAMP, T. and STUETZLE, W. (1993). The geometry of principal curves in the plane. Technical Report 250, Dept. Statistics, Univ. Washington (revised in August 1995).

DUCHAMP, T. and STUETZLE, W. (1995a). Extremal properties of principal curves in the plane. Preprint.

DUCHAMP, T. and STUETZLE, W. (1995b). Geometric properties of principal curves in the plane. *Robust Statistics, Data Analysis, and Computer Intensive Methods* 135–152. Springer, Berlin.

EGOROV, YU. V. and SHUBIN, M. A., eds. (1992). *Partial differential equations I. Encyclopedia of Mathematical Sciences* **30**. Springer, Berlin.

FLURY, B. D. (1994). Discussion of "Analysis of additive dependencies and concurtivities using smallest additive principal components," by D. J. Donnell, A. Buja and W. Stuetzle, *Ann. Statist.* **22** 1668–1670.

GIFI, A. (1990). *Nonlinear Multivariate Analysis*. Wiley, New York.

GOTTLIEB, H. P. W. (1985). Eigenvalues of the Laplacian with Neumann boundary conditions. *J. Austral. Math. Soc. Ser. B* **26** 293–309.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York.

HASTIE, T. and STUETZLE, W. (1989). Principal curves, *J. Amer. Statist. Assoc.*, **84** 502–516.

HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **26** 417–441, 498–520.

KOYAK, R. A. (1987). On measuring internal dependence in a set of random variables. *Ann. Statist.* **15** 1215–1228.

JOLLIFFE, I. T. (1986). *Principal Component Analysis*. Springer, Berlin.

MIKHAILOV, V. P. (1978). *Partial Differential Equations*. MIR, Moscow.

MIZUTA, M. (1984). Generalized principal components analysis invariant under rotations of a coordinate system. *J. Japan Statist. Soc.* **14** 1–9.

NEČAS, J. (1967). *Les méthodes directes en Théorie des Équations Elliptiques*. Masson, Paris.

OKAMOTO, M. and KANAZAWA, M. (1968). Minimization of eigenvalues of a matrix and optimality of principal components. *Ann. Math. Statist.* **39** 850–863.

PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2** 559–572.

RAO, C. R. (1973). Linear statistical inference and its applications. Wiley, New York.

SLEEMAN, B. D. and ZAYED, E. M. E. (1984). Trace formulae for the eigenvalues of the Laplacian. *Z. Angew. Math. Phys.* **35** 106–115.

TREVES, F. (1975). *Basic Linear Partial Differential Equations*. Academic Press, New York.

DIPARTIMENTO DI STATISTICA
E MATEMATICA "DIEGO DE CASTRO"
UNIVERSITÀ DI TORINO
PIAZZA ARBARELLO, 8
10122-TORINO
ITALIA
E-MAIL: salinelli@econ.unito.it