Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2010, Vol. 46, No. 3, 653–676
DOI: 10.1214/09-AIHP326

ANNALES DE L'INSTITUT HENRI POINCARÉ PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Optimal nonlinear transformations of random variables

Aldo Goia and Ernesto Salinelli

Dipartimento di Scienze Economiche e Metodi Quantitativi, Università del Piemonte Orientale "A. Avogadro," Via Perrone 18, 28100 Novara, Italy. E-mail: aldo.goia@eco.unipmn.it; ernesto.salinelli@eco.unipmn.it

Received 16 December 2008; revised 4 June 2009; accepted 22 June 2009

Abstract. In this paper we deepen the study of the nonlinear principal components introduced by Salinelli in 1998, referring to a real random variable. New insights on their probabilistic and statistical meaning are given with some properties. An estimation procedure based on spline functions, adapting to a statistical framework the classical Rayleigh–Ritz method, is introduced. Asymptotic properties of the estimator are proved, providing an upper bound for the rate of convergence under suitable mild conditions. Some applications to the goodness-of-fit test and the construction of bivariate distributions are proposed.

Résumé. Dans cet article nous étudions les composantes principales non linéaires définies par Salinelli en 1998, dans le cas d'une variable aléatoire réelle. La signification probabiliste et statistique est approfondie et des proprietés sont illustrées. Une procédure d'estimation basée sur les fonctions splines, qui adapte la méthode classique de Rayleigh–Ritz, est présentée. Des propriétés asymptotiques de cet estimateur sont établies, et on donne une borne pour la vitesse de convergence sous des conditions générales. Des applications aux tests d'ajustement et à la construction de distributions bivariées sont proposées.

MSC: Primary 60E05; secondary 49J05; 47A75; 62G05; 62G10

Keywords: Covariance operator; Chernoff–Poincaré inequality; Nonlinear principal components; Splines estimates; Sturm–Liouville problems

1. Introduction

Linear Principal Components Analysis (LPCA) of a real random vector is a well-known multivariate statistical technique mainly appreciated as a dimensional reduction tool in data analysis. Many different but equivalent definitions of LPCs are known (see [23]) concerning in any case the goal of finding a linear change of coordinates such that the transformed random vector has uncorrelated components or, in other words, its covariance operator is in diagonal form. The more general definition of LPCs for functional variables (see [15]), based on the Karhunen–Loève decomposition, follows the same idea.

Starting from the 70's, a rich literature which extends LPCA to the more general context of nonlinear transformations has been developed, motivated by the aim of detecting nonlinear structures in random vectors. Several different but not equivalent definitions of "nonlinear" principal components have been proposed (for a short survey see the introduction in [31]) each one based on the generalization of some property of LPCs. Many of these definitions give countable many nonlinear PCs for a random vector, excluding a priori the possibility of using them in the reduction of dimensionality and then raising the need to deepen their statistical meaning and applicability.

To this group of definitions belongs the one proposed in [30]: nonlinear principal components (in the sequel NLPCs) of a random vector with zero mean and positive definite covariance matrix were introduced as a solution of a variance maximization problem over the weighted Sobolev space of real valued nonlinear transformations which are centered, square integrable and differentiable in the weak sense, with square integrable first derivatives. Some results on the existence of NLPC transformations and their properties were proved for densities bounded away from zero and infinity, whereas in [31] the NLPCs of a Gaussian random vector were considered, obtaining a characterization result and giving new insight on the so-called “horse-shoe” effect. However, in these works, the probabilistic and statistical meaning of NLPCs and the problem of their estimation and applications were not investigated.

The aim of this paper is to face these questions in the one-dimensional case, that is for random variables (r.v.s). This choice is justified by several different reasons: first, given the depart of NLPCs from the classical idea of LPC as a reduction of dimensionality, in order to explore the statistical meaning of NLPCs it is not too restrictive to consider the one-dimensional case where PCA make no sense, gaining however in simplicity and mathematical tractability; second, as shown later on, it is possible to prove some results on the transformations of random variables which do not appear easily extensible to a multivariate context; furthermore, we feel that the analysis of the one-dimensional case could give some insight on the possibility to expand the idea, only sketched in [30], of introducing marginal nonlinear principal components of a random vector.

Our work is organized in three parts. In the first one (Sections 2 and 3), specializing the definition given in [30], we introduce optimal transformations of a random variable as solution of a variance maximization problem with a normalization differential constraint, calling them, by analogy with the multivariate case, nonlinear principal components. We show how this maximization problem on a functional space can be solved with the so-called direct method of the calculus of variations on which, in the second part, we base the estimation procedure proposed.

The probabilistic-statistical meaning of NLPCs is clarified under two different, but related, points of view. From one part we show that the first (i.e. with maximum variance) NLPC transformation is equal to the optimal transformation (i.e. the one that realize the equality) in the Chernoff–Poincaré inequality (see e.g. [7,12,13]): thus, all the NLPC transformations can be considered optimal transformations for suitable restricted Chernoff–Poincaré inequalities. From another point of view, we show how finding NLPCs is equivalent to determining a complete basis of the weighted Sobolev space on which is posed the maximization variance problem which diagonalizes the extension to this space of the covariance operator associated to the random variable considered: hence NLPCs transformations and their variances are the eigenfunctions and the eigenvalues respectively of this extended covariance operator.

Besides, we have obtained further improvements with respect to what obtained in [30] and [31]: the assumptions on the density of the initial r.v. are weakened, new examples (with a nonexistence case) are presented, we prove a symmetry result for the NLPCs transformations and the monotonicity of the first one, a result that do not extend to the multivariate case. Using this last result, we are able to prove a weak version of a characterization result known in the Chernoff–Poincaré literature.

The second part (Sections 4, 5 and 7) of our work is devoted to introducing and analyzing an estimation procedure of NLPCs based on splines, adapting in a statistical perspective the classical Rayleigh–Ritz method. After defining the estimator, we derive its explicit expression showing how the problem of estimating the NLPC transformations reduces to a generalized eigenvalue one. Then we give some asymptotic results. In particular, we prove that the empirical estimates of NLPCs exist and they are unique except on an event whose probability approaches zero as the sample size tends to infinity. Moreover, under mild regularity conditions on the NLPCs transformations, by choosing the degree of the splines and the number of knots properly, we obtain an upper bound for the rate of convergence in probability. A simulation study completes the analysis, showing the performances of the proposed estimator.

In the third part (Section 6) we suggest how to use in practice NLPCs, discussing two possible applications: the aim is to show how some properties of NLPCs turn out interesting in statistical frameworks involving distributional aspects. More precisely, we apply the characterization property in defining a goodness-of-fit test. Here we illustrate this idea in testing uniformity: we carry out a simulation study from which the good performances of this procedure emerge. Moreover, since NLPCs transformations represent a suitable basis of the functional space associated to the density of a r.v., they may help in the study of the dependence structure between two r.v.s. In this perspective we dwell on the study of the properties of a family of bivariate dependent variables, constructed using NLPCs.

2. Notation, definitions and existence results

We start by recalling briefly the definition of nonlinear principal component for a real, absolutely continuous (a.c.) random vector $\mathbf{X}$ with density $f_{\mathbf{X}}$ having support a domain (an open and connected set) $D \subseteq \mathbb{R}^p$ given in [30]. The starting point was to note that the *normalization constraint* $\|\mathbf{a}_j\|^2_{\mathbb{R}^p} = 1$ on the vector transformations to find the jth LPC $Y_j = \mathbf{a}_j^{\mathrm{T}}\mathbf{X}$, $j = 1, 2, \ldots, p$, of $\mathbf{X}$ can be equivalently replaced by $\|\nabla a_j\|^2_{\mathbb{R}^p} = 1$ where $a_j : \mathbb{R}^p \to \mathbb{R}$ is the linear functional represented by $\mathbf{a}_j$. Thus, the jth *nonlinear principal component* of $\mathbf{X}$ was defined as the r.v. $Z_j = \varphi_j(\mathbf{X})$

where φ_j solves

$$\begin{aligned}
&\max \mathbb{E}\big[u(\mathbf{X})^2\big],\\
&\text{sub} \quad \mathbb{E}\big[\big\|\nabla u(\mathbf{X})\big\|^2_{\mathbb{R}^p}\big]=1, \quad u \in \dot{W}^{1,2}_{\mathbf{X}},\\
&\qquad\quad \mathbb{E}\big[u(\mathbf{X})\varphi_s(\mathbf{X})\big]=0, \quad s=1,2,\dots,j-1, j>1,
\end{aligned} \tag{1}$$

and $\dot{W}^{1,2}_{\mathbf{X}}$ is the weighted Sobolev space of real valued nonlinear transformations u centered, square integrable (that is belonging to $\dot{\mathcal{L}}^2_{\mathbf{X}}$) and differentiable in the weak sense, with square integrable derivatives.

Denoting by $\Upsilon(D)$ the set of a.c. r.v.s X with zero mean, finite variance and density f_X having support the closure $\overline{D}$ of the interval $D=(a,b)$, $-\infty \le a < b \le +\infty$, we introduce the following definition.

Definition 2.1. *The r.v. $Z_j=\varphi_j(X)$ is said to be the jth nonlinear principal component of $X \in \Upsilon(D)$ if φ_j solves*

$$\begin{aligned}
&\max \mathbb{E}\big[u(X)^2\big],\\
&\text{sub} \quad \mathbb{E}\big[u'(X)^2\big]=1, \quad u \in \dot{W}^{1,2}_{X},\\
&\qquad\quad \mathbb{E}\big[u(X)\varphi_s(X)\big]=0, \quad s=1,2,\dots,j-1, j>1.
\end{aligned} \tag{2}$$

By $\dot{W}^{1,2}_X=\dot{W}^{1,2}(D,f_X)$ and $\dot{\mathcal{L}}^2_X=\dot{\mathcal{L}}^2(D,f_X)$ we denote the one-dimensional versions of spaces $\dot{W}^{1,2}_{\mathbf{X}}$ and $\dot{\mathcal{L}}^2_{\mathbf{X}}$ respectively. As known, $\dot{\mathcal{L}}^2_X$ is a separable Hilbert space with respect to the inner product $\langle u,v\rangle_0=\mathbb{E}[uv]$ with induced norm $\|u\|_0=\sqrt{\langle u,u\rangle_0}$, whereas $\dot{W}^{1,2}_X$ is an inner product space with respect to $\langle u,u\rangle_1=\mathbb{E}[u'v']$. A sufficient condition for $\dot{W}^{1,2}_X$ to be a separable Hilbert space (see [21]) that will assume in the sequel is

$$(1/f_X) \in \mathcal{L}^1_{\mathrm{loc}}(D). \tag{3}$$

Note that several usual distributions as e.g. the uniform, exponential or normal ones, satisfy (3).

As well known, in classical PCA the normalization constraint $\|\mathbf{a}_j\|^2_{\mathbb{R}^p}=1$ implies the boundedness of the variance $\mathbb{E}[\mathbf{a}^{\mathrm{T}}\mathbf{X}]$ of the transformed variables, which is a necessary condition for the existence of LPCs. The differential normalization constraint $\mathbb{E}[u'(X)^2]=1$ plays the same role in the nonlinear framework of Definition 2.1, but its effectiveness cannot be taken for granted: the existence of a constant C, depending on the domain D, such that

$$\mathbb{E}\big[u(X)^2\big] \le C\,\mathbb{E}\big[u'(X)^2\big] \quad \forall u \in \dot{W}^{1,2}_X \tag{4}$$

is not assured for any a.c. r.v. The inequality (4) is known as the *Chernoff–Poincaré inequality* (in the sequel CP-inequality). The interest for it in the probabilistic-statistical literature has notably grown starting from the seminal work of Chernoff [13] on the CP-inequality for Gaussian random variables. Nowadays several results on the validity of (4) for different univariate and multivariate distributions and some different applications to the characterization of the normal and uniform distributions, or to the central limit theorem are known (see e.g. [7–10,20,22]).

Note that, fixed $X \in \Upsilon(D)$, if there exists $u_* \in \dot{W}^{1,2}_X$ which realizes equality in (4) with $\mathbb{E}[u'_*(X)^2]=1$, then $u_*=\varphi_1$ that is, $Z=u_*(X)$ is the first NLPC of X and the (optimal) Poincaré constant C represents its variance. The same reasoning extend to the others NLPC transformations. The jth NLPC transformation φ_j (if there exists) of $X \in \Upsilon(D)$ is the optimal transformation (that is $C_j=\mathbb{E}[\varphi_j(X)^2]$) of the *restricted* CP-inequality:

$$\mathbb{E}\big[u(X)^2\big] \le C_j\,\mathbb{E}\big[u'(X)^2\big] \quad \forall u \in \Psi_j,$$

where

$$\Psi_j=\big\{u \in \dot{W}^{1,2}_X\colon \mathbb{E}\big[u'(X)^2\big]=1, \mathbb{E}\big[u(X)\varphi_s(X)\big]=0, s=1,2,\dots,j-1\big\}. \tag{5}$$

A final remark: since the CP-constant C can be defined as

$$C=\sup_{u \in \dot{W}^{1,2}_X} \frac{\mathbb{E}[u(X)^2]}{\mathbb{E}[u'(X)^2]}$$

it is possible to find r.v.s (see Example 3.4 later on) for which the CP-inequality holds but that do not have NLPCs.

Now we prove an existence result recurring to the so-called *direct methods of the calculus of variations* (see [6]), on which is based the estimation procedure introduced in Section 5. We set, for the sake of convenience,

$$\mathcal{F}[u] = \mathbb{E}\big[u^2(X)\big]; \qquad \mathcal{Q}[u] = \mathbb{E}\big[\big(u'(X)\big)^2\big].$$

Theorem 2.1. *If the embedding $\mathcal{I}_D : \dot{W}_X^{1,2} \to \dot{\mathcal{L}}_X^2$ is compact, then problem (2) admits countably many solutions φ_j.*

Proof. The first step to solve (2) is to look for solutions of the problem

$$\max \mathcal{F}[u] \quad \text{on } \Psi_1 = \big\{u \in \dot{W}_X^{1,2}\colon\ \mathcal{Q}[u] = 1\big\}. \tag{6}$$

Note that $\mathcal{F}[u] \geq 0$ and, assuming the validity of (4), there exists a constant C_1 such that $\sup \mathcal{F}[u] = C_1$. This implies the existence of a *maximizing sequence* $\{u_s\}$ in Ψ_1, i.e. $\mathcal{F}[u_s] \to C_1$. By definition of Ψ_1 the sequence $\{u_s\}$ is uniformly bounded (since $\mathcal{Q}$ is the norm of $\dot{W}_X^{1,2}$) hence there exists a subsequence $\{u_{s_j}\}$ such that u_{s_j} weakly converges in $\dot{W}_X^{1,2}$ to u. Since the validity of (4) means that the embedding $\mathcal{I}_D$ is continuous, it follows that $\mathcal{F}$ is continuous with respect to the weak convergence in $\dot{W}_X^{1,2}$. Furthermore, if $\mathcal{I}_D$ is compact, then u_{s_j} strongly converges to u in $\dot{\mathcal{L}}_X^2$ and this implies that the convergence is strong in $\dot{W}_X^{1,2}$ too.

Proved the existence of an element φ_1 in Ψ_1 such that

$$\mathcal{F}[\varphi_1] = \lambda_1 = \max\big\{\mathcal{F}[u]\colon\ u \in \Psi_1\big\}$$

the previous approach is applied to the sets Ψ_j defined in (5) giving the existence of two sequences $\{\lambda_j\}$ and $\{\varphi_j\}$ such that

$$\mathcal{F}[\varphi_j] = \lambda_j = \max\big\{\mathcal{F}[u]\colon\ u \in \Psi_j\big\}, \quad j \in \mathbb{N} \setminus \{0\}. \tag{7}$$

□

Remark 2.1. *The relation (7) can be stated equivalently in terms of the so-called Rayleigh quotient $\mathcal{Q}[u]/\mathcal{F}[u]$ as*

$$\mathcal{F}[\varphi_j] = \lambda_j = \max\big\{\mathcal{F}[u] \setminus \mathcal{Q}[u]\colon\ u \in \overline{\Psi}_j, u \neq 0\big\},$$

where $\overline{\Psi}_j = \{u \in \dot{W}_X^{1,2}\colon\ \mathbb{E}[u(X)\varphi_s(X)] = 0, s = 1, 2, \ldots, j-1\}$. Note that, by construction, λ_j represents the variance of the r.v. $Z_j = \varphi_j(X)$.

Next theorem gives some further insights on the statistical meaning of NLPCs. We recall that the *covariance operator* V of a random variable with values in a real separable Hilbert space H, having finite second moments is the unique linear operator from the dual H' to H such that

$$\big\langle V(\varphi), \psi\big\rangle_{H',H} = \mathbb{E}[\varphi\psi].$$

Theorem 2.2. *The solutions λ_j and φ_j of problem (2) are the eigenvalues and the corresponding eigenfunctions of the extension G to $\dot{\mathcal{L}}_X^2$ of the covariance operator V associated to X.*

If $\mathcal{I}_D$ is compact it holds:

(i) *there are countably many eigenvalues λ_j which are real positive, simple with 0 as unique limit point*;
(ii) *the eigenfunctions φ_j of G are mutually orthogonal and form a complete set in $\dot{W}_X^{1,2}$*;
(iii) *the jth eigenfunction φ_j has exactly j zeros in the open interval D*;
(iv) *if f_X and $1/f_X$ are in $\mathcal{L}^1(D)$ the following asymptotic formula holds*:

$$\lim_{j\to+\infty} j^2\lambda_j = \frac{\operatorname{meas}(D^2)}{\pi^2}.$$

Proof. We prove only the first statement, the others being standard consequences of the spectral theory of linear compact operators in Hilbert spaces (see [6,14]).

If φ_1 is the first solution of (2), applying the Lagrangian multipliers method, there exists a real positive constant λ_1 such that

$$\lambda_1\langle\varphi_1, h\rangle_1 = \langle\varphi_1, h\rangle_0 \quad \forall h \in \dot{W}_X^{1,2} \tag{8}$$

with the natural boundary conditions (possibly to be intended as limit):

$$f_X(a)\varphi_1'(a) = f_X(b)\varphi_1'(b) = 0. \tag{9}$$

The continuity of $l(h) = \langle u, h\rangle_0$, $h \in \dot{W}_X^{1,2}$, for any fixed $u \in \dot{\mathcal{L}}_X^2$ that follows by the CP-inequality, implies by Riesz representation theorem, the existence of a continuous linear operator $G : \dot{\mathcal{L}}_X^2 \to \dot{W}_X^{1,2}$ defined by

$$\langle\varphi_1, h\rangle_0 = \langle G\varphi_1, h\rangle_1 \quad \forall h \in \dot{W}_X^{1,2}. \tag{10}$$

Thus Eq. (8) can be written as

$$\langle G\varphi_1, h\rangle_1 = \lambda_1\langle\varphi_1, h\rangle_1 \quad \forall h \in \dot{W}_X^{1,2} \tag{11}$$

i.e., equality (8) is equivalent to state that λ_1 and φ_1 are the dominant eigenvalue and the corresponding eigenfunction of the operator G. The same argument can be repeated for all other solutions of (2). Hence finding NLPCs is equivalent to looking for a basis of $\dot{W}_X^{1,2}$ that diagonalizes G. Note that the restriction of G to $\dot{W}_X^{1,2}$ is invertible, positive and self-adjoint. Furthermore, the isometric isomorphism between the space of continuous bilinear forms on $\dot{W}_X^{1,2}$ and the one of continuous linear operator from $\dot{W}_X^{1,2}$ to its topological dual, guarantees that G represents in an unique way the extension to $\dot{W}_X^{1,2}$ of the covariance operator that represents V. Note that when $\dot{W}_X^{1,2}$ is densely contained in $\dot{\mathcal{L}}_X^2$ we can consider the (unique) extension of G to $\dot{\mathcal{L}}_X^2$. □

The last part of the previous proof shows that looking for the NLPCs of a r.v. X is equivalent to looking for a complete basis of $\dot{\mathcal{L}}_X^2$ that diagonalizes the extension to this space of the covariance operator of X. Hence, since $\{\varphi_j\}$ is an orthonormal basis of $\dot{W}_X^{1,2}$, by a density argument, $\{\lambda_j^{-1/2}\varphi_j\}$ represents an orthonormal basis of $\dot{\mathcal{L}}_X^2$ and, expanding with respect to this basis the identity function $i_D(X) = X$, we find

$$X = \sum_{j=0}^{+\infty}\varphi_j(X)\mathbb{E}[\varphi_j'].$$

Note that the elements φ_j are ordered by their variances. An extensive use of Fourier expansions of r.v.s was made in the work of Lancaster and its school (see e.g. [24]): in the framework of bivariate r.v.s (X, Y) they generalized the canonical linear analysis to a nonlinear context. We will see in Section 7 that many ideas for the applications of the nonlinear canonical variables in the study of the dependence structure between two r.v.s may be extended also for NLPCs.

The previous discussion puts in evidence that a key result is the compactness of the embedding $\mathcal{I}_D$: several conclusions about it are collected in the following theorem (see [1,6]).

Theorem 2.3. *For $X \in \Upsilon(D)$, with $D \subseteq \mathbb{R}$, the following conclusions hold*:

(i) *if D is bounded and there exist two constants c_1 and c_2 such that*

$$0 < c_1 \le f_X(x) \le c_2 \quad \forall x \in D \tag{12}$$

then the embedding $\mathcal{I}_D$ is compact;

(ii) *if* $f_X \in C^1(D)$ *is positive in every compact interval* $K \subset D$, *and near the boundary* ∂D *can be expressed as* $f_X(x) = g(\delta)$, *where* $g \in C^1$ *is positive*, *nondecreasing*, *with bounded derivative*, *such that* $\lim_{\delta\to 0^+} g(\delta) = 0$ *and* $\delta(x) = \operatorname{dist}(x, \partial D)$, *then* $\mathcal{I}_D$ *is compact*;

(iii) *if* $f_X \in C^0(D)$ *is bounded from above*, *and* $D_r = \{x \in D\colon |x| > r\}$, *then the compactness of* $\mathcal{I}_D$ *implies that for every* $k \in \mathbb{Z}$

$$\lim_{r\to+\infty} \mathrm{e}^{kr} \int_{D_r} f_X(x)\,\mathrm{d}x = 0; \tag{13}$$

(iv) *if* $D = \mathbb{R}$ *with symmetric density* $f_X(x) = g(|x|)$ *where* $g \in C^1([0,+\infty))$ *is positive*, *nonincreasing*, *with bounded derivative*, *then* $\mathcal{I}_D$ *is compact if and only if*

$$\lim_{s\to+\infty} \frac{g(s+\varepsilon)}{g(s)} = 0 \quad \forall \varepsilon > 0.$$

Remark 2.2. *As pointed out in* [30], *condition* (iii) *in Theorem* 2.3 *can be reformulated in terms of the moment generating function* (*m.g.f.*) *of* $X \in \Upsilon(D)$: *the existence on* $\mathbb{R}$ *of the m.g.f. is a necessary condition for the compactness of* $\mathcal{I}_D$.

We conclude this section observing that the regularity of the quadratic function which defines the functional $\mathcal{F}$ implies that the regularity of the eigenfunctions φ_j strictly depends on the regularity of the density f_X. More precisely, if $f_X \in C^m$ then $\varphi_j \in C^{m+1}$ for all j and $m \geq 1$.

3. Some examples and further results

The examples presented in this section are based on the possibility to analytically compute the NLPC transformations as solutions of the *Euler equation* associated to problem (2). It is a standard fact (see e.g. [6]) that a normalized (in $\dot{W}_X^{1,2}$) solution φ_j of $\lambda\langle u, h\rangle_1 = \langle u, h\rangle_0$ (the general version of (8)) is a *weak solution* corresponding to the constant $\xi_j = \lambda_j^{-1}$ of the *Sturm–Liouville problem* (SLP) with homogeneous Neumann boundary conditions (see [34])

$$\begin{cases} -\big(f_X u'\big)' = \xi f_X u, \\ f_X(a)u'(a) = f_X(b)u'(b) = 0. \end{cases} \tag{14}$$

From an operatorial point of view this means that φ_j is the jth eigenfunction corresponding to the eigenvalue ξ_j of the differential operator $\mathcal{G}$ (the inverse of G) defined by

$$\mathcal{G}[u] = -\frac{1}{f_X}\big(f_X u'\big)'. \tag{15}$$

A *strong solution* of (14) is a function $u : D \to \mathbb{R}$ such that u and its *quasi-derivative* $f_X u'$ are absolutely continuous on each compact subinterval of D and the equation is satisfied a.e. on D. The endpoint $a \in \mathbb{R} \cup \{\pm\infty\}$ is called *regular* if there exists $d \in D$ such that $(1/f_X) \in \mathcal{L}^1((a, d))$, otherwise it is called *singular*. The same definitions hold for b. Note that the regularity condition is stronger than (3) that guarantees $\dot{W}_X^{1,2}$ is Hilbert.

Remark 3.1. *If* ξ_j *and* φ_j *solve* (14) *then*, *by an integration by parts*, *we obtain*

$$\begin{aligned}
\xi_j \int_D x\varphi_j(x) f_X(x)\,\mathrm{d}x &= -\int_D x\big(\varphi_j'(x) f_X(x)\big)'\,\mathrm{d}x \\
&= -\big[x\varphi_j'(x) f_X(x)\big]_{\partial D} + \int_D \varphi_j'(x) f_X(x)\,\mathrm{d}x \\
&= \int_D \varphi_j'(x) f_X(x)\,\mathrm{d}x,
\end{aligned}$$

where the last equality descends from $\varphi_j' \in \dot{\mathcal{L}}_X^2$. Hence for any NLPCs transformation it holds:

$$\mathbb{E}[X\varphi_j] = \lambda_j \mathbb{E}\big[\varphi_j'\big]. \tag{16}$$

The simplest example of what exposed above is the following.

Example 3.1 (see [30]). *A r.v. X uniformly distributed on $(-a/2, a/2)$, $a \in \mathbb{R}_+$, by condition* (i) *in Theorem* 2.3, *admits NLPCs*. *Solving* (14) *we obtain*

$$\varphi_j(x) = -\frac{\sqrt{2}a}{j\pi}\cos\left(\frac{j\pi}{a}\left(x + \frac{a}{2}\right)\right), \quad j \in \mathbb{N} \setminus \{0\}, x \in (-a/2, a/2), \tag{17}$$

with corresponding variances $\lambda_j = a^2/j^2\pi^2$.

In particular, $\varphi_1(x) = \frac{\sqrt{2}a}{\pi}\sin(\frac{\pi}{a}x)$ with $\lambda_1 = a^2/\pi^2$: this last is the optimal CP-constant obtained by direct computation in [29].

The following example concerns a non-symmetric r.v. which satisfies (12).

Example 3.2. *Consider the r.v. X with density*

$$f_X(x) = K\mathrm{e}^{-\alpha x} \quad x \in (b, c), \tag{18}$$

where, fixed b and c such that $-\infty < b < 0 < -b < c < +\infty$, the parameters K and α are such that $\int_b^c f_X(x)\,\mathrm{d}x = 1$ and $\mathbb{E}[X] = 0$.

Solving the associated Sturm–Liouville problem

$$y'' - \alpha y' + \xi y = 0 \tag{19}$$

with Neumann condition $y'(b) = y'(c) = 0$, for $j \in \mathbb{N} \setminus \{0\}$ one finds

$$\begin{aligned}\varphi_j(x) = {}& \sqrt{2}\mathrm{e}^{\alpha x/2}\left(\frac{j\pi}{c-b}\cos\frac{j\pi(b-x)}{c-b} + \frac{\alpha}{2}\sin\frac{j\pi(b-x)}{c-b}\right) \\ & \times \left(\sqrt{a(c-b)}\xi_j \cdot \operatorname{signum}\left(\frac{j\pi}{c-b}\sin\left(\frac{jb\pi}{c-b}\right) - \frac{\alpha}{2}\cos\left(\frac{jb\pi}{c-b}\right)\right)\right)^{-1}\end{aligned}$$

with corresponding NLPCs variances

$$\lambda_j = \left(\frac{\alpha^2}{4} + \frac{j^2\pi^2}{(c-b)^2}\right)^{-1}.$$

Note that, *as prescribed by Theorem* 2.2, *we have*

$$j^2\lambda_j \sim \frac{(c-b)^2}{\pi^2} \quad \text{as } j \to +\infty.$$

The next example, treating a singular case on a bounded domain, shows that even when it is not immediate to derive an explicit form of NLPC transformations, their existence can be however stated.

Example 3.3. *Let us consider the family of generalized beta distributions with density f_X defined by*

$$f_X(x) = C\big(1 - x^2\big)^{\alpha}, \quad x \in (-1, 1), \alpha \in \mathbb{R}_+, \tag{20}$$

where C is the normalization constant. Note that for $\alpha = 0$ we obtain the uniform distribution, for $\alpha > 0$ we have the density of an unimodal symmetric distribution (*in particular, for $\alpha = 1/2$ the Wigner one*).

All the r.v.s of this family admit NLPCs. In fact, for $\alpha = 0$ *the result is stated in Example* 3.1. *Noting that*

$$\left(1 - x^2\right)^{\alpha} = \delta^{\alpha}(x)\left(2 - \delta(x)\right)^{\alpha}$$

having set $\delta(x) = 1 - |x|$ *and applying* (ii) *of Theorem* 2.3, *the existence of NLPC transformations is proved for* $\alpha \geq 1$. *With a similar reasoning the same conclusion holds for* $0 < \alpha < 1$ *thanks to Theorem* 8.8 *in* [19].

If the support D is unbounded, the existence of NLPCs is not guaranteed, as showed in the following example.

Example 3.4. *The r.v.* $X \in \Upsilon([-\alpha^{-1}, +\infty))$ *with* $\alpha > 0$ *and density*

$$f_X(x) = \alpha \exp\{-1 - \alpha x\}$$

does not satisfy the necessary condition of compactness of Remark 2.2.

Operating as in Example 3.2, *it is easy to prove that the associated Sturm–Liouville problem has no solutions*, *i.e.* X *has not optimal transformations satisfying* (2).

We note that for an exponential distribution with density $f_X(x) = \alpha \mathrm{e}^{-\alpha x}$ *on* $(0, +\infty)$ *the CP-inequality*

$$\operatorname{Var}\left[u(X)\right] \leq \frac{4}{\alpha^2} \mathbb{E}\left[\left(u'(X)\right)^2\right] \tag{21}$$

was proved in [20] (*for a related but different inequality see also* [8]). *Observe that equality in* (21) *cannot be reached by any* $\mathcal{L}_X^2$ *function*; *furthermore*, $4/\alpha^2$ *is the reciprocal of the value of* ξ *which corresponds to a null discriminant of the characteristic equation associated to* (19) *and* (*see Example* 3.2):

$$\lim_{c \to +\infty} \lambda_j = \lim_{c \to +\infty} \left(\frac{\alpha^2}{4} + \frac{j^2 \pi^2}{(c - b)^2}\right)^{-1} = \frac{4}{\alpha^2}.$$

We show now that the normal distribution admits NLPCs.

Example 3.5 (see [31]). *Let us consider* $X \sim \mathcal{N}(0, \sigma^2)$. *By* (iv) *of Theorem* 2.3, X *admits NLPCs. To compute them, observe that the differential operator* $\mathcal{G}[u] = -\mathrm{e}^{x^2/2\sigma^2}(\mathrm{e}^{-x^2/2\sigma^2} u')'$ *associated to* (14) *is the well-know Ornstein–Uhlenbeck operator having* (*nonconstant*) *eigenfunctions*, *orthonormal in* $\dot{W}_X^{1,2}$, *given by*

$$\varphi_j(x) = \left(j \sigma^{2(j-1)}\right)^{-1/2} H_j(x), \quad j \in \mathbb{N} \setminus \{0\},$$

where H_j *is the* j*th Hermite polynomial*

$$H_j(x) = \frac{(-\sigma^2)^j}{\sqrt{j!}} \mathrm{e}^{x^2/2\sigma^2} \frac{\mathrm{d}^j}{\mathrm{d}x^j}\left(\mathrm{e}^{-x^2/2\sigma^2}\right), \quad j \in \mathbb{N} \setminus \{0\}.$$

The variances λ_j *of the NLPCs* $Z_j = \varphi_j(X)$ *are expressed in a simple form in terms of the variance of* X:

$$\lambda_j = \frac{\sigma^2}{j}, \quad j \in \mathbb{N} \setminus \{0\}.$$

The first NLPC transformation is the identity function $\varphi_1(x) = x$: *as proved in* [31] *in a multivariate framework, this happens only in the Gaussian case. The optimal Poincaré constant* $\lambda_1 = \sigma^2$ *coincides with the one found in* [13].

Note furthermore that λ_2 *coincides with the restricted CP-constant* R_X^* *introduced in* [22] *as*

$$R_X^* = \sup_{u \in {}_*\dot{W}_X^{1,2}} \frac{\mathbb{E}[u(X)^2]}{\mathbb{E}[u'(X)^2]},$$

where ${}_*\dot{W}_X^{1,2} = \dot{W}_X^{1,2} \cap \{u\colon \mathbb{E}[u'] = 0\}$, *that is, to be orthogonal to the first NLPC transformation* φ_1 *and to have centered first derivative coincides*:

$$\mathbb{E}\bigl[\varphi_1(X)\varphi_2(X)\bigr] = 0 \quad \Leftrightarrow \quad \mathbb{E}\bigl[\varphi_2'\bigr] = 0.$$

This is a direct consequence of the Stein equality (*compare with* (16)) *for Gaussian distributions*

$$\mathbb{E}\bigl[u'(X)\bigr] = \mathbb{E}\bigl[Xu(X)\bigr], \quad u \in \dot{W}_X^{1,2},$$

and $\varphi_1(x) = x$. *This result is not true for any distribution. For example*: *the transformation* φ_2 *in Example* 3.2 *does not satisfy* $\mathbb{E}[\varphi_2'(X)] = 0$.

Theorem 2.2 of Section 2 states many properties of the NLPCs and their corresponding variances. We investigate some further properties particularly interesting for their statistical meaning and applicability.

We start with a result which relates the symmetry of the density f_X of a r.v. X with the one of its NLPC transformations.

Theorem 3.1. *Let* $X \in \Upsilon(D)$ *be a r.v. that admits NLPCs. If* f_X *is even then the NLPC transformations* φ_j *are alternatively even and odd*, *with* φ_1 *odd*.

Proof. If f_X is even, then the operator $\mathcal{G}$ defined in (15) is even in the following sense:

$$\begin{aligned}\mathcal{G}(-x) &= -\frac{1}{f_X(-x)}\frac{\mathrm{d}}{\mathrm{d}(-x)}\left(f_X(-x)\frac{\mathrm{d}}{\mathrm{d}(-x)}y(x)\right)\\ &= -\frac{1}{f_X(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left(f_X(x)\frac{\mathrm{d}}{\mathrm{d}x}y(x)\right) = \mathcal{G}(x)\end{aligned}$$

and this implies (see [2]) that its eigenfunctions are odd or even. Since the dominant eigenfunction, by Theorem 2.2(iii), has one and only one zero, it must be odd. The orthogonality properties of φ_j's imply the general part of the thesis. □

Remark 3.2. *Theorem* 3.1 *implies that if* f_X *is even then* φ_2 *is the transformation which corresponds to the optimal CP-constant*

$$\lambda_2 = \sup_{u \in {}_*\dot{W}_X^{1,2} \cap \Psi_2^*} \frac{\mathbb{E}[u(X)^2]}{\mathbb{E}[u'(X)^2]},$$

where

$$\Psi_2^* = \bigl\{u \in \dot{W}_X^{1,2}\colon \mathbb{E}\bigl[u(X)\varphi_1(X)\bigr] = 0\bigr\}.$$

The second result we present concerns a monotonicity property of the first NLPC transformation, result which cannot be extended to the others NLPCs transformations.

Theorem 3.2. *Let* $X \in \Upsilon(D)$ *be a r.v. which admits NLPCs. If* $f_X \in C^1(D)$ *then the first NLPC transformation* φ_1 *of* X *is strictly monotone*.

Proof. We assume for simplicity $\overline{D} = [a, b]$, since the proof can be straightforwardly extended to the unbounded case. By the assumptions it follows the existence of the first NLPC transformation $\varphi_1 \in C^2([a, b])$. Since φ_1 is a solution of (14), setting $g(x) = \varphi'(x) f_X(x)$, it follows the existence of $c \in (a, b)$ such that

$$g'(c) = -\bigl(f_X\varphi_1'\bigr)'(c) = \xi_1 f_X(c)\varphi_1(c) = 0$$

and the positivity of ξ_1 and f_X implies $\varphi_1(c) = 0$. From Theorem 2.2(iii) c is the unique interior zero of φ_1, hence φ_1' has constant sign on D. In fact, if there existed $d \in (a, b)$, $d \neq c$, such that $\varphi_1'(d) = 0$, then $g(d) = 0$ and hence it would exist $h \in (a, d)$ such that $g'(h) = 0$ and consequently $\varphi_1(h) = 0$. Since φ_1 is the dominant solution of (14) it follows $h = c$. □

The following proposition shows that, in several cases, the knowledge of only the first NLPC transformation univocally determines the density of X. The conclusion and the proof (which we omit) substantially agree with the ones presented in Lemma 4.2 in [12] and Theorem 3.8 in [10] regarding the optimal transformation for the CP-inequality but, differently from these authors, we do not need to make any assumption on the first derivative of φ_1 as a consequence of our Theorem 3.2.

Theorem 3.3. *If $\varphi_1 \in C^2(D)$ is the first NLPC transformation of $X \in \Upsilon(D)$, then its density f_X is of the form $f_X = g/\int_D g$, where*

$$g(x) = \frac{1}{\varphi_1'(x)} \exp\left\{-\xi_1 \int \frac{\varphi_1}{\varphi_1'}\right\}, \tag{22}$$

$\int \varphi_1/\varphi_1'$ is an indefinite integral and ξ_1 is the eigenvalue associated to φ_1 by (14).

4. Spline estimates of NLPCs

In this section we introduce estimators for the NLPCs transformations φ_j and their corresponding variances λ_j of a given r.v. X, stating the main results about their asymptotic properties.

To define an estimation procedure we adapt in a statistical perspective the classical *Rayleigh–Ritz method* (see e.g. [18]): we find an approximate version of the maximizers and the corresponding maxima of the reciprocal of the Rayleigh quotient $\mathcal{F}[u]/\mathcal{Q}[u]$ (see Remark 2.1) on a sequence of appropriate finite-dimensional subspaces of $\dot{W}_X^{1,2}$ generated by B-splines (see e.g. [16] and [32]). In such a way we obtain nonparametric estimates of the eigenfunctions φ_j and the corresponding eigenvalues λ_j of the covariance operator G defined in (10).

We will refer to a r.v. $X \in \Upsilon(D)$ that admits NLPCs, assuming $D = (a, b)$ is bounded. Fixed two integers $k > 0$ and $d \geq 2$, we denote by $\mathcal{S}_{k,d}$ the $(k + d)$-dimensional linear space of spline functions of order $d + 1$ defined on $\overline{D}$, having $k - 1$ interior knot points $r_0 = a < r_1 < \cdots < r_{k-1} < b = r_k$.

Consider now a sample $\{X_i, i = 1, 2, \ldots, n\}$ of i.i.d. r.v.s drawn from X, and denote by $\widehat{\dot{W}}_k^{1,2}$ the space of spline functions $u \in \mathcal{S}_{k,d}$ such that $n^{-1} \sum_{i=1}^n u(X_i) = 0$, a.s. This space is Hilbert with respect to the inner product $n^{-1} \sum_{i=1}^n u'(X_i) v'(X_i)$.

We introduce the estimator of λ_j and φ_j by the following:

Definition 4.1. *Let j, k and $d \geq 2$ be positive integers such that $j \leq k + d - 1 < n$ and*

$$\mathcal{F}_n[u] = \frac{1}{n} \sum_{i=1}^n \big(u(X_i)\big)^2, \qquad \mathcal{Q}_n[u] = \frac{1}{n} \sum_{i=1}^n \big(u'(X_i)\big)^2.$$

We define the estimator $\widehat{\lambda}_{j,k,n}$ of the jth eigenvalue λ_j of G as

$$\widehat{\lambda}_{j,k,n} = \max_{u \in \widehat{\dot{W}}_k^{1,2}} \left\{ \frac{\mathcal{F}_n[u]}{\mathcal{Q}_n[u]} \colon \frac{1}{n} \sum_{i=1}^n u(X_i) \widehat{\varphi}_{h,k,n}(X_i) = 0 \text{ a.s.}, 1 \leq h \leq j - 1 \right\} \tag{23}$$

and the estimator of the associated eigenfunction φ_j as the maximizer $\widehat{\varphi}_{j,k,n} \in \widehat{\dot{W}}_k^{1,2}$ corresponding to $\widehat{\lambda}_{j,k,n}$.

The technical aspects which illustrate how to convert problem (23) into a generalized eigenvalue problem and thus to obtain an explicit expression for the estimates of NLPCs transformations by means of B-splines are discussed in the Section 7.

We state the main asymptotic results on the estimators $\widehat{\lambda}_{j,k,n}$ and $\widehat{\varphi}_{j,k,n}$.

Theorem 4.1. *Let $X \in \Upsilon(D)$ be a r.v. which admits NLPCs transformations φ_j. We assume that*:

(H1) *knots are equispaced*;
(H2) *f_X satisfies condition* (12);
(H3) *there exists $m > 2$ such that $\varphi_j \in C^m(\overline{D})$ with mth derivative Lipschitz continuous of order α, $0 < \alpha \leq 1$.*

Let $\beta = m + \alpha$ and consider $k \leq n^{(1-\delta)/(2\beta+5)}$ for $0 < \delta < 1$. Fixed j, under conditions (H1)–(H3) *and taken $d \geq m + \alpha$*:

(a) *a unique solution to problem* (23) *exists except on an event whose probability goes to zero as $n \to +\infty$*;
(b) *when $n \to +\infty$ the following results hold*:

$$|\widehat{\lambda}_{j,k,n} - \lambda_j| \longrightarrow 0 \quad \textit{in probability},$$

$$\|\widehat{\varphi}_{j,k,n} - \varphi_j\|_0 \longrightarrow 0 \quad \textit{in probability}.$$

The proof of this theorem directly follows from Theorems 7.1 and 7.2 in Section 7. More precisely, in Theorem 7.1 we obtain the convergence of the population spline approximation $\widetilde{\lambda}_{j,k}$ and $\widetilde{\varphi}_{j,k}$ to λ_j and φ_j respectively. Then Theorem 7.2 concerns the convergence in probability of the sequence of estimators $\widehat{\lambda}_{j,k,n}$ and $\widehat{\varphi}_{j,k,n}$ to their approximation counterparts.

Some remarks on assumptions (H1)–(H3) are in order. Condition (H1) is standard in spline framework; however the results of Theorem 4.1 remain true when (H1) is replaced by the following asymptotic condition: the distance between two contiguous knots r_s and r_{s+1} satisfies

$$\max_s |r_{s+1} - r_s| = \mathrm{O}(k^{-1}) \quad \text{and} \quad \frac{1}{\min_s |r_{s+1} - r_s|} = \mathrm{O}(k).$$

The assumption (H2) and the regularity hypothesis (H3) are typical in many studies of asymptotic properties of nonparametric estimators as, for example, in the estimation of the regression function (see among the others [5]), in the estimation of optimal transformations of variables (see [4]) or in deriving rates of convergence of splines estimates of additive principal components (see [17]).

Remark 4.1. *Let $\beta = m + \alpha$. Under the same assumptions of Theorem 4.1 and choosing $k \sim n^{1/(4\beta+1)}$, an upper bound for the rate of convergence for both the estimators of λ_j and φ_j is $n^{-\beta/(4\beta+1)}$. This bound cannot be considered optimal if compared with the usual optimal rate of convergence in a nonparametric framework* (*see, for instance*, [17] *and* [33]) *and the one of the approximations in Theorem 7.1. Probably, it is necessary to use some other arguments in the proof of Theorem 7.2 to improve our bound. These aspects need further investigations in a future work.*

5. Simulation study

To asses the practical performance of the proposed estimator we run a simulation study: we make comparison between the estimations of the NLPC transformations obtained in the Examples 3.1, 3.2 and 3.5 when sample size varies. Our aim is to compare some empirical measures for the distributions of the estimated eigenvalues (mean, standard deviation and mean square error) and the following error criterion for the eigenfunctions

$$E_j^2 = \frac{1}{\lambda_j} \frac{1}{M} \sum_{h=1}^{M} \|{}_h\widehat{\varphi}_{j,k,n} - \varphi_j\|_0^2,$$

where M is the number of simulations in each case study and ${}_h\widehat{\varphi}_{j,k,n}$ denotes the estimation of φ_j at the hth simulation.

We base simulations on the following operative conditions:

1. sample sizes are $n = 50, 100, 200, 500, 1000, 2000$ and 5000;
2. we provide the results on the first two eigenvalues λ_j and eigenfunctions φ_j $(j = 1, 2)$ for the uniform distribution on $(-1, 1)$, the truncate exponential on $(-1, 2)$ and the standardized normal $\mathcal{N}(0, 1)$;

3. the number of simulations for each setting of experimental factors is $M = 2000$.

As in the greatest part of the applications of spline functions, we work with cubic splines ($d = 3$): this choice is sufficient to guarantee the wished degree of regularity of the estimates. The number of knots varies between 3 and 15. Differently from condition (H1) in Theorem 4.1, the knots are placed at sample quantiles: this choice does not invalidate the results of the theorem, and it is desiderable in practice, as pointed out for example in [4]. For evidently reasons of synthesis in the next we collect only some selected results. The norm $\|\cdot\|_0^2$ in the above defined error E_j^2 is evaluated by the trapezoidal rule integration.

5.1. *Uniform distribution*

Tables 1 and 2 provide the results when the samples are drawn from a uniform r.v. on $(-1, 1)$ (see also Example 3.1) with 3 interior knots. Note that the estimators show good performances for reasonably large sample sizes both for eigenvalues and for eigenfunctions. For $n \geq 2000$, we would get similar results using 4 knots. This confirms the intuition that the parameter k has to grow quite slowly with respect to the sample size although.

From Tables 1 and 2 it emerges the presence of a positive bias that decreases when n increases: in relative terms, the bias for the first eigenvalue is of the order of 10% when the sample is small and decreases under 1% when n is largest than 500. The results are better for the second eigenvalue: for $n = 50$ the relative bias is around 7% and it is less then 0.5% for $n \geq 500$.

To complete this example, we have drawn an estimate of the first two NLPCs obtained from a sample of 200 elements and 3 interior knots: the graphics of Fig. 1 tends to confirm the good performances of the estimation procedure for a reasonably large sample size.

Table 1
Estimation errors for the first eigenvalue and associated eigenfunction of a uniform distribution

	λ_1			φ_1
n	Mean	St. dev	MSE	E_1
50	0.4460	0.0978	1.12×10^{-2}	4.29×10^{-2}
100	0.4231	0.0619	4.15×10^{-3}	2.04×10^{-2}
200	0.4147	0.0426	1.90×10^{-3}	9.69×10^{-3}
500	0.4090	0.0258	6.80×10^{-4}	3.94×10^{-3}
1000	0.4065	0.0178	3.19×10^{-4}	1.88×10^{-3}
2000	0.4061	0.0128	1.65×10^{-4}	9.70×10^{-4}
5000	0.4059	0.0083	6.84×10^{-5}	3.80×10^{-4}

Table 2
Estimation errors for the second eigenvalue and associated eigenfunction of a uniform distribution

	λ_2			φ_2
n	Mean	St. dev	MSE	E_2
50	0.1086	0.0238	6.18×10^{-4}	1.22×10^{-1}
100	0.1047	0.0149	2.33×10^{-4}	6.15×10^{-2}
200	0.1026	0.0106	1.14×10^{-4}	3.03×10^{-2}
500	0.1018	0.0064	4.13×10^{-5}	1.22×10^{-2}
1000	0.1017	0.0044	1.97×10^{-5}	6.07×10^{-3}
2000	0.1013	0.0033	1.06×10^{-5}	3.08×10^{-3}
5000	0.1013	0.0020	3.88×10^{-6}	1.27×10^{-3}

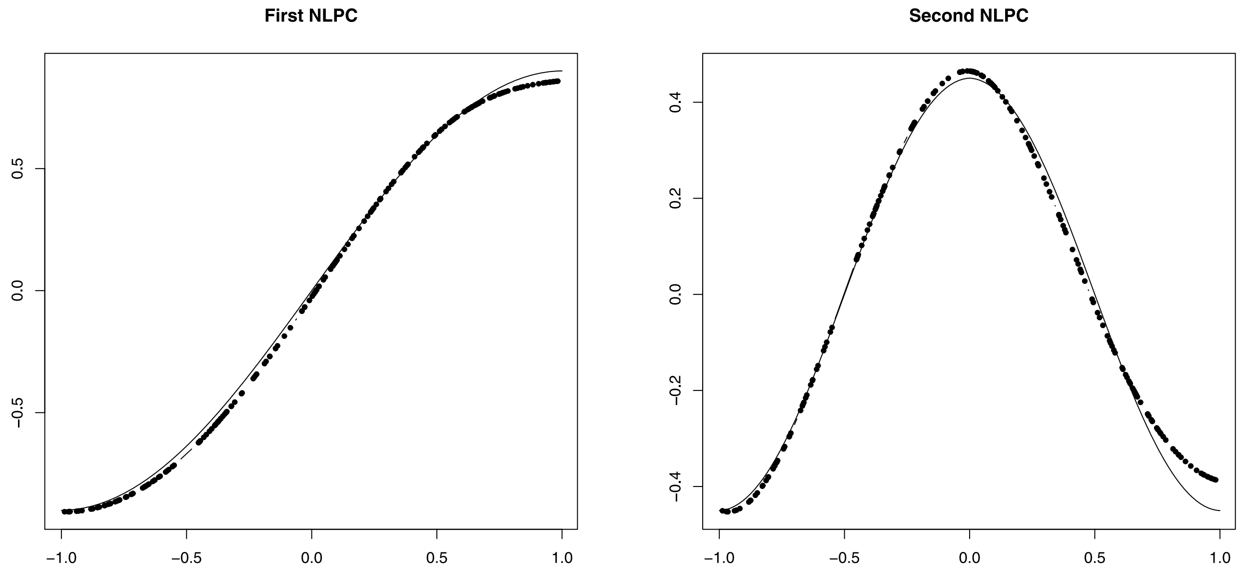

Fig. 1. Estimates of the first and second NLPC transformations for the uniform distribution on $(-1, 1)$.

5.2. *Truncated exponential distribution*

Also for the truncated exponential distribution on $(-1, 2)$ (see Example 3.2), results are encouraging: from Tables 3 and 4, which summarize the results obtained with 5 knots, we can confirm the goodness of the estimator of the λs with a bias that decreases rapidly when the sample size increases and the goodness of fit of the estimated eigenfunctions. An estimate when we use a sample of size $n = 200$ of the first two NLPC transformations is plotted in Fig. 2: as we can see, the obtained curves are smooth and fit very well the shape of the true NLPC transformations.

Table 3
Estimation errors for the first eigenvalue and associated eigenfunction of a truncated exponential distribution

	λ_1			φ_1
n	Mean	St. dev	MSE	E_1
50	0.9470	0.2620	8.57×10^{-2}	5.95×10^{-2}
100	0.8692	0.1598	2.83×10^{-2}	2.61×10^{-2}
200	0.8450	0.1063	1.21×10^{-2}	1.27×10^{-2}
500	0.8265	0.0638	4.17×10^{-3}	4.82×10^{-3}
1000	0.8214	0.0445	2.01×10^{-3}	2.48×10^{-3}
2000	0.8181	0.0315	9.96×10^{-4}	1.22×10^{-3}
5000	0.8164	0.0204	4.16×10^{-4}	4.90×10^{-4}

5.3. *Normal distribution*

We conclude the collection of examples by proposing the results about the estimates of eigenvalues and associated eigenfunctions when the samples come from a standard normal distribution (see Example 3.5). Since D is unbounded, the boundary points are chosen as the minimum and the maximum of the observed data. The results (see Tables 5 and 6), with 3 knots, are good both for the eigenvalues and the eigenfunctions as is evidenced by the estimates, displayed in Fig. 3, of the first two NLPC transformations with $n = 200$. These results show that the estimation procedure may be used with success also when we consider a r.v. admitting NLPCs and defined on an unbounded set.

Table 4
Estimation errors for the second eigenvalue and associated eigenfunction of a truncated exponential distribution

	λ_2			φ_2
n	Mean	St. dev	MSE	E_2
50	0.2429	0.0607	4.14×10^{-3}	1.83×10^{-1}
100	0.2311	0.0402	1.70×10^{-3}	8.87×10^{-2}
200	0.2271	0.0275	7.87×10^{-4}	4.38×10^{-2}
500	0.2237	0.0169	2.89×10^{-4}	1.75×10^{-2}
1000	0.2224	0.0120	1.46×10^{-4}	8.59×10^{-3}
2000	0.2221	0.0085	7.28×10^{-5}	4.26×10^{-3}
5000	0.2219	0.0054	2.89×10^{-5}	1.82×10^{-3}

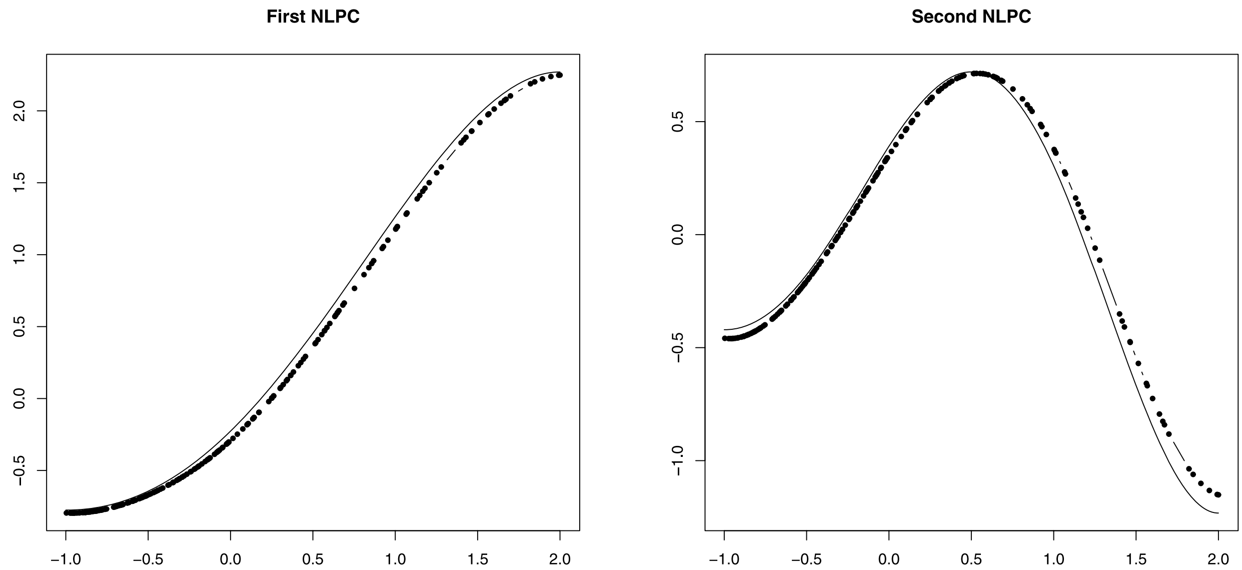

Fig. 2. Estimates of the first and second NLPC transformations for the truncated exponential on $(-1, 2)$.

6. Some applications of NLPCs

The properties discussed in Section 3 represent a theoretical support to apply NLPCs in some statistical contexts; the estimation procedure, introduced in Section 4 provides the tool that permits to make operational the NLPCs in practice.

The aim of this Section is to show some examples of these possible applications, developing principally two streams: the first one based on the characterization property of the first transformation φ_1 stated in Theorem 3.3, the second one on the fact that the set of NLPCs transformations is the basis of $\mathring{\mathcal{L}}_X^2$ that diagonalizes the covariance operator acting on this space.

In particular, the characterization property is the ground in defining a goodness-of-fit test: the statistic used in the test is based on the comparison between the estimate of λ_1 and its value under the null hypothesis. About the second aspect, we show how to use the transformations in study the dependence structure of a bivariate r.v., illustrating how to construct a bivariate distribution with fixed marginals using NLPCs.

Table 5
Estimation errors for the first eigenvalue and associated eigenfunction of a standardized normal distribution

	λ_1			φ_1
n	Mean	St. dev	MSE	E_1
50	1.1534	0.2815	1.03×10^{-1}	1.16×10^{-1}
100	1.0836	0.1813	3.99×10^{-2}	6.94×10^{-2}
200	1.0452	0.1212	1.67×10^{-2}	4.68×10^{-2}
500	1.0175	0.0678	4.91×10^{-3}	1.97×10^{-2}
1000	1.0070	0.0498	2.53×10^{-3}	1.20×10^{-2}
2000	1.0035	0.0364	1.33×10^{-3}	6.17×10^{-3}
5000	1.0009	0.0220	4.84×10^{-4}	2.47×10^{-3}

Table 6
Estimation errors for the second eigenvalue and associated eigenfunction of a standardized normal distribution

	λ_2			φ_2
n	Mean	St. dev	MSE	E_2
50	0.5180	0.1720	2.99×10^{-2}	3.78×10^{-1}
100	0.5258	0.1363	1.92×10^{-2}	2.58×10^{-1}
200	0.5279	0.1145	1.39×10^{-2}	1.90×10^{-1}
500	0.5228	0.0784	6.66×10^{-3}	1.23×10^{-1}
1000	0.5177	0.0667	4.76×10^{-3}	1.00×10^{-1}
2000	0.5141	0.0514	2.84×10^{-3}	7.67×10^{-2}
5000	0.5080	0.0327	1.13×10^{-3}	4.46×10^{-2}

6.1. *Goodness-of-fit test*

We have seen in Theorem 3.3 that when the couple (λ_1, φ_1) exists, it characterizes the distribution of a given r.v. $X \in \Upsilon(D)$: this fact may be exploited for constructing a goodness-of-fit test. Moreover, in some cases, some additional results linked to particular families of distributions, permit to use only λ_1 with a considerable simplification in defining the statistic in the test.

We expose this idea referring to the uniform case: suppose we have to test that the distribution of $X \in \Upsilon((-1,1))$ is uniform

$$\mathcal{H}_0\colon\ X \sim \mathcal{U}(-1,1)$$

against the alternative

$$\mathcal{H}_1\colon\ X \sim \mathcal{S}(-1,1),$$

where $\mathcal{S}$ is a unimodal symmetric distribution on $(-1,1)$ such that $\mathcal{S} \not\sim \mathcal{U}$.

We know (see Example 3.1) that the variance of the first NLPC of a uniform distribution on $(-1,1)$ is $\lambda_1 = 4/\pi^2$ and, by the result of [29], that it is the largest Poincaré constant for all unimodal symmetric r.v.s in $\Upsilon((-1,1))$. This leads to the equivalent hypotheses

$$\mathcal{H}_0\colon\ \lambda_1 = 4/\pi^2 \quad \text{vs.} \quad \mathcal{H}_1\colon\ \lambda_1 < 4/\pi^2.$$

We thus define the test statistic

$$D_{n,k} = \sqrt{n}\big(\widehat{\lambda}_{1,k,n} - 4/\pi^2\big)$$

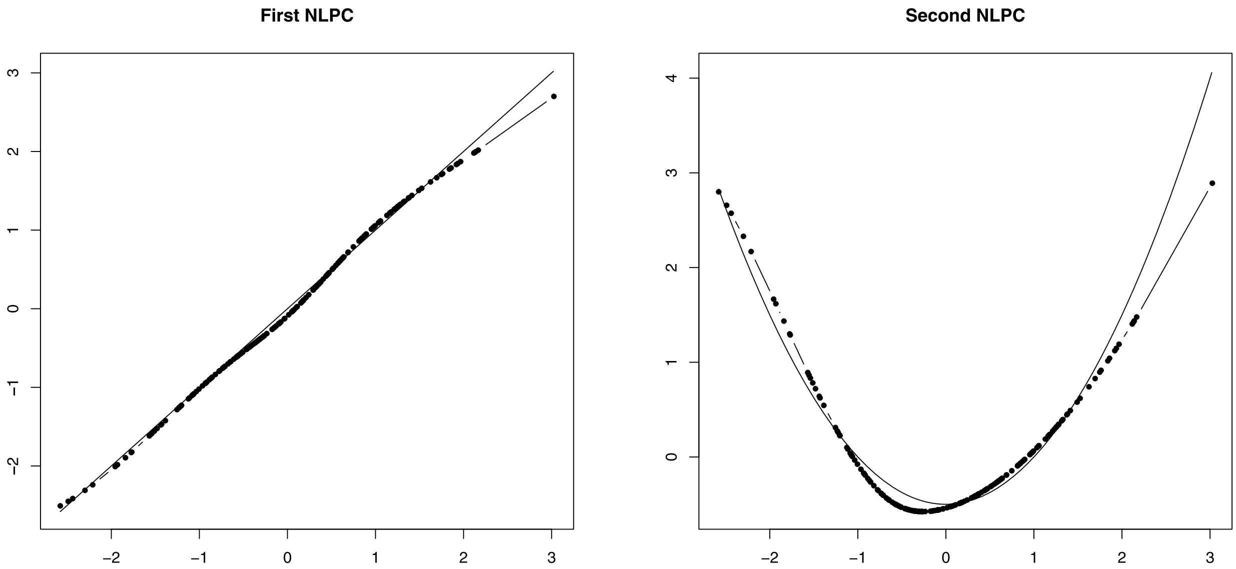

Fig. 3. Estimate of the first and second NLPC transformations for the standard normal distribution.

Table 7
Probabilities of rejecting hypothesis of uniformity using the $D_{n,k}$ statistic (in bold) and the Kolmogorov–Smirnov statistic (in italics) when sample size is 200

Underlying distributions	Level of significance					
	0.10		0.05		0.01	
$\mathcal{U}(-1, 1)$	**0.097**	*0.093*	**0.048**	*0.043*	**0.008**	*0.008*
$\mathcal{N}^{\mathcal{T}}(0, 2)$	**0.210**	*0.096*	**0.125**	*0.048*	**0.034**	*0.008*
$\mathcal{N}^{\mathcal{T}}(0, 3/2)$	**0.348**	*0.111*	**0.223**	*0.056*	**0.076**	*0.010*
$\mathcal{N}^{\mathcal{T}}(0, 1)$	**0.752**	*0.221*	**0.619**	*0.118*	**0.361**	*0.024*
$\mathcal{N}^{\mathcal{T}}(0, 3/4)$	**0.981**	*0.551*	**0.963**	*0.353*	**0.865**	*0.102*
$\mathcal{N}^{\mathcal{T}}(0, 1/2)$	**1.000**	*0.999*	**1.000**	*0.995*	**1.000**	*0.911*

and the critical region at level of significance α, $0 < \alpha \le 1$, is $\{D_{n,k} \le d_n^\star(\alpha)\}$.

In order to give a short but meaningful exemplification, we conduct the study for selected levels in the i.i.d. case with sample size fixed to 200. We defer to a future work a deeper analysis.

We obtain the critical values of the test by a Monte Carlo calculation: fifty hundred samples were generated from a uniform distribution on $(-1, 1)$ and the null distribution of $D_{n,k}$ was thus estimated, by using 3 knots. The values $d_n^\star(\alpha)$ are the empirical quantiles of order α of these distributions, with $\alpha = 0.10$, 0.05 and 0.01.

A Monte Carlo investigation was carried out for the level and the power of this test. The alternatives consist of truncated normal distributions $\mathcal{N}^{\mathcal{T}}(0, \sigma)$ on $(-1, 1)$ with $\sigma = 2$, $4/3$, 1, $3/4$ and $1/2$. For each distribution ten hundred samples of size $n = 200$ were generated and the probabilities of rejection using the test based on $D_{n,k}$ was determined by the proportion of cases falling into the critical region. The power estimates are thus compared with that ones obtained by the Kolmogorov–Smirnov test.

The results, collected in Table 7, show that for all the selected distributions, our test performs reasonably well and it is better than the known Kolmogorov–Smirnov. This fact encourages further developments: investigations need for the study of the exact or asymptotic distribution of the statistic proposed and for a more complete study of the power of this test.

6.2. *Construction of bivariate distributions*

The specification of a continuous joint bivariate distribution with fixed margins is required in many statistical applications. For instance, examining the robustness of multivariate tests, comparing the multivariate goodness-of-fit tests, simulating in a portfolio analysis, constructing a multivariate prior in Bayesian inference.

In the literature various methods to form some families of bivariate distributions with given marginal and a dependence structure specified by a vector of parameters have been investigated (see e.g. [26]). In the following, we illustrate a method to generate some bivariate distributions by using NLPC transformations which is based on Lancaster's thought [24].

Let $X \in \Upsilon(D_X)$ and $Y \in \Upsilon(D_Y)$ be two r.v.s with densities f_X and f_Y respectively, $\{a_h\}$ and $\{b_j\}$ some normalized elements of $\dot{\mathcal{L}}_X^2$ and $\dot{\mathcal{L}}_Y^2$ respectively. Then

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)\left(1 + \sum_{h=1}^{N}\sum_{j=1}^{N} \rho_{h,j} a_h(x) b_j(y)\right) \tag{24}$$

is a joint density distribution on $D_X \times D_Y$ if $\rho_{h,j}$ are chosen such that

$$\sup_{x,y}\left[-\sum_{h=1}^{N}\sum_{j=1}^{N} \rho_{h,j} a_h(x) b_j(y)\right] \le 1.$$

Clearly $\{a_h\}$ and $\{b_j\}$ can be selected into the orthonormal bases of $\dot{\mathcal{L}}_X^2$ and $\dot{\mathcal{L}}_Y^2$ respectively.

To describe the dependence structure between X and Y one usually uses the Pearson's ϕ^2 index of independence, the covariance $\mathbb{E}[XY]$ and the regression functions $\mathbb{E}[Y|x]$ and $\mathbb{E}[X|y]$. When the density is of the type (24), we have

$$\phi^2 = \sum_{h=1}^{N}\sum_{j=1}^{N} \rho_{h,j}^2, \qquad \mathbb{E}[XY] = \sum_{h=1}^{N}\sum_{j=1}^{N} \rho_{h,j}\alpha_h\beta_j,$$

where $\alpha_h = \mathbb{E}[Xa_h(X)]$ and $\beta_j = \mathbb{E}[Yb_j(Y)]$ and

$$\mathbb{E}[Y|x] = \sum_{h=1}^{N}\sum_{j=1}^{N} \rho_{h,j}\beta_j a_h(x), \qquad \mathbb{E}[X|y] = \sum_{h=1}^{N}\sum_{j=1}^{N} \rho_{h,j}\alpha_h b_j(y).$$

If X and Y admit NLPCs, we can use the NLPC transformations φ_h of X and ψ_j of Y, normalized with respect to their standard deviations, to construct a bivariate distribution having a given structure of dependence in terms of ϕ^2, covariance and regression functions.

We illustrate the aforesaid idea in the case $N = 1$ obtaining some densities in the class of the Sarmanov family, which is a generalization of the popular Farlie–Gumbel–Morgenstern (FGM) family (see [26], Chapter 5). This type of distributions is suited in modelling when dependence is small or moderate: the search of some extensions is motivated in increasing the maximal value of the correlation coefficient between the r.v.s involved.

For fixed (h, j) we have

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)\bigl(1 + \rho a_h(x) b_j(y)\bigr),$$

where ρ satisfies the condition that $|\rho| \le 1/\sup_{x,y}[-a_h(x)b_j(y)]$, for all x, y, and we choose $a_h(x) = (\lambda_h^X)^{-1/2}\varphi_h(x)$ and $b_j(y) = (\lambda_j^Y)^{-1/2}\psi_j(y)$ where $\lambda_h^X = \mathbb{E}[\varphi_h^2(X)]$ and $\lambda_j^Y = \mathbb{E}[\psi_j^2(Y)]$.

Note that the dependence structure of the bivariate distribution is described through just the parameter ρ, i.e. the index of linear dependence between two selected NLPC transformations. In particular we have $\phi^2 = \rho^2$ and, thanks to (16),

$$\mathbb{E}[XY] = \rho\bigl(\lambda_h^X\lambda_j^Y\bigr)^{1/2}\mathbb{E}\bigl[\varphi_h'(X)\bigr]\mathbb{E}\bigl[\psi_j'(Y)\bigr].$$

Besides the regression function is expressed in term of the NLPC transformations used: for example the regression function of Y on X is

$$\mathbb{E}[Y|x] = \rho\left(\frac{\lambda_j^Y}{\lambda_h^X}\right)^{1/2} \mathbb{E}\big[\psi_j'(Y)\big]\varphi_h(x).$$

It is proportional to $\varphi_h(x)$ and its sign depends on those of ρ and $\mathbb{E}[\psi_j'(Y)]$. The same arguments are valid for the regression function of X on Y.

In the following example we consider a family of bivariate joint distribution with uniform marginals for which the range of the correlation coefficient may be wider than the one for the FGM distribution.

Example 6.1. *Consider a bivariate distributions whose marginals are uniforms on* $(-1, 1)$ (*see the Example* 3.1). *We obtain*

$$f_{X,Y}(x, y) = \frac{1}{4} + \frac{\rho}{2}\cos\left(\frac{h\pi(x+1)}{2}\right)\cos\left(\frac{j\pi(y+1)}{2}\right), \quad x, y \in (-1, 1). \tag{25}$$

The function (25) *will be non-negative if* $|\rho| \le 1/2$. *It is easy to see that* $\phi^2 = \rho^2$, *and the covariance is equal to* $32\rho/(jh\pi^2)^2$ *for* $j = 2m + 1$ *and* $h = 2n + 1$, $m, n \in \mathbb{N}$, *and* 0 *otherwise. Denoting by* $\mathrm{Cor}(X, Y)$ *the linear correlation coefficient*, *if we take* $j = h = 1$ *we obtain* $|\mathrm{Cor}(X, Y)| \le 48/\pi^4$. *We note that this bound is appreciably wider than the one of the extension proposed in* [25] *which is the same as for the classical FGM*: $|\mathrm{Cor}(X, Y)| \le 1/3$.

Finally, the regression function of Y *on* X *is*

$$\mathbb{E}[Y|x] = -\frac{8\rho}{(j\pi)^2}\cos\left(\frac{h\pi}{2}(x+1)\right)$$

for $j = 2m + 1$, $m \in \mathbb{N}$, *and equal to zero otherwise.*

7. Technical results and proofs

This section is devoted to illustrate how to convert problem (23) into a generalized eigenvalue problem and thus to obtain an explicit expression for the estimates of NLPCs transformations by means of B-splines. The discussion is developed by two steps: first, we construct the spline approximations $\widetilde{\lambda}_{j,k}$ and $\widetilde{\varphi}_{j,k}$ of λ_j and φ_j respectively, assuming the density f_X known and we derive the rate of convergence; second, we solve the problem of estimating $\widetilde{\lambda}_{j,k}$ and $\widetilde{\varphi}_{j,k}$ from a sample, and we study the behaviour of the estimators when the sample size goes to infinity.

Definition 7.1. *Fixed integers* $k > 0$, $d \ge 2$ *and given the* $(k + d - 1)$*-dimensional linear space of spline functions with zero mean* $\widetilde{\dot{W}}{}_k^{1,2} = \dot{W}_X^{1,2} \cap \mathcal{S}_{k,d}$, *for any* j, *with* $j \le k + d - 1$, *we define* $\widetilde{\lambda}_{j,k}$ *as*

$$\widetilde{\lambda}_{j,k} = \max_{u \in \widetilde{\dot{W}}{}_k^{1,2}} \left\{\frac{\mathcal{F}[u]}{\mathcal{Q}[u]} \colon \mathbb{E}\big[u(X)\widetilde{\varphi}_{h,k}(X)\big] = 0, 1 \le h \le j - 1\right\} \tag{26}$$

and $\widetilde{\varphi}_{j,k} \in \widetilde{\dot{W}}{}_k^{1,2}$ *as the corresponding maximizer.*

To find solutions of (26) we look for the stationary values of $\mathcal{F}[u]/\mathcal{Q}[u]$ on $\widetilde{\dot{W}}{}_k^{1,2}$. In this perspective, we preliminarily consider the basis of $\widetilde{\dot{W}}{}_k^{1,2}$ constituted by the functions

$$\widetilde{B}_t(x) = \sum_{s=1}^{k+d} \frac{e_{t,s}B_s(x)}{kb_s}, \quad t = 1, \ldots, k + d - 1,$$

where the B_s are normalized B-splines forming a basis of $\mathcal{S}_{k,d}$, the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{k+d-1}$ of $\mathbb{R}^{k+d}$ satisfy $\mathbf{e}_r^{\mathrm{T}}\mathbf{e}_{r'} = \delta_{r,r'}$ and $\mathbf{e}_r^{\mathrm{T}}\mathbf{1} = 0$ for all r, r', and $b_s = \mathbb{E}[B_s(X)] > 0$. Thus, any function $u \in \tilde{\dot{W}}_k^{1,2}$ has the representation

$$u(x) = \theta^{\mathrm{T}}\mathbf{D}_k\mathbf{B}_{d,k}(x),$$

where $\theta \in \mathbb{R}^{k+d-1}$, $\mathbf{D}_k$ is a $(k+d-1) \times (k+d)$ matrix with entries $e_{t,s}/(kb_s)$ and $\mathbf{B}_{d,k}$ is the $(k+d)$-dimensional vector of the B-splines.

The functionals $\mathcal{F}$ and $\mathcal{Q}$ on $\tilde{\dot{W}}_k^{1,2}$ are respectively represented by the symmetric and positive definite matrices

$$\mathbf{F}_k = \mathbf{D}_k\mathbf{M}_k\mathbf{D}_k^{\mathrm{T}}; \qquad \mathbf{Q}_k = \mathbf{D}_k\mathbf{M}_k^{\star}\mathbf{D}_k^{\mathrm{T}},$$

where $\mathbf{M}_k$ and $\mathbf{M}_k^{\star}$ have entries (B' denotes the first derivative of a B-spline):

$$M_{s,t} = \mathbb{E}\big[B_s(X)B_t(X)\big], \qquad M_{s,t}^{\star} = \mathbb{E}\big[B_s'(X)B_t'(X)\big], \quad 1 \le s, t \le k+d.$$

The stationary values of $\mathcal{F}[u]/\mathcal{Q}[u]$ on $\tilde{\dot{W}}_k^{1,2}$ are the solutions of the equations

$$\frac{\partial\theta^{\mathrm{T}}\mathbf{F}_k\theta}{\partial\theta_t} = \ell\frac{\partial\theta^{\mathrm{T}}\mathbf{Q}_k\theta}{\partial\theta_t}, \quad 1 \le t \le k+d-1,$$

that is the solutions of the generalized eigenvalue problem $\mathbf{F}_k\theta = \ell\mathbf{Q}_k\theta$ which can be rewritten as

$$\mathfrak{M}_k\eta = \ell\eta, \tag{27}$$

where $\mathfrak{M}_k = \mathbf{Q}_k^{-1/2}\mathbf{F}_k\mathbf{Q}_k^{-1/2}$, $\eta = \mathbf{Q}_k^{1/2}\theta$, and $\mathbf{Q}_k^{1/2}$ is the symmetric square root of the positive definite matrix $\mathbf{Q}_k$. The matrix $\mathfrak{M}_k$ is symmetric, positive definite and has $k+d-1$ positive eigenvalues $\ell_{j,k}$ with corresponding orthonormal eigenvectors $\tilde{\eta}_{j,k}$. Thus the approximate solutions we look for are:

$$\tilde{\lambda}_{j,k} = \ell_{j,k}; \qquad \tilde{\varphi}_{j,k}(x) = \tilde{\theta}_{j,k}^{\mathrm{T}}\mathbf{D}_k\mathbf{B}_{d,k}(x),$$

where $\tilde{\theta}_{j,k} = \mathbf{Q}_k^{-1/2}\tilde{\eta}_{j,k}$.

In the following theorem we summarize the most important convergence results about $\tilde{\lambda}_{j,k}$ and $\tilde{\varphi}_{j,k}$ adapting the issues in [28].

Theorem 7.1. *Fixed the positive integers j and d, under conditions* (H1)–(H3) *of Theorem* 4.1:

(a) *the sequence $\{\tilde{\lambda}_{j,k}\}$ converges* (*from below*) *to λ_j and*

$$|\lambda_j - \tilde{\lambda}_{j,k}| \le ck^{-\beta} \quad \textit{for all } k \ge k_0,$$

(b) *if $\beta = m + \alpha$, there exist a positive integer k_0 and a positive constant c, depending only on j, such that*

$$\|\varphi_j - \tilde{\varphi}_{j,k}\|_0 \le ck^{-\beta} \quad \textit{for all } k \ge k_0.$$

Proof. Part (a) is a direct consequence of [28].

To prove part (b), observe that by (H2) the norm $\|\cdot\|_0$ is equivalent to the standard Lebesgue norm $\|g\|_{\star} = (\int g^2(x)\,\mathrm{d}x)^{1/2}$ and thus there exists $c_1 > 0$ such that

$$\|\varphi_j - \tilde{\varphi}_{j,k}\|_0 \le c_1\|\varphi_j - \tilde{\varphi}_{j,k}\|_{\star}.$$

Let $\overline{\varphi}_{j,k}$ be the projections of φ_j on the subspaces $\tilde{\dot{W}}_k^{1,2}$. There exists (see [28]) a constant $c_2 > 0$ and a positive integer k_0 such that

$$\|\varphi_j - \tilde{\varphi}_{j,k}\|_{\star} \le c_2\|\varphi_j - \overline{\varphi}_{j,k}\|_{\star} \quad \text{for all } k \ge k_0.$$

From Theorem XII.6 in [16] and thanks to condition (H3) the result follows. □

We discuss now the sample version of (26) in the $(k+d-1)$-dimensional space $\widehat{\widetilde{W}}_k^{1,2}$ defined above. The functions

$$\widehat{B}_t(x) = \sum_{s=1}^{k+d} \frac{e_{t,s} B_s(x)}{k\widehat{b}_{s,n}}, \quad t = 1, \dots, k+d-1, \tag{28}$$

where $\widehat{b}_{s,n} = \frac{1}{n}\sum_{i=1}^{n} B_s(X_i)$, provide a basis for it. As shown in [4], there exists $\gamma > 0$ such that we have

$$\mathbb{P}(\widehat{b}_{s,n} = 0 \text{ for some } s = 1, \dots, k+d) \le k\exp(-\gamma n/k)$$

and this probability goes to zero for $n \to +\infty$.

The functionals $\mathcal{F}_n$ and $\mathcal{Q}_n$ in (23) have the following matrix representation with respect to the basis (28):

$$\widehat{\mathbf{F}}_{k,n} = \widehat{\mathbf{D}}_{k,n}\widehat{\mathbf{M}}_{k,n}\widehat{\mathbf{D}}_{k,n}^{\mathrm{T}}; \qquad \widehat{\mathbf{Q}}_{k,n} = \widehat{\mathbf{D}}_{k,n}\widehat{\mathbf{M}}_{k,n}^{\star}\widehat{\mathbf{D}}_{k,n}^{\mathrm{T}},$$

where $\widehat{\mathbf{D}}_{k,n} = [e_{t,s}/(k\widehat{b}_{s,n})]$, $\widehat{\mathbf{M}}_{k,n} = [\widehat{M}_{s,t}]$ and $\widehat{\mathbf{M}}_{k,n}^{\star} = [\widehat{M}_{s,t}^{\star}]$ with:

$$\widehat{M}_{s,t} = \frac{1}{n}\sum_{i=1}^{n} B_s(X_i)B_t(X_i); \qquad \widehat{M}_{s,t}^{\star} = \frac{1}{n}\sum_{i=1}^{n} B_s'(X_i)B_t'(X_i).$$

Analogously to the approximation case, we have to solve the generalized eigenvalue problem $\widehat{\mathbf{F}}_{k,n}\theta = \ell\widehat{\mathbf{Q}}_{k,n}\theta$ which, if $\widehat{\mathbf{Q}}_{k,n}$ is nonsingular, is equivalent to

$$\widehat{\mathfrak{M}}_{k,n}\eta = \ell\eta, \tag{29}$$

where $\widehat{\mathfrak{M}}_{k,n} = \widehat{\mathbf{Q}}_{k,n}^{-1/2}\widehat{\mathbf{F}}_{k,n}\widehat{\mathbf{Q}}_{k,n}^{-1/2}$. In this case, the symmetric matrix $\widehat{\mathfrak{M}}_{k,n}$ has $k+d-1$ eigenvalues $\ell_{j,k,n}$ with corresponding orthonormal eigenvectors $\widehat{\eta}_{j,k,n}$. The estimates we look for are thus

$$\widehat{\lambda}_{j,k,n} = \ell_{j,k,n}; \qquad \widehat{\varphi}_{j,k,n}(x) = \widehat{\theta}_{j,k}^{\mathrm{T}}\widehat{\mathbf{D}}_k\mathbf{B}_{d,k}(x)$$

with $\widehat{\theta}_{j,k,n} = \widehat{\mathbf{Q}}_{k,n}^{-1/2}\widehat{\eta}_{j,k,n}$.

The existence and unicity of the solutions of (23) and the relevant results about the rate of convergence in probability of the sequence of the estimators to the spline approximates are stated in the following theorem.

Theorem 7.2. *Let* $\tau_n = (kn^{1-\delta})^{-1/2}$, $0 < \delta < 1$, *and consider* $k = k(n)$ *that satisfies* $k \le n^{(1-\delta)/6}$, j *and* $d \ge 2$ *fixed. Under conditions* (H1)–(H3) *of Theorem* 4.1:

(a) *a unique solution to the sample problem* (23) *exists except on an event whose probability goes to zero as* $n \to +\infty$;
(b) *when* $n \to +\infty$ *the following results hold*:

$$|\widehat{\lambda}_{j,k,n} - \widetilde{\lambda}_{j,k}| = \mathrm{o}_P\big(k^3\tau_n\big), \tag{30}$$

$$\|\widetilde{\varphi}_{j,k} - \widehat{\varphi}_{j,k,n}\|_0 = \mathrm{o}_P\big(k^3\tau_n\big). \tag{31}$$

The proof of Theorem 7.2 makes use of arguments similar to the ones in [4], and is based on asymptotics for matrices $\widehat{\mathbf{F}}_{k,n}$, $\widehat{\mathbf{Q}}_{k,n}$ and $\widehat{\mathfrak{M}}_{k,n}$, summarized in the next Lemmas 7.1 and 7.2.

For the reader's convenience, we collect here the most important properties of B-splines to which we will refer (see [16] and [32]):

(P1) $0 < B_s(x) \le 1$, for all $x \in (r_s, r_{s+d})$ and $B_s(x) = 0$ otherwise;
(P2) $\int_D B_s(x)B_t(x)\,\mathrm{d}x = 0$ for $|s-t| > d+1$ and positive otherwise;
(P3) $\int_D B_s(x)\,\mathrm{d}x = (r_{s+d}^{\star} - r_s^{\star})/d$ for all s (with $r_1^{\star}, r_2^{\star}, \dots,$ we denote the extended knots sequence);

(P4) there exists a positive constant c such that $\sup_{x\in D}|B_s'(x)| \le ck$ for all s;
(P5) for any spline function $u \in \mathcal{S}_{k,d}$, $u(x) = \alpha^{\mathrm{T}}\mathbf{B}_{d,k}(x)$, with $\alpha \in \mathbb{R}^{k+d}$, we have

$$u'(x) = \alpha^{\mathrm{T}}\boldsymbol{\Delta}^{\mathrm{T}}\mathbf{B}_{d-1,k}(x),$$

where $\mathbf{B}_{d-1,k}$ is the $(k+d-1)$-dimensional vector of B-splines belonging from the basis of $\mathcal{S}_{k,d-1}$ and $\boldsymbol{\Delta}$ is the weighted differentiation $(k+d-1)\times(k+d)$ matrix such that, when the knots are equally spaced, for $\mathbf{w} \in \mathbb{R}^{k+d}$,

$$\boldsymbol{\Delta}\mathbf{w} = (d-1)k(w_2 - w_1, w_3 - w_2, \dots, w_{k+d} - w_{k+d-1})^{\mathrm{T}}.$$

Denoting by $\|\cdot\|$ the euclidean norm on $\mathbb{R}^p$, as usual the norm $|||\mathbf{A}|||$ of a matrix $\mathbf{A}$ is defined as

$$|||\mathbf{A}||| = \max_{\|\mathbf{w}\|=1}\|\mathbf{A}\mathbf{w}\|.$$

In the following lemma we prove the convergence in probability of the "empirical matrices" $\widehat{\mathbf{F}}_{k,n}$ and $\widehat{\mathbf{Q}}_{k,n}$ to the "approximation matrices" $\mathbf{F}_k$ and $\mathbf{Q}_k$, respectively.

Lemma 7.1. *The following conclusions hold*:

$$|||\widehat{\mathbf{M}}_{k,n} - \mathbf{M}_k||| = \mathrm{o}_P(\tau_n), \tag{32}$$

$$|||\widehat{\mathbf{M}}^{\star}_{k,n} - \mathbf{M}^{\star}_k||| = \mathrm{o}_P(k^2\tau_n), \tag{33}$$

$$|||\widehat{\mathbf{F}}_{k,n} - \mathbf{F}_k||| = \mathrm{o}_P(\tau_n), \tag{34}$$

$$|||\widehat{\mathbf{Q}}_{k,n} - \mathbf{Q}_k||| = \mathrm{o}_P(k^2\tau_n). \tag{35}$$

Proof. Since

$$|||\widehat{\mathbf{M}}_{k,n} - \mathbf{M}_k||| \le \sup_{1\le s\le k+d}\sum_{t=1}^{k+d}|\widehat{M}_{s,t} - M_{s,t}|$$

and since from property (P2), for $|s-t| > d+1$ we have $\widehat{M}_{s,t} \overset{\text{a.s.}}{=} M_{s,t} = 0$, to prove (32) it is sufficient to get $|\widehat{M}_{s,t} - M_{s,t}| = \mathrm{o}_P(\tau_n)$ for any s, t, with $|s-t| \le d+1$.

Setting $Z_i = B_s(X_i)B_t(X_i)$, obviously $\mathbb{E}[Z_i] = M_{s,t}$ and, for (P1), $|Z_i| \le 1$ a.s. for all i. By (H1) the support of each B-spline is of the order of k^{-1} and by (H2) the density f_X is bounded from above. The boundedness of B-splines implies that $\mathrm{Var}(Z_i)$ vanishes for $|s-t| > d+1$ and is $\mathrm{O}(k^{-1})$ otherwise. Choosing $\varepsilon = \varepsilon_0\tau_{k,n}$, where ε_0 is a positive constant, the Chebychev inequality states that for $|s-t| \le d+1$ there exists $c > 0$ such that

$$\mathbb{P}\big(|\widehat{M}_{s,t} - M_{s,t}| \ge \varepsilon\big) \le cn^{-\delta}$$

and the relation (32) follows.

For what concerns (33), by (P5), for any $u \in \widetilde{\dot{W}}_k^{1,2}$, we have

$$u'(x) = \theta^{\mathrm{T}}\mathbf{D}_k\boldsymbol{\Delta}^{\mathrm{T}}\mathbf{B}_{d-1,k}(x).$$

It follows that $\mathbf{M}^{\star}_k = \boldsymbol{\Delta}^{\mathrm{T}}\mathbf{M}_{d-1,k}\boldsymbol{\Delta}$ and $\widehat{\mathbf{M}}^{\star}_{k,n} = \boldsymbol{\Delta}^{\mathrm{T}}\widehat{\mathbf{M}}_{d-1,k,n}\boldsymbol{\Delta}$ where $\mathbf{M}_{d-1,k} = \mathbb{E}[\mathbf{B}_{d-1,k}\mathbf{B}^{\mathrm{T}}_{d-1,k}]$ and $\widehat{\mathbf{M}}_{d-1,k}$ is its empirical version. Denoting $\zeta = \max(|||\boldsymbol{\Delta}|||, |||\boldsymbol{\Delta}^{\mathrm{T}}|||)$, we obtain

$$\big|||\widehat{\mathbf{M}}^{\star}_{k,n} - \mathbf{M}^{\star}_k\big||| \le \zeta^2|||\widehat{\mathbf{M}}_{d-1,k,n} - \mathbf{M}_{d-1,k}|||.$$

By (32) we get

$$|||\widehat{\mathbf{M}}_{d-1,k,n} - \mathbf{M}_{d-1,k}||| = \mathrm{o}_P(\tau_{k,n})$$

and, since by condition (H1) it is $\zeta = \mathrm{O}(k)$, relation (33) follows.

To derive (34) we use the decomposition

$$\widehat{\mathbf{F}}_{k,n} - \mathbf{F}_k = (\widehat{\mathbf{D}}_{k,n} - \mathbf{D}_k)\widehat{\mathbf{M}}_{k,n}\widehat{\mathbf{D}}_{k,n}^{\mathrm{T}} + \widehat{\mathbf{D}}_{k,n}(\widehat{\mathbf{M}}_{k,n} - \mathbf{M}_k)\mathbf{D}_k^{\mathrm{T}} + \mathbf{D}_k\mathbf{M}_k\big(\widehat{\mathbf{D}}_{k,n}^{\mathrm{T}} - \mathbf{D}_k^{\mathrm{T}}\big)$$

which allows to bound from above $|||\widehat{\mathbf{F}}_{k,n} - \mathbf{F}_k|||$ by

$$\begin{aligned} &|||\widehat{\mathbf{D}}_{k,n} - \mathbf{D}_k|||\,|||\widehat{\mathbf{M}}_{k,n}|||\,|||\widehat{\mathbf{D}}_{k,n}||| + |||\widehat{\mathbf{D}}_{k,n}|||\,|||\widehat{\mathbf{M}}_{k,n} - \mathbf{M}_k|||\,|||\mathbf{D}_k||| \\ &\quad + |||\mathbf{D}_k|||\,|||\mathbf{M}_k|||\,|||\widehat{\mathbf{D}}_{k,n} - \mathbf{D}_k|||. \end{aligned} \tag{36}$$

Thanks to (H1), (H2) and property (P3), for $s = 1, \ldots, k+d-1$, there exists a positive constant c such that $kb_s \geq c$ and so $|||\mathbf{D}_k||| = \mathrm{O}(1)$.

Since $|||\widehat{\mathbf{D}}_{k,n} - \mathbf{D}_k||| = \mathrm{o}_P(k\tau_n)$ (see [4]), it follows $|||\widehat{\mathbf{D}}_k||| = \mathrm{O}_P(1)$. Besides (H1), (H2) and (P1) imply that $|||\mathbf{M}_k||| = \mathrm{O}(k^{-1})$ and, thanks to (32), we get $|||\widehat{\mathbf{M}}_k||| = \mathrm{O}_P(k^{-1})$. So, we conclude that all the terms in (36) are $\mathrm{o}_P(\tau_n)$ and (34) follows.

The result (35) is obtained by observing that, operating similarly to the proof of (34), it holds

$$\begin{aligned} |||\widehat{\mathbf{Q}}_{k,n} - \mathbf{Q}_k||| \leq{} & |||\widehat{\mathbf{D}}_{k,n} - \mathbf{D}_k|||\,\big|\big|\big|\widehat{\mathbf{M}}^{\star}_{k,n}\big|\big|\big|\,|||\widehat{\mathbf{D}}_{k,n}||| + |||\widehat{\mathbf{D}}_{k,n}|||\,\big|\big|\big|\widehat{\mathbf{M}}^{\star}_{k,n} - \mathbf{M}^{\star}_k\big|\big|\big|\,|||\mathbf{D}_k||| \\ & + |||\mathbf{D}_k|||\,\big|\big|\big|\mathbf{M}^{\star}_k\big|\big|\big|\,|||\widehat{\mathbf{D}}_{k,n} - \mathbf{D}_k|||. \end{aligned} \tag{37}$$

Then, from conditions (H1), (H2) and property (P4), we get $|||\mathbf{M}^{\star}_k||| = \mathrm{O}(k)$ and, thanks to (33), $|||\widehat{\mathbf{M}}^{\star}_k||| = \mathrm{O}_P(k)$. It follows that all the terms in (37) are $\mathrm{o}_P(k^2\tau_{k,n})$ and this concludes the proof. □

Note that, by (35), the probability that $\widehat{\mathbf{Q}}_{k,n}$ is definite positive tends to one when n tends to infinity and in this sense $\widehat{\mathbf{Q}}_{k,n}^{-1/2}$ exists.

In the next lemma we show the convergence of the "empirical matrix" $\widehat{\mathfrak{M}}_{k,n}$ to the "population approximation matrix" $\mathfrak{M}_k$ involved in the eigenvalues problems (29) and (27), respectively.

Lemma 7.2. *The following conclusions hold*:

$$\big|\big|\big|\widehat{\mathbf{Q}}_{k,n}^{-1/2} - \mathbf{Q}_k^{-1/2}\big|\big|\big| = \mathrm{o}_P\big(k^{7/2}\tau_n\big), \tag{38}$$

$$|||\widehat{\mathfrak{M}}_{k,n} - \mathfrak{M}_k||| = \mathrm{o}_P\big(k^3\tau_n\big). \tag{39}$$

Proof. Observe preliminarily that matrix $\mathbf{Q}_k = \mathbf{D}_k\mathbf{M}^{\star}_{d,k}\mathbf{D}_k^{\mathrm{T}}$ is positive definite and consider $\mathbf{w}^{\mathrm{T}}\mathbf{Q}_k\mathbf{w}$ where $\mathbf{w} \in R^{k+d-1}$ with $\|\mathbf{w}\|^2 = 1$. Let $\mathcal{K} = \{\alpha \in \mathbb{R}^{k+d} : \mathbf{M}^{\star}_{d,k}\alpha = 0\}$. From Lemma 5.2 in [11] and (H1), it follows that there exist two positive constants c_1 and c_2 such that $c_1k^{-1}\|\alpha\|^2 \leq \alpha^{\mathrm{T}}\mathbf{M}^{\star}_{d,k}\alpha$, for any $\alpha \in \mathcal{K}^{\perp}$, and $\alpha^{\mathrm{T}}\mathbf{M}^{\star}_{d,k}\alpha \leq c_2k\|\alpha\|^2$, for any $\alpha \in \mathbb{R}^{k+d}$. Because, from (H1), (H2) and (P3), $|||\mathbf{D}_k|||^2$ is bounded by positive constants, then there are $0 < c_3 < c_4$ such that all the eigenvalues of $\mathbf{Q}_k$ lie between c_3k^{-1} and c_4k.

Let ϑ be the minimum of the smallest eigenvalues of $\widehat{\mathbf{Q}}_{k,n}$ and $\mathbf{Q}_k$. There exists (see [4]) a positive constant c_5 such that:

$$\big|\big|\big|\widehat{\mathbf{Q}}_{k,n}^{-1/2} - \mathbf{Q}_k^{-1/2}\big|\big|\big| \leq c_5\vartheta^{-3/2}|||\widehat{\mathbf{Q}}_{k,n} - \mathbf{Q}_k|||. \tag{40}$$

By the preliminary remark, the smallest eigenvalue of $\mathbf{Q}_k$ is not smaller than c_3k^{-1}; furthermore, by (35) we obtain $\vartheta^{-1} = \mathrm{O}_P(k)$. Finally, combining (35) and (40) the (38) follows.

About (39), using the decomposition

$$\begin{aligned} \widehat{\mathfrak{M}}_{k,n} - \mathfrak{M}_k = {}& \big(\widehat{\mathbf{Q}}_{k,n}^{-1/2} - \mathbf{Q}_k^{-1/2}\big)\widehat{\mathbf{F}}_{k,n}\widehat{\mathbf{Q}}_{k,n}^{-1/2} \\ & + \mathbf{Q}_{k,n}^{-1/2}(\widehat{\mathbf{F}}_{k,n} - \mathbf{F}_k)\widehat{\mathbf{Q}}_{k,n}^{-1/2} + \mathbf{Q}_k^{-1/2}\mathbf{F}_k\big(\widehat{\mathbf{Q}}_{k,n}^{-1/2} - \mathbf{Q}_k^{-1/2}\big) \end{aligned}$$

we can bound from above $|||\widehat{\mathfrak{M}}_{k,n} - \mathfrak{M}_k|||$ by

$$\left|\left|\left|\widehat{\mathbf{Q}}_{k,n}^{-1/2} - \mathbf{Q}_k^{-1/2}\right|\right|\right| |||\widehat{\mathbf{F}}_{k,n}||| \left|\left|\left|\widehat{\mathbf{Q}}_{k,n}^{-1/2}\right|\right|\right| + |||\mathbf{Q}_k^{-1/2}||| \, |||\widehat{\mathbf{F}}_{k,n} - \mathbf{F}_k||| \left|\left|\left|\widehat{\mathbf{Q}}_{k,n}^{-1/2}\right|\right|\right| + \left|\left|\left|\mathbf{Q}_k^{-1/2}\right|\right|\right| |||\mathbf{F}_k||| \left|\left|\left|\widehat{\mathbf{Q}}_{k,n}^{-1/2} - \mathbf{Q}_k^{-1/2}\right|\right|\right|. \tag{41}$$

Note that conditions (H1), (H2) and properties (P1) and (P3) imply that $|||\mathbf{F}_k||| = \mathrm{O}(k^{-1})$ and, by Lemma 7.1, $|||\widehat{\mathbf{F}}_{k,n}||| = \mathrm{O}_P(k^{-1})$. Besides, $|||\mathbf{Q}_k^{-1/2}||| = \mathrm{O}(k^{1/2})$ and from (38) we have $|||\widehat{\mathbf{Q}}_{k,n}^{-1/2}||| = \mathrm{O}_P(k^{1/2})$. Thanks to these results and to relations (38) and (34), we obtain that the first and third term in (41) are $\mathrm{o}_P(k^3\tau_n)$ and the second one is $\mathrm{o}_P(k\tau_n)$; thus (39) follows. □

We are now ready to prove Theorem 7.2.

Proof of Theorem 7.2. (a) Since, by Lemma 7.1, the matrix $\widehat{\mathbf{Q}}_{k,n}$ is nonsingular except on a event whose probability tends to zero as $n \to +\infty$, there exists a solution for (23) whose unicity is guaranteed with probability that goes to one as $n \to +\infty$, thanks to the result proved in [27].

(b) The result (30) follows directly from (39) since, for all j,

$$|\widehat{\lambda}_{j,k,n} - \widetilde{\lambda}_{j,k}| \le |||\widehat{\mathfrak{M}}_{k,n} - \mathfrak{M}_k|||.$$

Now we prove (31). Recalling that

$$\widehat{\varphi}_{j,k,n}(x) = \widehat{\eta}_{j,k,n}^{\mathrm{T}} \widehat{\mathbf{Q}}_{k,n}^{-1/2} \widehat{\mathbf{D}}_k \mathbf{B}_{d,k}(x),$$
$$\widetilde{\varphi}_{j,k,n}(x) = \widetilde{\eta}_{j,k,n}^{\mathrm{T}} \mathbf{Q}_{k,n}^{-1/2} \mathbf{D}_k \mathbf{B}_{d,k}(x),$$

we obtain the following majorization

$$\begin{aligned} \|\widehat{\varphi}_{j,k,n} - \widetilde{\varphi}_{j,k}\|_0 \le{} & \|\widetilde{\eta}_{j,k,n}\| \left|\left|\left|\mathbf{Q}_{k,n}^{-1/2} - \widehat{\mathbf{Q}}_{k,n}^{-1/2}\right|\right|\right| |||\mathbf{D}_k||| \, |||\mathbf{M}_k|||^{1/2} \\ & + \|\widehat{\eta}_{j,k,n} - \widetilde{\eta}_{j,k,n}\| \left|\left|\left|\widehat{\mathbf{Q}}_{k,n}^{-1/2}\right|\right|\right| |||\widehat{\mathbf{D}}_k||| \, |||\mathbf{M}_k|||^{1/2} \\ & + \|\widetilde{\eta}_{j,k,n}\| \left|\left|\left|\widehat{\mathbf{Q}}_{k,n}^{-1/2}\right|\right|\right| |||\mathbf{D}_k - \widehat{\mathbf{D}}_k||| \, |||\mathbf{M}_k|||^{1/2}. \end{aligned}$$

By the Lemma 3.1 in [3], we have

$$\|\widehat{\eta}_{j,k,n} - \widetilde{\eta}_{j,k,n}\| \le \alpha_j |||\widehat{\mathfrak{M}}_{k,n} - \mathfrak{M}_k|||,$$

where

$$\begin{cases} \alpha_1 = \frac{2\sqrt{2}}{\widetilde{\lambda}_{1,k} - \widetilde{\lambda}_{2,k}}, \\ \alpha_j = \frac{2\sqrt{2}}{\min(\widetilde{\lambda}_{j-1,k} - \widetilde{\lambda}_{j,k}, \widetilde{\lambda}_{j,k} - \widetilde{\lambda}_{j+1,k})}, \quad 1 < j \le k + q - 1. \end{cases}$$

Thanks to the part (a) of Theorem 7.1, Lemma 7.2, and because $|||\mathbf{M}_k|||^{1/2} = \mathrm{O}(k^{-1/2})$, $|||\widehat{\mathbf{Q}}_{k,n}^{-1/2}||| = \mathrm{O}(k^{1/2})$, $|||\mathbf{D}_k||| = \mathrm{O}(1)$, $|||\widehat{\mathbf{D}}_k||| = \mathrm{O}_P(1)$, $|||\mathbf{D}_k - \widehat{\mathbf{D}}_k||| = \mathrm{o}_P(k\tau_n)$ and $\|\widehat{\eta}_{j,k,n}\| \stackrel{\text{a.s.}}{=} \|\widetilde{\eta}_{j,k}\| = 1$, the result (31) follows. □

Acknowledgment

We are grateful to Pascal Sarda for several fruitful discussions during our visit to the working group STAPH of the Université Paul Sabatier in Toulouse. We would like also to express our thanks to Elvise Berchio, Silvano Fiorin and

Franco Tomarelli. Finally, we wish to thank an anonymous referee for his helpful comments and suggestions which have led to substantial improvement in the presentation of this paper. In any case, every mistake is ascribed entirely only to the authors.

References

[1] F. Antoci. Some necessary and some sufficient conditions for the compactness of the embedding of weighted Sobolev spaces. *Ric. Mat.* **LII** (2003) 55–71. MR2091081
[2] G. Arfken. *Mathematical Methods for Physicists*. Academic Press, New York, 1966. MR0205512
[3] D. Bosq. Modelization, nonparametric estimation and prediction for continuous time process. In *Nonparametric Functional Estimation and Related Topics* 509–529. G. Roussas, (Ed.). *Nato, Asi Series*. Kluwer Academic, Dordrecht, 1991. MR1154349
[4] P. Burman. Rates of convergence for the estimate of the optimal transformations of variables. *Ann. Statist.* **19** (1991) 702–723. MR1105840
[5] P. Burman and K. W. Chen. Nonparametric estimation of a regression function. *Ann. Statist.* **17** (1989) 1567–1596. MR1026300
[6] G. Buttazzo, M. Giaquinta and S. Hildebrandt. *One-Dimensional Variational Problems*. *Oxford Lecture Series in Mathematics and Its Applications* **15**. Clarendon Press, New York, 1998. MR1694383
[7] T. Cacoullos. On upper and lower-bounds for the variance of a function of a random variable. *Ann. Probab.* **10** (1982) 799–809. MR0659549
[8] T. Cacoullos and V. Papathanasiou. On upper bounds for the variance of functions of random variables. *Statist. Probab. Lett.* **3** (1985) 175–184. MR0801687
[9] T. Cacoullos and V. Papathanasiou. Characterizations of distributions by variance bounds. *Statist. Probab. Lett.* **7** (1989) 351–356. MR1001133
[10] T. Cacoullos and V. Papathanasiou. Characterization of distributions by generalizations of variance bounds and simple proofs of the CLT. *J. Statist. Plann. Inference* **63** (1997) 157–171. MR1491576
[11] H. Cardot. Spatially adaptive splines for statistical linear inverse problems. *J. Multivariate Anal.* **81** (2002) 100–119. MR1901208
[12] L. H. Y. Chen and J. H. Lou. Characterization of probability distributions by Poincaré-type inequalities. *Ann. Inst. H. Poincaré Probab. Statist* **23** (1987) 91–110. MR0877386
[13] H. Chernoff. A note on an inequality involving the normal distribution. *Ann. Probab.* **9** (1981) 533–535. MR0614640
[14] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. Wiley, New York, 1989. MR1013360
[15] J. Dauxois, A. Pousse and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** (1982) 136–154. MR0650934
[16] C. De Boor. *A Practical Guide to Splines*. Springer, New York, 2001. MR1900298
[17] N. E. El Faouzi and P. Sarda. Rates of convergence for spline estimates of additive principal components. *J. Multivariate Anal.* **68** (1999) 120–137. MR1668907
[18] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Prentice-Hall, New Jersey, 1963. MR0160139
[19] P. Gurka and B. Opic. Continuous and compact imbeddings of weighted Sobolev Spaces II. *Czechoslovak Math. J.* **39** (1989) 78–94. MR0983485
[20] C. A. J. Klaassen. On an inequality of Chernoff. *Ann. Probab.* **13** (1985) 966–974. MR0799431
[21] A. Kufner and B. Opic. How to define reasonably weighted Sobolev Spaces. *Comment. Math. Univ. Carolin.* **25** (1984) 537–554. MR0775568
[22] O. Johnson and A. Barron. Fisher information inequalities and the central limit theorem, *Probab. Theory Related Fields* **129** (2004) 391–409. MR2128239
[23] I. T. Jolliffe. *Principal Component Analysis*. Springer, Berlin, 2004. MR0841268
[24] H. O. Lancaster. *The Chi-Squared Distribution*. Wiley, New York, 1969. MR0253452
[25] M.-L. T. Lee. Properties and applications of the Samarov family of bivariate distributions. *Comm. Statist. Theory Methods* **25** (1996) 1207–1222. MR1394279
[26] D. D. Mari and S. Kotz. *Correlation and Dependence*. Imperial College Press, London, 2001. MR1835042
[27] M. Okamoto. Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann. Statist.* **1** (1973) 763–765. MR0331643
[28] J. G. Pierce and R. S. Varga. Higher order convergence results for the Rayleigh–Ritz method applied to eigenvalue problems. I: Estimates relating Rayleigh–Ritz and Galerkin approximations to eigenfunctions. *SIAM J. Numer. Anal.* **9** (1972) 137–151. MR0395268
[29] S. Purkayastha and S. K. Bhandari. Characterization of uniform distributions by inequality of Chernoff-type, *Sankhyā* **52** (1990) 376–382. MR1178045
[30] E. Salinelli. Nonlinear principal components I. Absolutely continuous random variables with positive bounded densities. *Ann. Statist.* **26** (1998) 596–616. MR1626079
[31] E. Salinelli. Nonlinear principal components II. Characterization of normal distributions. *J. Multivariate Anal.* **100** (2009) 652–660. MR2478188
[32] L. L. Schumaker. *Spline Functions: Basic Theory*. Wiley, New York, 1981. MR0606200
[33] C. J. Stone. Optimal global rate of convergence for nonparametric regression. *Ann. Statist.* **10** (1982) 1040–1053. MR0673642
[34] A. Zettl. *Sturm–Liouville Theory. Mathematical Survey and Monographs* **121**. Amer. Math. Soc., Providence, RI, 2005. MR2170950