The Annals of Statistics
1991, Vol. 19, No. 2, 702–723

RATES OF CONVERGENCE FOR THE ESTIMATES OF THE OPTIMAL TRANSFORMATIONS OF VARIABLES

BY PRABIR BURMAN

University of California, Davis

We consider here spline estimates of the optimal transformations of variables for multiple correlation and regression as dealt with in a recent paper by Breiman and Friedman. We show that we can construct estimates of the optimal transformations which have the same optimal rate of convergence as in the usual nonparametric estimation of a univariate function.

1. Introduction. Ever since the publication of the paper by Box and Cox (1964), there has been a considerable interest in data transformation. Recently, Breiman and Friedman (1985) have substantially generalized the idea of Box and Cox and provided a very powerful tool for data analysis. Box and Cox considered only power transformations, but Breiman and Friedman considered arbitrary transformations and provided an elegant method called the alternating conditional expectation (ACE) to calculate such transformations. When the random variables are discrete, there is a methodology closely related to that of Breiman and Friedman and it is known in the literature as Optimal Scaling [see Greenacre (1984)].

Let $(Y, X_1, \dots, X_d)$ be a $(d+1)$-dimensional vector of univariate random variables. Breiman and Friedman considered transformations $h(Y)$, $\phi_1(X_1), \dots, \phi_d(X_d)$ so that these transformed variables have zero means, $Eh^2(Y) = 1$, $E\phi_j^2(X_j) < \infty$, $j = 1, \dots, d$. Let $e^2(h, \tilde{\phi}) = E(h(Y) - \tilde{\phi}(\mathbf{X}))^2$, where $\tilde{\phi}(\mathbf{x}) = \phi_1(x_1) + \cdots + \phi_d(x_d)$, be the unexplained error for regression of $h(Y)$ on $\tilde{\phi}(\mathbf{X})$. Transformations $(h^*, \tilde{\phi}^*)$ are called optimal if they minimize e^2. Breiman and Friedman showed that under certain conditions, optimal transformations exist though not necessarily uniquely.

In this paper, we address the question of rates of convergence. We consider spline approximations to $h, \phi_1, \dots, \phi_d$, and the problem of estimating these splines from a sample of size n. The ACE algorithm requires iterative computations, whereas in our case the problem of estimating the optimal transformations reduces to a matrix eigenvalue problem. We show that the empirical estimates $\hat{h}, \hat{\phi}_1, \dots, \hat{\phi}_d$, exist and except on an event whose probability approaches zero as $n \to \infty$, these estimates are unique. By choosing the degree of the splines and the number of knots properly, we can show that

$$\inf\{\|\hat{h} - h^*\|: h^* \text{ is an optimal transformation}\}$$

Received September 1986; revised March 1990.
AMS 1980 *subject classifications*. Primary 62G20; secondary 62G05.
Key words and phrases. B-splines, compact operators, optimal transformations of variables, perturbation theory, rate of convergence.

and

$$\inf\left\{\left\|\hat{\phi}_j - \phi_j^*\right\|: \phi_j^* \text{ is an optimal transformation}\right\}, \qquad j = 1, \dots, d,$$

have the same optimal rate of convergence as in the usual nonparametric estimation of a univariate function (e.g, a density). Here $\| \; \|$ refers to the L^2 distance with respect to the joint density of $(Y, X_1, \dots, X_d)$. The arguments of this paper could be easily adapted to the case of orthogonal series estimators.

We would like to mention here a related work by Koyak (1990). He has recently shown that the sieve method provides consistent estimates of the transformations.

The organization of the paper is as follows. In Section 2, we present some basic results which are borrowed from the paper by Breiman and Friedman. We discuss spline estimates and present the main result on the rates of convergence in Section 3. In Section 4, we present a few technical results. We prove all our technical results in Sections 5 and 6b. In Section 6a, we write down a matrix version of the problem of estimating the optimal transformations. We would like to point out here that we use some perturbation theory of linear operators in our proofs.

2. Some basic results. In this section, we will write down some definitions and assumptions which we will need later on. Let $(Y, X_1, \dots, X_d)$ be a $(d+1)$-dimensional vector of univariate random variables on a compact set (on R^{d+1}) which we will assume to be $[0,1]^{d+1}$ without loss of generality (see, however, Assumption 4 in Section 3). Let $h, \phi_1, \dots, \phi_d$, be any transformations of $Y, X_1, \dots, X_d$ satisfying

$$(2.1)\qquad \begin{aligned} &Eh(Y) = 0, \qquad E\phi_j(X_j) = 0, \qquad j = 1, \dots, d,\\ &Eh^2(Y) < \infty, \qquad E\phi_j^2(X_j) < \infty, \qquad j = 1, \dots, d, \end{aligned}$$

Let $\tilde{\phi}(\mathbf{X}) = \sum_1^d \phi_j(X_j)$ and $e^2(h, \tilde{\phi}) = E[h(Y) - \tilde{\phi}(\mathbf{X})]^2$, where $h, \phi_1, \dots, \phi_d$, satisfy (2.1) and $\|h\| = 1$. $(h^*, \tilde{\phi}^*)$ are called optimal if they minimize e^2. Breiman and Friedman proved that optimal transformations exist under certain conditions, but are not necessarily unique. The results presented in this section are taken from Sections 5.2 and 5.3 of Breiman and Friedman's paper.

DEFINITION 2.1. (a) Let H^0 be the linear space of functions f of the form $f(y, \mathbf{x}) = h(y) + \tilde{\phi}(\mathbf{x})$, where h and $\tilde{\phi}$ are as in (2.1), with the norm and inner product defined as

$$\|f\|^2 = Ef^2, \qquad \langle f_1, f_2 \rangle = Ef_1 f_2.$$

(b) Let H_Y and H_{X_j}, $j = 1, \dots, d$, be the usual subspaces of H^0 with the same norm and inner product,

$$h \in H_Y \quad \text{if } Eh(Y) = 0 \text{ and } Eh^2(Y) < \infty,$$

$$\phi_j \in H_{X_j} \quad \text{if } E\phi_j(X_j) = 0 \text{ and } E\phi_j^2(X_j) < \infty.$$

(c) Let $H_{\mathbf{X}}$ be the space of functions of the form $\tilde{\phi}(\mathbf{x}) = \sum_1^d \phi_j(x_j)$, $\phi_j \in H_{X_j}$.

It is easy to see that H_Y and H_{X_j}, $j = 1, \dots, d$, are Hilbert spaces. Breiman and Friedman showed that under Assumptions 1 and 2, H^0 is a Hilbert space and H_Y, $H_{\mathbf{X}}$ and H_{X_j}, $j = 1, \dots, d$, are closed subspaces.

Throughout the paper we will make the following two crucial assumptions which guarantee the existence of the optimal transformations.

ASSUMPTION 1. If the transformations $h, \phi_1, \dots, \phi_d$ with zero means and finite variances satisfy

$$h(Y) + \sum_1^d \phi_j(X_j) = 0 \quad \text{a.e., then each of them is zero a.e.}$$

ASSUMPTION 2. The conditional expectation operators

$$E\big(\phi_j(X_j)\big|Y\big)\colon H_{X_j} \to H_Y,$$
$$E\big(\phi_j(X_j)\big|X_i\big)\colon H_{X_j} \to H_{X_i}, \qquad i \neq j,$$
$$E\big(h(Y)\big|X_j\big)\colon H_Y \to H_{X_j},$$

are all compact operators.

It can be shown that a sufficient condition for Assumption 2 to hold is that $\int f_{\mathbf{X}Y}^2/(f_{\mathbf{X}} f_Y) < \infty$, where $f_{\mathbf{X}}$ is the joint density of $X_1, \dots, X_d$, $f_{\mathbf{X}Y}$ is the joint density of $\mathbf{X}$ and Y and f_Y is the marginal density of Y.

THEOREM 2.2. *Under Assumptions* 1 *and* 2, *optimal transformations exist.*

Now we will characterize the optimal transformations. We will see that the optimal transformations are the eigenfunctions corresponding to the largest eigenvalues of certain operators.

DEFINITION 2.3. In H^0, let P_Y and $P_{\mathbf{X}}$ be the orthogonal projections on the subspaces H_Y and $H_{\mathbf{X}}$, respectively. Let $U = P_Y P_{\mathbf{X}}$ and $V = P_{\mathbf{X}} P_Y$.

Breiman and Friedman showed that U and V are compact, self-adjoint and nonnegative definite. Let λ_1 be the largest eigenvalue of U and let M be the corresponding eigenspace. It is easy to see that

$$\inf\{e^2(h, \tilde{\phi})\colon h \in H_Y, \tilde{\phi} \in H_{\mathbf{X}}, \|h\| = 1\}$$
$$= 1 - \sup\{\langle h, Uh \rangle\colon h \in H_Y, \|h\| = 1\} = 1 - \lambda_1.$$

So if $h \in M$, $\|h\| = 1$, then $(h, \tilde{\phi})$, $\tilde{\phi} = P_{\mathbf{X}} h$, are optimal transformations. Indeed, all the optimal transformations are of this form. Let us note that $\lambda_1 < 1$, by Assumption 1. Following Breiman and Friedman, we will assume throughout that $\lambda_1 > 0$. The case $\lambda_1 = 0$ is a trivial one, because then for any h in H_Y with $\|h\| = 1$, $(h, P_{\mathbf{X}} h)$ would be optimal. However, the behaviour of the sample estimate of λ_1 in that case is still an open question. Since U is compact and $\lambda_1 > 0$, $m = \dim(M) < \infty$. This tells us that there are exactly m linearly independent optimal transformations.

3. Spline estimates of the optimal transformations and rates of convergence. Since we will discuss the rate of convergence, we will first set some smoothness conditions on the optimal transformations. Let us recall that the number of linearly independent optimal transformations $(h^*, \tilde{\phi}^*)$ is finite, which we assume to be m, and M is the Hilbert space generated by the h^*'s, the optimal transformations of the Y variable. Let N be generated by the optimal $\tilde{\phi}^*$'s. Let us note that $\dim(M) = \dim(N) = m$.

ASSUMPTION 3. (a) If $h \in M$, $\|h\| = 1$, then h is p' times continuously differentiable and $|h^{(p')}(y_1) - h^{(p)}(y_2)| \le c|y_1 - y_2|^{\nu}$ for all y_1 and y_2, for some $0 < \nu \le 1$ and $c > 0$.

(b) If $\tilde{\phi} \in N$ and $\|\tilde{\phi}\| = 1$, then each ϕ_j is p' times continuously differentiable and $|\phi_j^{(p')}(x_1) - \phi_j^{(p')}(x_2)| \le c|x_1 - x_2|^{\nu}$ for all x_1 and x_2.

We would like to point out that the constants c and ν in Assumption 3 do not depend on h and $\tilde{\phi}$. Let us note that if m basis elements of M and N satisfy Assumption 3, then every element of M and N will satisfy Assumption 3. It is possible to find simple sufficient conditions which imply Assumption 3 [see Remark (c) at the end of this section].

Let $p = p' + \nu$ and we will assume that $p > 0$. Let q be any integer larger than or equal to p. We will seek spline estimates of the optimal transformations. More specifically, we will minimize e^2 over the class of spline functions $(h, \tilde{\phi})$. To make everything formal, let us first define spline functions. h is a spline function on $[0, 1]$ of degree q with k knots if (a) h is a polynomial of degree q on each interval $[(t-1)k^{-1}, tk^{-1}]$, $t = 1, \dots, k$, (b) h is $(q-1)$ times continuously differentiable on $[0, 1]$.

Let S_{qk} be the class of all such splines on $[0, 1]$ satisfying (a) and (b). It is well known that the dimension of S_{qk} is $q + k$ and there exists a basis of S_{qk} consisting of normalized B-splines $\{B_{kj}, j = 1, \dots, q + k\}$ [see de Boor (1978)].

In order to prove the main result of this paper we need the following additional assumption.

ASSUMPTION 4. The joint density of $Y, X_1, \dots, X_d$ is bounded and the marginal densities of $Y, X_1, \dots, X_d$ are bounded away from zero.

Now let us define our spline estimates. Suppose we have n observations $(Y_i, \mathbf{X}_i)$, $i = 1, \dots, n$. We want to estimate optimal transformations from the data. Let

$$e_{kn}^2(h, \tilde{\phi}) = n^{-1} \sum_{t=1}^{n} \left(h(Y_t) - \sum_{j=1}^{d} \phi_j(X_{jt}) \right)^2, \tag{3.1}$$

where $\tilde{\phi} = \phi_1 + \cdots + \phi_d$ and $h, \phi_1, \ldots, \phi_d$, are in S_{qk} with the constraints

$$(3.2)\quad \begin{aligned} &n^{-1} \sum_{t=1}^{n} h^2(Y_t) = 1, \qquad n^{-1} \sum_{t=1}^{n} h(Y_t) = 0, \\ &n^{-1} \sum_{t=1}^{n} \phi_j(X_{jt}) = 0, \qquad j = 1, \ldots, d. \end{aligned}$$

(h, ϕ) are estimates of the optimal transformations if they are a solution to the following minimization problem:

$$(3.3)\quad e_{kn}^{*2} = \inf\left\{n^{-1} \sum \left(h(Y_t) - \tilde{\phi}(\mathbf{X}_t)\right)^2 : h \text{ and } \tilde{\phi} \text{ satisfy } (3.2)\right\}$$

Let us note that the minimization problem in (3.3) is a finite dimensional problem and could be converted into a problem of finding the eigenvector corresponding to the largest eigenvalue of a matrix [see Section 6a]. Well-known computer packages are available for solving such a matrix problem.

Now we state the main result of this paper, which is an immediate consequence of Theorems 4.1 and 4.2 given in the next section. We will write $k \sim n^{\gamma}$, $\gamma > 0$, to mean that $c'n^{\gamma} \le k \le c''n^{\gamma}$ for positive constants c' and c''.

THEOREM 3.1. (a) *Let $k \le n^{1-\delta}$ for some $\delta > 0$.*

A solution $(\hat{h}_k, \hat{\tilde{\phi}}_k)$ to the optimization problem in (3.3) *exists and except on an event whose probability goes to zero as $n \to \infty$, this solution is unique.*

(b) *Let $k \sim n^{1/(2p+1)}$ and $r = p/(2p+1)$. Let $(\hat{h}_k, \hat{\tilde{\phi}}_k)$ be a solution of* (3.3), *where $\hat{\tilde{\phi}}_k = \hat{\phi}_{k1} + \cdots + \hat{\phi}_{kd}$. Then*

$$\inf\{\|\hat{h}_k - h^*\| : h^* \textit{ is an optimal transformation for variable } Y\} = O_P(n^{-r})$$

and

$$\inf\left\{\|\hat{\phi}_{kj} - \phi_j^*\| : \phi_j \textit{ is an optimal transformation variable } X_j\right\} = O_P(n^{-r}),$$
$$j = 1, \ldots, d.$$

(c) *Let $e^{*2} = \inf\{\|h - \tilde{\phi}\|^2$: where h and $\tilde{\phi}$ satisfy* (2.1) *and $\|h\| = 1\}$. Then*

$$e_{kn}^{*2} = e^{*2} + O_P(n^{-r}).$$

The implication of Theorem 3.1 is that we can construct estimates of the transformations which achieve the optimal rate of convergence associated with the estimation of univariate functions. To clarify this point, let us consider the following example. Suppose we are interested in estimating the regression function $\mu(x) = E(Y|X = x)$ from a sample of size n, where Y and X are univariate random variables. As in Assumption 3, let us assume that μ is p' times differentiable and the p'th derivative satisfies $|\mu^{(p')}(x_1) - \mu^{(p')}(x_2)| \le c|x_1 - x_2|^{\nu}$ for all x_1 and x_2 for some $c > 0$ and $0 < \nu \le 1$. Let $r = p/(2p+1)$, where $p = p' + \nu$. Stone (1982) showed that no estimate $\hat{\mu}$ of μ can achieve a rate better than n^{-r}, when the distance between $\hat{\mu}$ and μ is assumed to be the

L^2 distance with respect to the marginal distribution of X. [See Stone (1982) for details.]

In our case, the optimal transformations $(h^*, \tilde{\phi}^*)$ are not unique. Still, we can show that for each variable, we can construct an estimate of the transformation which is close to one of the optimal transformations with the optimal rate n^{-r}.

REMARKS. (a) It may be desirable in practice to place the knots at the sample quantiles. For variable Y, the knots could be placed at the (j/k)th sample quantiles, $j = 1, \dots, k$. For each X variable, the knots could be similarly placed at the corresponding sample quantiles. We believe that all the results of this paper would remain true if we use spline estimates with the type of variable knots described here, however, the proofs would be far more complicated.

(b) The smoothness properties of the optimal transformations are not known in practice. Since we are approximating the optimal transformations by splines, it is important to select the knots in a data-dependent manner. Some computationally inexpensive methods for tackling such problems are given in Burman (1990b).

(c) It is not difficult to find sufficient conditions which imply Assumption 3. If $(h^*, \tilde{\phi}^*)$ are optimal transformations, then it can be shown that

$$h^*(Y) = \lambda_1^{-1} \sum E\{\phi_j^*(X_j)|Y\} \quad \text{and}$$

$$\phi_i^*(X_i) = E\{h^*(Y)|X_i\} - \sum_{j \neq i} E\{\phi_j^*(X_j)|X_i\}, \qquad i = 1, \dots, d.$$

Let $f_{X_i|Y}$ be the conditional density of X_i given Y. It can be shown that Assumption 3 holds if for any $i = 1, \dots, d$, and for any fixed x, $f_{X_i|Y}(x|\cdot)$ is p' times differentiable and the p'th derivative, which we will denote by $f_{X_i|Y}^{(p')}(x|\cdot)$ for notational simplicity, satisfies

$$\left| f_{X_i|Y}^{(p')}(x|y_1) - f_{X_i|Y}^{(p')}(x|y_2) \right| \le c'|y_1 - y_2|^{\nu}$$

for all y_1, y_2 and x, for some $c' > 0$. Similarly, we can find sufficient conditions which imply Assumption 3.

4. Some important results. We begin this section with a few definitions.

DEFINITION 4.1. (a) Let H_{Yk}, H_{X_jk}, $j = 1, \dots, d$, and $H_{\mathbf{X}k}$ be the subspaces of H_Y, H_{X_j}, $j = 1, \dots, d$, and $H_{\mathbf{X}}$, respectively, defined as

$$H_{Yk} = S_{qk} \cap H_Y, \qquad H_{X_jk} = S_{qk} \cap H_{X_j}, \qquad j = 1, \dots, d,$$

$$\tilde{\phi} = \phi_1 + \cdots + \phi_d \in H_{\mathbf{X}k} \quad \text{if } \phi_j \in H_{X_jk}, \qquad j = 1, \dots, d.$$

(b) In H^0, let P_{Yk}, $P_{\mathbf{X}k}$, P_{X_jk}, $j = 1, \dots, d$, be the orthogonal projections on the subspaces H_{Yk}, $H_{\mathbf{X}k}$ and H_{X_jk}, $j = 1, \dots, d$, respectively.

Let F be the joint distribution function of $(Y, X_1, \dots, X_d)$ and let F_n be its empirical estimate. Now we will define the Hilbert spaces similar to the ones before but with respect to the norm defined by the empirical d.f. F_n.

If f_1 and f_2 are two functions on $[0,1]^{d+1}$, we define

$$\begin{aligned} \|f_1\|_n^2 &= n^{-1} \sum_{t=1}^{n} f_1^2(Y_t, \mathbf{X}_t) \quad \text{and} \\ \langle f_1, f_2 \rangle_n &= n^{-1} \sum_{t=1}^{n} f_1(Y_t, \mathbf{X}_t) f_2(Y_t, \mathbf{X}_t). \end{aligned} \tag{4.1}$$

DEFINITION 4.2. (a) Let $\hat{H}_{Yk}$, $\hat{H}_{X_jk}$, $j = 1, \dots, d$, and $\hat{H}_{\mathbf{X}k}$ be the following Hilbert spaces with respect to the norm and inner product defined in (4.1):

$$h \in \hat{H}_{Yk}, \quad \text{if } h \in S_{qk},\ n^{-1} \sum_{t=1}^{n} h(Y_t) = 0 \text{ and } \|h\|_n < \infty,$$

$$\phi_j \in \hat{H}_{X_jk}, \quad \text{if } h \in S_{qk},\ n^{-1} \sum_{t=1}^{n} \phi_j(X_{jt}) = 0 \text{ and } \|\phi_j\|_n < \infty,$$

$$\tilde{\phi} \in \hat{H}_{\mathbf{X}k}, \quad \text{if } \tilde{\phi} = \phi_1 + \cdots + \phi_d \text{ and } \phi_j \in \hat{H}_{X_jk},\ j = 1, \dots, d.$$

Finally, let $\hat{H}_k^0$ be the Hilbert space of all functions f of the form $f(y, \mathbf{x}) = h(y) + \tilde{\phi}(\mathbf{x})$, $h \in \hat{H}_{Yk}$ and $\tilde{\phi} \in \hat{H}_{\mathbf{X}k}$, with the norm and inner product defined in (4.1).

(b) In $\hat{H}_k^0$, let $\hat{P}_{Yk}$, $\hat{P}_{\mathbf{X}k}$, $\hat{P}_{X_jk}$, $j = 1, \dots, d$, be the orthogonal projections on the subspaces $\hat{H}_{Yk}$, $\hat{H}_{\mathbf{X}k}$ and $\hat{H}_{X_jk}$, $j = 1, \dots, d$, respectively.

The following series of expressions will show the main structure of the proof.

Let e^* be the same as in Theorem 3.1. Let

$$e_k^{*2} = \inf\left\{ \|h - \tilde{\phi}\|^2 : h \in H_{Yk}, \tilde{\phi} \in H_{\mathbf{X}k}, \|h\| = 1 \right\}. \tag{4.2}$$

Let $U_k = P_{Yk} P_{\mathbf{X}k} P_{Yk}$ and $\hat{U}_k = \hat{P}_{Yk} \hat{P}_{\mathbf{X}k} \hat{P}_{Yk}$, then

$$e^{*2} = 1 - \sup\{\langle h, Uh \rangle : h \in H_Y, \|h\| = 1\}, \tag{4.3a}$$

$$e_k^{*2} = 1 - \sup\{\langle h, U_k h \rangle : h \in H_{Yk}, \|h\| = 1\}, \tag{4.3b}$$

$$e_{kn}^{*2} = 1 - \sup\{\langle h, \hat{U}_k h \rangle_n : h \in \hat{H}_{Yk}, \|h\|_n = 1\}. \tag{4.3c}$$

So, in each case the solution is an eigenfunction corresponding to the largest eigenvalue, (4.3a) is the original problem, (4.3b) is the spline approximation to the original problem and (4.3c) is the sample version of the approximation problem in (4.3b).

It is easy to see that Theorem 3.1 follows immediately from the following two theorems. Let us recall from Section 2 that the number of linearly independent optimal transformations $(h, \tilde{\phi})$ is m.

THEOREM 4.1. (a) *There exists an integer k_0 such that for $k > k_0$, the dimension of the solution space for the problem in* (4.3b) *is no larger than m. Let $\lambda_{k1} \geq \lambda_{k2} \geq \cdots$ be the eigenvalues of U_k with the corresponding eigenfunctions $h_{k1}, h_{k2}, \dots$. Then*

$$\sup_{1 \leq t \leq m} \inf_{h^*} \{\| h_{kt} - h^* \| : h^* \text{ is an optimal transformation for variable } Y\}$$
$$\leq ck^{-p},$$

for some $c > 0$.

(b) *Let $\tilde{\phi}_{kt} = P_{\mathbf{X}k} h_{kt} = \phi_{k1t} + \cdots + \phi_{kdt}$, $t = 1, \dots, m$. Then, for $j = 1, \dots, d$,*

$$\sup_{1 \leq t \leq m} \inf_{\phi_j^*} \{\| \phi_{kjt} - \phi_j^* \| : \phi_j^* \text{ is an optimal transformation for variable } X_j\}$$
$$\leq ck^{-p}.$$

Let M_k be the Hilbert space spanned by h_{kt}, $t = 1, \dots, m$, where h_{kt}'s are the same as in Theorem 4.1. If $\tilde{\phi}(\mathbf{x}) = \phi_1(x_1) + \cdots + \phi_d(x_d)$, then sometimes we will refer to ϕ_j as the jth functional component of $\tilde{\phi}$. The following result is true for any $0 < \delta < 1$.

THEOREM 4.2. *Let $k \leq n^{1-\delta}$ for some $\delta > 0$.*

(a) (i) *A solution $(\hat{h}_k, \hat{\tilde{\phi}}_k)$ to the optimization problem in* (3.3) *exists.* (ii) *Except on an event whose probability approaches zero as $n \to \infty$, the solution to the optimization problem in* (3.3) *is unique.*

(b) *Let $(\hat{h}_k, \hat{\tilde{\phi}}_k)$ be the solution to the minimization problem in* (3.3), *$\|\hat{h}_k\|_n = 1$, and $\hat{\tilde{\phi}}_k = \hat{P}_{\mathbf{X}k} \hat{h}_k = \hat{\phi}_{k1} + \cdots + \hat{\phi}_{kd}$. Then each of the following quantities is $O_P((k/n)^{1/2})$,*

$$\inf\left\{\left\| \hat{h}_k - h \right\| : h \in M_k, \| h \| = 1\right\},$$

$$\inf\left\{\left\| \hat{\phi}_{kj} - \phi_j \right\| : \phi_j \text{ is the } j\text{th functional component of } \tilde{\phi} = P_{\mathbf{X}k} h,\ h \in M_k \text{ and } \|h\| = 1\right\}, \qquad j = 1, \dots, d.$$

The following important result follows from Theorem XII.1 in de Boor (1978).

THEOREM 4.3. *Let h be a function on* [0, 1] *which is p' times differentiable and the p'th derivative satisfies $|h^{(p')}(x_1) - h^{(p')}(x_2)| \leq c|x_1 - x_2|^{\nu}$, for some $c > 0$ and $0 < \nu \leq 1$. Let $p = p' + \nu$ and q be any integer greater than or* *equal to p. Then there exists a function* $h_1 \in S_{qk}$ *such that* $\sup\{|h_1(x) - h(x)|: 0 \le x \le 1\} \le c_0 k^{-p}$ *for some* $c_0 > 0$ *(*c_0 *depends on h).*

5. Proof of Theorem 4.1. Before we prove Theorem 4.1, we need the following useful lemmas.

First let us recall that the largest eigenvalue of U is $\lambda_1 > 0$ and the corresponding eigenspace is M. Q_1 is the orthogonal projection on M and $\dim(M) = m < \infty$. Let λ_2 be the second largest eigenvalue of U. Let $U_k = \sum_{j=1}^{k+q} \lambda_{kj} Q_{kj}$, where $\lambda_{k1} \ge \lambda_{k2} \ge \cdots, \lambda_{ki} \ge 0$. The following lemma relates the eigenvalues of U_k to those of U.

LEMMA 5.1. *There is a constant* $c_1 > 0$ *such that* $\lambda_1 - c_1 k^{-p} \le \lambda_{kj} \le \lambda_1$ *for* $j = 1, \dots, m$ *and* $\lambda_{k,m+1} \le \lambda_2$.

PROOF. From Theorems 7.1 and 7.2 in Weinberger (1974) we conclude that

$$\lambda_{kj} \le \lambda_1 \quad \text{for } j = 1, \dots, m \text{ and } \lambda_{k,m+1} \le \lambda_2. \tag{5.1}$$

Now we will show that $\lambda_{kj} > \lambda_1 - c_1 k^{-p}$ for $j = 1, \dots, d$. Let h_t, $t = 1, \dots, m$, be an orthonormal basis of M. By Theorem 4.3, there exists a constant $c_0 > 0$ such that $\|P_{Yk} h_t - h_t\| \le c_0 k^{-p}$ for all $t = 1, \dots, m$. Since any $h \in M$, $\|h\| = 1$, can be written as $\sum_1^m \alpha_t h_t$, where $\sum_1^m \alpha_t^2 = 1$, we conclude that for some $c_2 > 0$,

$$\sup\{\| P_{Yk} h - h \| : h \in M, \|h\| = 1\} \le c_2 k^{-p}. \tag{5.2}$$

A similar argument will show that for some $c_3 > 0$,

$$\sup\Big\{\Big\| P_{X_j k} \phi_j - \phi_j \Big\| : \phi_j \text{ is the } j\text{th functional component of } \tilde{\phi} = P_{\mathbf{X}k} h,\ h \in M, \|h\| = 1\Big\} \le c_3 k^{-p}. \tag{5.3}$$

Let $h \in M$ and $\|h\| = 1$, then

$$\begin{aligned} &|\langle h, U_k h\rangle - \langle h, Uh \rangle| \\ &= \Big| \| P_{\mathbf{X}k} P_{Yk} h \|^2 - \| P_{\mathbf{X}} h \|^2 \Big| \le 2 \| P_{\mathbf{X}k} P_{Yk} h - P_{\mathbf{X}} h \| \\ &\qquad\qquad (\text{since } \| P_{\mathbf{X}k} \|, \| P_{\mathbf{X}} \| \text{ and } \| P_{Yk} \| \text{ are } \le 1) \\ &\le 2\{\| P_{\mathbf{X}k} h - P_{\mathbf{X}} h \| + \| P_{Yk} h - h \|\}. \end{aligned} \tag{5.4}$$

From (5.2) we get

$$\| P_{Yk} h - h \| \le c_2 k^{-p}. \tag{5.5}$$

It is easy to see that

$$\| P_{\mathbf{X}k} h - P_{\mathbf{X}} h \| = \inf\Big\{\Big\| \tilde{\phi}_k - P_{\mathbf{X}} h \Big\| : \tilde{\phi}_k \in H_{\mathbf{X}k}\Big\}. \tag{5.6}$$

Let $P_{\mathbf{X}}h = \phi_1 + \cdots + \phi_d$, then the right-hand side of (5.6) is less than

$$\left\| \sum_1^d P_{X_{jk}}\phi_j - \sum_1^d \phi_j \right\| \le \sum_1^d \left\| P_{X_{jk}}\phi_j - \phi_j \right\|. \tag{5.7}$$

Now, (5.3) tells us that each term on the right-hand side of (5.7) is less than $c_3 k^{-p}$ and so (5.5), (5.6) and (5.7) prove that for some constant $c_4 > 0$,

$$|\langle h, U_k h\rangle - \langle h, Uh\rangle| \le c_4 k^{-p}. \tag{5.8}$$

Let us note that for any $h \in M$, $\langle h, Uh\rangle = \lambda_1$ and that all the inequalities in (5.5) to (5.8) could be proved uniformly for all $h \in M$, $\|h\| = 1$. Consequently,

$$\inf\{\langle h, U_k h\rangle : h \in M, \|h\| = 1\} \ge \lambda_1 - c_4 k^{-p} \tag{5.9}$$

and this proves that $\lambda_{km} \ge \lambda_1 - c_4 k^{-p}$. Since $\lambda_{kj} \ge \lambda_{km}$ for $j = 1, \dots, m$, this proves the result. □

Proof of Theorem 4.1. We will prove part (a) later, but first we will show how part (b) could be proved from part (a).

(5.10) Let $h_k \in M_k$, $\|h_k\| = 1$ and $\tilde{\phi}_k = P_{\mathbf{X}k} h_k = \phi_{k1} + \cdots + \phi_{kd}$.

Because of Proposition (5.2) in Breiman and Friedman, it is enough to show

$$\inf\left\{\|\tilde{\phi}_k - P_{\mathbf{X}}h\| : h \in M, \|h\| = 1\right\} \le c_5 k^{-p}. \tag{5.11}$$

For $h \in M$ and $\|h\| = 1$,

$$\|\tilde{\phi}_k - P_{\mathbf{X}}h\| \le \|P_{\mathbf{X}k}h - P_{\mathbf{X}}h\| + \|h_k - h\|. \tag{5.12}$$

We have shown in the proof of Lemma 5.1 [see (5.6) and (5.7)] that

$$\sup\{\|P_{\mathbf{X}k}h - P_{\mathbf{X}}h\| : h \in M, \|h\| = 1\} \le c_6 k^{-p}. \tag{5.13}$$

(5.12), (5.13) and the part (a) of the theorem together prove the second part.

Now let us prove part (a). Let Γ be a positively-oriented closed curve in the complex plane which encloses only one eigenvalue of U, namely λ_1, and every point λ on Γ is at least $(\lambda_1 - \lambda_2)/2$ distance away from λ_1 and λ_2. Because of Lemma 5.1, there exist an integer k_0 such that for any $k > k_0$, $\lambda_{k1}, \dots, \lambda_{km}$ are all inside Γ and all of them are at least ε_0 distance away from Γ for some $\varepsilon_0 > 0$. So the resolvents $R_k(\lambda) = (U_k - \lambda)^{-1}$ and $R(\lambda) = (U - \lambda)^{-1}$ exist and are bounded for each λ on Γ.

Since U and U_k are self-adjoint, the following are orthogonal projections [see the proof of Theorem 2.27 and Section 2.7.7 in Chatelin (1983)].

$$S_k = (2\pi i)^{-1} \int_\Gamma R_k(\lambda)\, d\lambda, \qquad S_k = \sum_1^m Q_{kj}, \tag{5.14a}$$

$$S = (2\pi i)^{-1} \int_\Gamma R(\lambda)\, d\lambda, \qquad S = Q_1. \tag{5.14b}$$

Let us note that part (a) is proved if we can show that

$$\|(S_k - S)S_k\| \le c_7 k^{-p}. \tag{5.15}$$

To show this we will prove that

$$\|(S_k - S)S\| \le c_8 k^{-p}. \tag{5.16}$$

Let us assume for the moment that (5.16) is correct. Then we can choose k_0 in such a way that $\|(S_k - S)S\| < 1$ for $k > k_0$. Because of (5.14a) and (5.14b), $M_k = S_k H_Y$ and $M = SH_Y$. We also know that $\dim(M_k) = \dim(M) = m$. By Lemma 5.2 given at the end of this section or by Theorem 6.34 in Kato [(1976), page 56],

$$\|(S_k - S)S_k\| = \|(S_k - S)S\|$$

and this proves (5.15). [Though the result in Kato (1976) is stated for finite-dimensional spaces, the author points out in the footnote that this result is true even when the dimension of the space is infinite. Indeed, the proof given in the book has nothing to do with the dimensionality of the space.]

So what is left to prove is (5.16). First let us state the following resolvent equation which is quite simply to verify:

$$R(\lambda) - R_k(\lambda) = R_k(\lambda)(U_k - U)R(\lambda). \tag{5.17}$$

Since $R(\lambda)S = (\lambda_1 - \lambda)^{-1}S$, we have

$$(S - S_k)S = (2\pi i)^{-1}\int_\Gamma (\lambda_1 - \lambda)^{-1} R_k(\lambda)\, d\lambda (U_k - U)S.$$

Since $(2\pi i)^{-1}\int_\Gamma(\lambda_1 - \lambda)^{-1}R_k(\lambda)\,d\lambda$ is a bounded operator and since for any $h \in M$, $\|h\| = 1$, $Sh = h$,

$$\|(S - S_k)h\| \le c_9\|(U_k - U)h\|. \tag{5.18}$$

Now, $\|(U_k - U)h\| = \|P_{Yk}P_{\mathbf{X}k}P_{Yk}h - P_Y P_{\mathbf{X}}h\|$ and this can be bounded above by

$$\begin{aligned} &\|P_{Yk}P_{\mathbf{X}k}(P_{Yk} - P_Y)h\| + \|P_{Yk}(P_{\mathbf{X}k} - P_{\mathbf{X}})h\| + \|(P_{Yk} - P_Y)P_{\mathbf{X}}h\| \\ &\quad \le \|P_{Yk}h - h\| + \|P_{\mathbf{X}k}h - P_{\mathbf{X}}h\| + \|(P_{Yk} - P_Y)P_{\mathbf{X}}h\|. \end{aligned} \tag{5.19}$$

We will show that each term in (5.19) is $O(k^{-p})$. (5.2) tells us that the first term in (5.19) is no larger than $c_2 k^{-p}$. We have already shown in (5.6) and (5.7) (in the proof of Lemma 5.1) that $\|P_{\mathbf{X}k}h - P_{\mathbf{X}}h\|$ is no larger than dc_3k^{-p}. The third term in (5.19) is bounded above by

$$\sum_1^d \|(P_{Yk} - P_Y)\phi_j\| = \sum_1^d \|P_{Yk}\bar{h}_j - \bar{h}_j\| = \sum_1^d \|P_{Yk}\tilde{h}_j - \tilde{h}_j\|\,\|\bar{h}_j\|,$$

where $\phi_1 + \cdots + \phi_d = P_{\mathbf{X}}h$, $\bar{h}_j = P_Y\phi_j$ and $\tilde{h}_j = \bar{h}_j/\|\bar{h}_j\|$. Using (5.2), we get

$$\sum_1^d \|P_{Yk}\tilde{h}_j - \tilde{h}_j\|\,\|\bar{h}_j\| \le c_2k^{-p}\sum_1^d \|\bar{h}_j\| \le c_2k^{-p}\sum_1^d \|\phi_j\|.$$

By Proposition 5.2 from Breiman and Friedman,

$$\sum_1^d \|\phi_j\|^2 \le c_{10} \left\| \sum_1^d \phi_j \right\|^2 .$$

where c_{10} depends only on F. Since $\|\phi_1 + \cdots + \phi_d\| = \|P_{\mathbf{X}} h\| \le 1$, it follows that the third term in (5.19) is bounded above by $c_{10} k^{-p}$ for some $c_{10} > 0$. This completes the proof of Theorem 4.1. □

LEMMA 5.2. *Let H_1 and H_2 be two finite-dimensional subspaces of a Hilbert space H. Let us assume that* $\dim(H_1) = \dim(H_2) < \infty$. *If P_1 and P_2 are the orthogonal projections onto H_1 and H_2, then*

$$\|(P_1 - P_2)P_1\| = \|(P_2 - P_1)P_2\|.$$

PROOF. Let $\dim(H_1) = \dim(H_2) = l$, say. Let $\xi_1, \dots, \xi_l$ be an orthonormal basis of H_1 and let $\eta_1, \dots, \eta_l$ be an orthonormal basis of H_2. It can be shown that

$$\|(P_1 - P_2)P_1\| = \sup\{\|\mathbf{u} - P_2\mathbf{u}\| : \mathbf{u} \in H_1, \|\mathbf{u}\| = 1\}.$$

So,

$$\begin{aligned} \|(P_1 - P_2)P_1\|^2 &= \sup\{\|\mathbf{u} - P_2\mathbf{u}\|^2 : \mathbf{u} \in H_1, \|\mathbf{u}\| = 1\} \\ &= 1 - \inf\{\|P_2\mathbf{u}\|^2 : \mathbf{u} \in H_1, \|\mathbf{u}\| = 1\} \\ &= 1 - \inf\left\{\sum \langle \eta_j, \mathbf{u}\rangle^2 : \mathbf{u} \in H_1, \|\mathbf{u}\| = 1\right\}. \end{aligned}$$

Let us note that any $\mathbf{u} \in H_1$, $\|\mathbf{u}\| = 1$, can be written as a linear combination of $\xi_1, \dots, \xi_l$, that is, $\mathbf{u} = \sum a_j \xi_j$ with $\mathbf{a}'\mathbf{a} = \sum a_j^2 = 1$. So we conclude that

$$\|(P_1 - P_2)P_1\|^2 = 1 - \inf\left\{\mathbf{a}'\Sigma\,\Sigma'\mathbf{a} : \mathbf{a} \in R^l, \mathbf{a}'\mathbf{a} = 1\right\},$$

where Σ is a $l \times l$ matrix with (i, j)th element $\langle \xi_i, \eta_j \rangle$. Similarly, we can show that

$$\|(P_2 - P_1)P_2\|^2 = 1 - \inf\left\{\mathbf{a}'\Sigma'\,\Sigma\,\mathbf{a} : \mathbf{a} \in R^l, \mathbf{a}'\mathbf{a} = 1\right\}.$$

The proof follows since $\Sigma'\Sigma$ and $\Sigma\Sigma'$ have the same eigenvalues. □

6a. Expressions for U_k and $\hat{U}_k$. This section is devoted to finding explicit expressions for U_k and $\hat{U}_k$. Since the transformation of any variable is approximated by a linear combination of $k + q$ B-splines and since the mean of this approximate transformation is zero, there are $k + q - 1$ free parameters. We use this fact to construct exact expressions for U_k and $\hat{U}_k$.

Let $\mathbf{e}_1, \dots, \mathbf{e}_{k+q-1}$ be $(k + q)$-dimensional vectors which are orthogonal to each other and orthogonal to $\mathbf{l}$, the vector of 1 and $\mathbf{e}_i'\mathbf{e}_i = 1$ for all i.

For $1 \le t \le d$ and $1 \le i \le k + q - 1$, we define

$$(6.1)\qquad \psi_{kti}(x_t) = \sum_j \frac{e_{ij} B_{kj}(x_t)}{k b_{ktj}}, \qquad \hat{\psi}_{kti}(x_t) = \sum_j \frac{e_{ij} B_{kj}(x_t)}{k \hat{b}_{ktj}},$$

where $b_{ktj} = EB_{kj}(X_t)$ and $\hat{b}_{ktj} = n^{-1}\sum_{u=1}^{n} B_{kj}(X_{tu})$.

Let us note that each B_{kj} is positive on its support which is an interval of length $(q+1)k^{-1}$ and consequently, $EB_{kj}(X_t) > 0$, by Assumption 4. For any variable, say X_t, $\hat{b}_{ktj} > 0$ for all j if each of the open subintervals $((l-1)k^{-1}, lk^{-1})$, $l = 1, \dots, k$, has at least one observation. By Assumption 4, $f_{X_t}(x) > \gamma_0 > 0$ for all $0 \le x \le 1$, where f_{X_t} is the marginal density of X_t. An easy calculation will show that

$$\operatorname{Prob}\left\{\hat{b}_{ktj} = 0 \text{ for some } j = 1, \dots, k+q\right\} \le k \exp(-\gamma_0 n/k).$$

This probability is small when $k \le n^{1-\delta}$ for some $\delta > 0$. Let us note that $\hat{b}_{ktj} > 0$ for all t and j, if the knots of the B-splines are placed at the sample quantiles (see the remark at the end of Section 3).

It is easy to check that $\int \psi_{kti}\, dF = 0$ and $\int \hat{\psi}_{kti}\, dF_n = 0$ for all t and i. It is also easy to see that for a transformation ϕ_{kt} of the tth X-variable which is a linear combination of B-splines,

if $\int \phi_{kt}\, dF = 0$,

then ϕ_{kt} is a linear combination of $\{\psi_{kti}\colon i = 1, \dots, k+q-1\}$,

if $\int \phi_{kt}\, dF_n = 0$,

then ϕ_{kt} is a linear combination of $\left\{\hat{\psi}_{kti}\colon i = 1, \dots, k+q-1\right\}$.

Let $\boldsymbol{\psi}_{kt}$ and $\hat{\boldsymbol{\psi}}_{kt}$ be the $(k+q-1)$-dimensional vector of the ψ functions and let $\mathbf{B}_{kt}$ be the $(k+q)$-dimensional vector of B-splines corresponding to variable X_t. So we can easily write

$$(6.2)\qquad \boldsymbol{\psi}_{kt}(x_t) = D_{kt}\mathbf{B}_{kt}(x_t) \quad \text{and} \quad \hat{\boldsymbol{\psi}}_{kt}(x_t) = \hat{D}_{kt}\mathbf{B}_{kt}(x_t),$$

where D_{kt} and $\hat{D}_{kt}$ are $(k+q-1) \times (k+q)$ matrices. The elements of the ith rows of D_{kt} and $\hat{D}_{kt}$ are $\{e_{ij}/(kb_{ktj}),\ j = 1, \dots, k+q\}$ and $\{e_{ij}/(k\hat{b}_{ktj}),\ j = 1, \dots, k+q\}$, respectively, $i = 1, \dots, k+q-1$.

Let $\boldsymbol{\psi}_k$ be the $(kd + qd - d)$-dimensional vectors of all the ψ functions. We define $\hat{\boldsymbol{\psi}}_k$ similarly. Then we can write

$$(6.3)\qquad \boldsymbol{\psi}_k(\mathbf{x}) = D_k\mathbf{B}_k(\mathbf{x}), \qquad \hat{\boldsymbol{\psi}}_k(\mathbf{x}) = \hat{D}_k\mathbf{B}_k(\mathbf{x}),$$

where $\mathbf{B}_k$ is the vector of all the B-splines corresponding to the X variables and D_k and $\hat{D}_k$ are $(kd + qd - d) \times (kd + qd)$ block diagonal matrices. D_{kt},

$t = 1, \dots, d$, are the diagonal blocks of D_k and $\hat{D}_{kt}$, $t = 1, \dots, d$, are the diagonal blocks of $\hat{D}_k$.

Let us define the following matrices,

$$A_{00} = E\mathbf{B}_k(Y)\mathbf{B}_k'(Y), \qquad A_{0\mathbf{X}} = E\mathbf{B}_k(Y)\boldsymbol{\psi}_k'(\mathbf{X}), \tag{6.4a}$$
$$A_{\mathbf{XX}} = E\boldsymbol{\psi}_k(\mathbf{X})\boldsymbol{\psi}_k'(\mathbf{X}),$$

$$\hat{A}_{00} = n^{-1}\sum_{j=1}^{n} \mathbf{B}_k(Y_j)\mathbf{B}_k'(Y_j), \qquad \hat{A}_{0\mathbf{X}} = n^{-1}\sum_{j=1}^{n} \mathbf{B}_k(Y_j)\hat{\boldsymbol{\psi}}_k'(\mathbf{X}_j), \tag{6.4b}$$

$$\hat{A}_{\mathbf{XX}} = n^{-1}\sum_{j=1}^{n} \hat{\boldsymbol{\psi}}_k(\mathbf{X}_j)\hat{\boldsymbol{\psi}}_k'(\mathbf{X}_j), \tag{6.4c}$$
$$A_{\mathbf{X}0} = A_{0\mathbf{X}}' \quad \text{and} \quad \hat{A}_{\mathbf{X}0} = \hat{A}_{0\mathbf{X}}'.$$

It is easy to see that, for $h(y) = \boldsymbol{\theta}'\mathbf{B}_k(y)$,

$$(U_k h)(y) = \boldsymbol{\theta}' A_{0\mathbf{X}} A_{\mathbf{XX}}^{-1} A_{\mathbf{X}0} A_{00}^{-1}\mathbf{B}_k(y), \tag{6.5a}$$

$$(\hat{U}_k h)(y) = \boldsymbol{\theta}' \hat{A}_{0\mathbf{X}} \hat{A}_{\mathbf{XX}}^{-1} \hat{A}_{\mathbf{X}0} \hat{A}_{00}^{-1}\mathbf{B}_k(y). \tag{6.5b}$$

In the last expression, if $\hat{A}_{\mathbf{XX}}$ is singular, we take $\hat{A}_{\mathbf{XX}}^{-1}$ to be a generalized inverse of $\hat{A}_{\mathbf{XX}}$. The same is true for $\hat{A}_{00}^{-1}$. It is easy to check that

$$E(U_k h)(Y) = 0 \quad \text{and} \quad n^{-1}\sum_{j=1}^{n} (\hat{U}_k h)(Y_j) = 0. \tag{6.6}$$

We end this section with the following remarks.

REMARK 1. For any $1 \le t \le d$, the sample estimate of the transformation of variable X_t is a linear combination of $\hat{\psi}_{ktj}$, $j = 1, \dots, k + q - 1$. By construction $\hat{\psi}_{ktj}$'s are well-defined if $\hat{b}_{ktj}$'s are positive. In practice this condition is satisfied, since the knots of the B-splines are usually placed at the sample quantiles (see the remark at the end of Section 3).

REMARK 2. The following scheme could also be used to estimate the optimal transformations. For each X variable, consider the first $k + q - 1$ B-splines with their sample means subtracted. Let $\tilde{\mathbf{B}}_k(\mathbf{x})$ be the $d(k + q - 1)$-dimensional vector of such sample mean adjusted B-splines for all the X variables. Let $L_{0\mathbf{X}} = \int \mathbf{B}_k(y)\tilde{\mathbf{B}}_k'(\mathbf{x})\, dF_n(y, \mathbf{x})$, $L_{\mathbf{X}0} = L_{0\mathbf{X}}'$ and $L_{\mathbf{XX}} = \int \tilde{\mathbf{B}}_k\tilde{\mathbf{B}}_k'\, dF_n$. Let $L_{\mathbf{XX}}^{-1}$ be the inverse of $L_{\mathbf{XX}}$. If $L_{\mathbf{XX}}$ is singular, then we take $L_{\mathbf{XX}}^{-1}$ to be the Moore–Penrose generalized inverse (or any symmetric and reflexive generalized inverse) of $L_{\mathbf{XX}}$ [see Section 1b.5 in Rao (1973)]. If $h(y) = \boldsymbol{\theta}'\mathbf{B}_k(y)$, $\boldsymbol{\theta} \in R^{k+q}$ and $\|h\|_n^2 = \boldsymbol{\theta}'\hat{A}_{00}\boldsymbol{\theta} = 1$, then $\|h - \hat{P}_{\mathbf{X}k}h\|_n^2 = 1 - \boldsymbol{\theta}'L_k\boldsymbol{\theta}$, where $L_k = L_{0\mathbf{X}}L_{\mathbf{XX}}^{-1}L_{\mathbf{X}0}$. Let $\hat{\boldsymbol{\theta}}_k$ be the vector at which the quadratic form $\boldsymbol{\theta}'L_k\boldsymbol{\theta}$ attains its supremum subject to the constraint $\boldsymbol{\theta}'\hat{A}_{00}\boldsymbol{\theta} = 1$. If $\hat{h}_k(y) = \hat{\boldsymbol{\theta}}_k'\mathbf{B}_k(y)$, then $(\hat{h}_k, \hat{P}_{\mathbf{X}k}\hat{h}_k)$ are estimates of the optimal transformations.

6b. The proof of Theorem 4.2. Before we prove Theorem 4.2, we need a few preliminary results. The first lemma can be easily derived from Lemma 4.4 in Section 4 of Burman (1985). Let $0 < \delta' < \delta/2$, where δ is the constant in Theorem 4.2.

LEMMA 6.1. (a) *There exist constants* $0 < c_{11} < c_{12}$ *such that all the eigenvalues of* A_{00} *and* $A_{\mathbf{XX}}$ *lie between* $c_{11}k^{-1}$ *and* $c_{12}k^{-1}$.

(b) *Let* $\delta_{nk} = (kn^{1-\delta'})^{-1/2}$. *Then* $\|\hat{A}_{00} - A_{00}\|$, $\|\hat{A}_{0\mathbf{X}} - A_{0\mathbf{X}}\|$ *and* $\|\hat{A}_{\mathbf{XX}} - A_{\mathbf{XX}}\|$ *are* $o_P(\delta_{nk})$ *uniformly for* $k \le n^{1-\delta'}$.

PROOF. (a) Let $\mathbf{u} \in R^{k+q}$ with $\|\mathbf{u}\|^2 = \sum u_j^2 = 1$. Then, $\mathbf{u}'A_{00}\mathbf{u} = \int\{\sum u_j B_{kj}\}^2\, dF$. Using the two-sided inequality (12) in Section 4 of Stone (1986), we get

$$c_{13}k^{-1} \sum u_j^2 \le \mathbf{u}'A_{00}\mathbf{u} \le c_{14}k^{-1} \sum u_j^2. \tag{6.7}$$

This proves the result for A_{00}. Now let us prove it for $A_{\mathbf{XX}}$.

Let us note that for any $\mathbf{u} \in R^{d(k+q-1)}$ which is composed of $\mathbf{u}_1, \dots, \mathbf{u}_d$, $\mathbf{u}_t$'s in R^{k+q-1}, $\mathbf{u}'A_{\mathbf{XX}}\mathbf{u}$ can be written as $\int[\phi_1 + \cdots + \phi_d]^2\, dF$, where $\phi_t(x_t) = \mathbf{u}_t'\boldsymbol{\psi}_{kt}(x_t)$, $t = 1, \dots, d$. By Proposition 5.2 of Breiman and Friedman,

$$c_{15} \sum \|\phi_t\|^2 \le \mathbf{u}'A_{\mathbf{XX}}\mathbf{u} \le c_{16} \sum \|\phi_t\|^2.$$

Arguing the same way as before we can show that for any $t = 1, \dots, d$,

$$c_{17}k^{-1}\|\mathbf{u}_t\|^2 \le \|\phi_t\|^2 \le c_{18}k^{-1}\|\mathbf{u}_t\|^2.$$

This proves the result for $A_{\mathbf{XX}}$.

(b) By Theorem 1.19 in Chatelin (1983),

$$\|\hat{A}_{00} - A_{00}\| \le \sup_{1 \le i \le k+q} \sum_{j=1}^{k+q} \left| \int B_{k0i} B_{k0j}\, d(F_n - F) \right|.$$

Here B_{k0i}'s are the B-splines for variable Y. Since $B_{k0i}B_{k0j} \equiv 0$ for $|i - j| > q + 1$ and $\operatorname{Var}(B_{k0i}(Y)B_{k0j}(Y)) = O(k^{-1})$ for $|i - j| \le q + 1$, an application of Hoeffding's inequality [Theorem 3 in Hoeffding (1963)] will prove the result.

We will give only an outline of the proof for $\|\hat{A}_{\mathbf{XX}} - A_{\mathbf{XX}}\| = o_P(\delta_{nk})$. The proof for the other result $\|\hat{A}_{0\mathbf{X}} - A_{0\mathbf{X}}\| = o_P(\delta_{nk})$ is similar. The detailed proofs are given in Burman (1990a). Here, all the probability statements like $o_P(\delta_{nk})$ and so on are understood to be uniform for $k \le n^{1-\delta'}$. Now, $A_{\mathbf{XX}} = D_k \int \mathbf{B}_k\mathbf{B}_k'\, dF\, D_k'$ and $\hat{A}_{\mathbf{XX}} = \hat{D}_k \int \mathbf{B}_k\mathbf{B}_k'\, dF_n\, \hat{D}_k'$, where $\mathbf{B}_k$ is a vector of all the B-splines for all the X variables. It can be easily shown that $\|\hat{D}_k - D_k\| = o_P(k\delta_{nk})$ and hence the proof follows if we can show that $\|\int \mathbf{B}_k\mathbf{B}_k'\, d(F_n - F)\| = o_P(\delta_{nk})$. Let $\mathbf{B}_{ki}$ be the vector of the B-splines for variable X_i, $i = 1, \dots, d$, and $G_{ij} = \int \mathbf{B}_{ki}\mathbf{B}_{kj}'\, d(F_n - F)$, $1 \le i, j \le d$. Arguments similar to those used in proving $\|\hat{A}_{00} - A_{00}\| = o_P(\delta_{nk})$ can be used to show that $\|G_{ii}\| = o_P(\delta_{nk})$, $i = 1, \dots, d$.

For $i \neq j$, $\|G_{ij}\|^{2s} \leq \operatorname{trace}((G_{ij}G_{ij})^s)$. It can be shown that

$$E\Big[\operatorname{trace}\Big(\big(G_{ij}G_{ij}\big)^s\Big)\Big] \leq c(s)(kn)^{-s},$$

where $c(s)$ is a positive constant depending only on s. By choosing $s > \delta'^{-1} - 1$, it can be shown that $\|G_{ij}\| = o_P(\delta_{nk})$ for $1 \leq i, j \leq d$. □

Let us recall that λ_{kj}'s are the eigenvalues of the operator U_k. Let $\hat{\lambda}_{kj}$'s be the eigenvalues of the operator $\hat{U}_k$.

LEMMA 6.2.

$$\max\Big\{\big|\hat{\lambda}_{kj} - \lambda_{kj}\big| : 1 \leq j \leq m+1\Big\} = o_P(k\delta_{nk}) \quad \textit{uniformly for } k \leq n^{1-\delta'}.$$

PROOF. Let us first note that the eigenvalues of the operator U_k are identical to the eigenvalues of the matrix $W_k = A_{00}^{-1/2}A_{0\mathbf{X}}A_{\mathbf{XX}}^{-1}A_{\mathbf{X}0}A_{00}^{-1/2}$. Also, the eigenvalues of the operator $\hat{U}_k$ are identical to those of the matrix $\hat{W}_k = \hat{A}_{00}^{-1/2}\hat{A}_{0\mathbf{X}}\hat{A}_{\mathbf{XX}}^{-1}\hat{A}_{\mathbf{X}0}\hat{A}_{00}^{-1/2}$.

This tells us that our lemma will be proved if we can show that $\|\hat{W}_k - W_k\| = o_P(k\delta_{nk})$ uniformly for $k \leq n^{1-\delta'}$, where $\|\cdot\|$ refers to the usual operator norm for matrices. Let us note that $\|\hat{W}_k - W_k\|$ is bounded above by

$$\begin{aligned}
&\big\|\hat{A}_{00}^{-1/2} - A_{00}^{-1/2}\big\|\,\big\|\hat{A}_{0\mathbf{X}}\hat{A}_{\mathbf{XX}}^{-1}\hat{A}_{\mathbf{X}0}\hat{A}_{00}^{-1/2}\big\| \\
&\quad+\big\|A_{00}^{-1/2}\big\|\,\big\|\hat{A}_{0\mathbf{X}} - A_{0\mathbf{X}}\big\|\,\big\|\hat{A}_{\mathbf{XX}}^{-1}\hat{A}_{\mathbf{X}0}\hat{A}_{00}^{-1/2}\big\| \\
&\quad+\big\|A_{00}^{-1/2}A_{0\mathbf{X}}\big\|\,\big\|\hat{A}_{\mathbf{XX}}^{-1} - A_{\mathbf{XX}}^{-1}\big\|\,\big\|\hat{A}_{\mathbf{X}0}\hat{A}_{00}^{-1/2}\big\| \\
&\quad+\big\|A_{00}^{-1/2}A_{0\mathbf{X}}A_{\mathbf{XX}}^{-1}\big\|\,\big\|\hat{A}_{\mathbf{X}0} - A_{\mathbf{X}0}\big\|\,\big\|\hat{A}_{00}^{-1/2}\big\| \\
&\quad+\big\|A_{00}^{-1/2}A_{0\mathbf{X}}A_{\mathbf{XX}}^{-1}A_{\mathbf{X}0}\big\|\,\big\|\hat{A}_{00}^{-1/2} - A_{00}^{-1/2}\big\|.
\end{aligned} \tag{6.8}$$

By part (b) of Lemma 6.1 and part (a) of Lemma 6.3, the following is true uniformly for $k \leq n^{1-\delta'}$:

$$\big\|\hat{A}_{00}^{-1/2} - A_{00}^{-1/2}\big\| = O_P(k^{3/2})\big\|\hat{A}_{00} - A_{00}\big\| = o_P\big(k^{3/2}\delta_{nk}\big).$$

Since by Lemma 6.1, $\|\hat{A}_{0\mathbf{X}}\| = O_P(k^{-1})$, $\|\hat{A}_{\mathbf{XX}}^{-1}\| = O_P(k)$ and $\|\hat{A}_{00}^{-1/2}\| = O_P(k^{1/2})$, we can conclude that the first term in (6.8) is $o_P(k\delta_{nk})$ uniformly for $k \leq n^{1-\delta'}$. Similar arguments will show that the other terms in (6.8) are $o_P(k\delta_{nk})$ uniformly for $k \leq n^{1-\delta'}$. This concludes the proof. □

Before we state the next result, let us note that by part (a) of Lemma 6.1, A_{00} is a positive definite matrix and consequently, $\rho(\lambda) = (A_{00} + \lambda)^{-1}$ exists for any $\lambda > 0$. In the next lemma, we will use the following representation of $A_{00}^{-1/2}$ [see relation (3.43), page 282 in Kato (1976)]

$$A_{00}^{-1/2} = \pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)\,d\lambda.$$

This integral is well-defined since A_{00} is a positive definite matrix [see Section 3.11, Chapter V in Kato (1976)].

LEMMA 6.3. *We have that*

(a) $\|\hat{A}_{00}^{-1/2} - A_{00}^{-1/2}\| = O_P(k^{3/2})\|\hat{A}_{00} - A_{00}\|$, *uniformly for* $k \le n^{1-\delta'}$,

(b) $\hat{A}_{00}^{-1/2} - A_{00}^{-1/2} = -\pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)(\hat{A}_{00} - A_{00})\rho(\lambda)\,d\lambda + G$,

where $\|G\| = o_P(k^{5/2}\delta_{nk}^2)$ *uniformly for* $k \le n^{1-\delta'}$.

PROOF. (a) In the proof of this lemma all the probability statements like $o_P(\ )$, $O_P(\ )$ and so on are understood to be uniform for $k \le n^{1-\delta'}$. For $\lambda > 0$, let $\hat{\rho}(\lambda) = (\hat{A}_{00} + \lambda)^{-1}$. By Lemma 6.1, $\hat{\rho}(\lambda)$ exists since $\hat{A}_{00}$ is positive definite (for any $k \le n^{1-\delta'}$) except on an event whose probability goes to zero as $n \to \infty$. The following resolvent equation is easy to verify:

$$\hat{\rho}(\lambda) - \rho(\lambda) = -\rho(\lambda)(\hat{A}_{00} - A_{00})\hat{\rho}(\lambda). \tag{6.9}$$

Using this resolvent equation we obtain:

$$\begin{aligned}\hat{A}_{00}^{-1/2} - A_{00}^{-1/2} &= \pi^{-1}\int_0^\infty \lambda^{-1/2}\{\hat{\rho}(\lambda) - \rho(\lambda)\}\,d\lambda \\ &= -\pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)(\hat{A}_{00} - A_{00})\hat{\rho}(\lambda)\,d\lambda.\end{aligned} \tag{6.10}$$

The following can be proved easily and is stated without a proof.

$$\pi^{-1}\int_0^\infty \lambda^{-1/2}(\lambda + t)^{-i}\,d\lambda \le c_{19}t^{-(2i-1)/2} \quad \text{for some } c_{19} > 0,\ t > 0,\ i = 1, 2, 3. \tag{6.11}$$

Let $\tilde{\gamma}$ be the minimum of the smallest eigenvalues of $\hat{A}_{00}$ and A_{00}. Then, by Lemma 6.1, $\tilde{\gamma}^{-1} = O_P(k)$. Noting that $\|\rho(\lambda)\| \le (\lambda + \tilde{\gamma})^{-1}$ and $\|\hat{\rho}(\lambda)\| \le (\lambda + \tilde{\gamma})^{-1}$ for any $\lambda > 0$, we get from (6.10) and (6.11)

$$\begin{aligned}\|\hat{A}_{00}^{-1/2} - A_{00}^{-1/2}\| &\le \pi^{-1}\int_0^\infty \lambda^{-1/2}(\lambda + \tilde{\gamma})^{-2}\,d\lambda\,\|\hat{A}_{00} - A_{00}\| \\ &\le c_{19}\tilde{\gamma}^{-3/2}\|\hat{A}_{00} - A_{00}\| = O_P(k^{3/2})\|\hat{A}_{00} - A_{00}\|.\end{aligned}$$

(b) Using the resolvent equation (6.9) once again in (6.10), we get

$$\hat{A}_{00}^{-1/2} - A_{00}^{-1/2} = -\pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)(\hat{A}_{00} - A_{00})\rho(\lambda)\,d\lambda + G,$$

where

$$G = \pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)(\hat{A}_{00} - A_{00})\rho(\lambda)(\hat{A}_{00} - A_{00})\hat{\rho}(\lambda)\,d\lambda.$$

So,

$$\begin{aligned}\|G\| &\le \pi^{-1}\int_0^\infty \lambda^{-1/2}\|\rho(\lambda)\|^2\|\hat{\rho}(\lambda)\|\,d\lambda\|\hat{A}_{00}-A_{00}\|^2\\ &\le \pi^{-1}\int_0^\infty \lambda^{-1/2}(\lambda+\tilde{\gamma})^{-3}\,d\lambda\|\hat{A}_{00}-A_{00}\|^2\\ &\le c_{19}\tilde{\gamma}^{-5/2}\|\hat{A}_{00}-A_{00}\|^2 &&\text{[by (6.11)]}\\ &= O_P(k^{5/2})o_P(\delta_{nk}^2). &&\text{[by Lemma 6.1]}\end{aligned}$$

This concludes the proof of this lemma. □

PROOF OF THEOREM 4.2. (a)(i) Let $\tilde{\mathbf{B}}_k(\mathbf{x})$, $L_{0\mathbf{X}}, L_{\mathbf{X}0}$, $L_{\mathbf{XX}}$ and L_k be the same as in Remark 2 in Section 6a. If $h(y)=\boldsymbol{\theta}'\mathbf{B}_k(y)$, $\boldsymbol{\theta}\in R^{k+q}$, then

$$(\hat{P}_{\mathbf{X}k}h)(\mathbf{x}) = \boldsymbol{\theta}'L_{0\mathbf{X}}L_{\mathbf{XX}}^{-1}\tilde{\mathbf{B}}_k(\mathbf{x}).$$

A simple calculation will show that

$$\begin{aligned}e_{kn}^{*2} &= \inf\left\{\|h-\hat{P}_{\mathbf{X}k}h\|_n^2: h\in S_{kq}, \|h\|_n=1\right\}\\ &= 1-\sup\left\{\boldsymbol{\theta}'L_k\boldsymbol{\theta}: \boldsymbol{\theta}\in R^{k+q}, \boldsymbol{\theta}'\hat{A}_{00}\boldsymbol{\theta}=1\right\}.\end{aligned}\tag{6.12}$$

If $\hat{\boldsymbol{\theta}}_k$ is a solution to this maximization problem and $\hat{h}_k(y)=\hat{\boldsymbol{\theta}}_k'\mathbf{B}_k(y)$ and $\hat{\phi}_k=\hat{P}_{\mathbf{X}k}\hat{h}_k$, then $(\hat{h}_k,\hat{\phi}_k)$ are the sample estimates of the optimal transformations. Clearly, (6.12) has a solution if $\hat{A}_{00}$ is nonsingular. We will now argue that (6.12) has a solution even if $\hat{A}_{00}$ is singular.

In R^{k+q}, let Ω be the vector space generated by the eigenvectors corresponding to the nonzero eigenvalues of $\hat{A}_{00}$. For $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in R^{k+q}, let $s_1(y)=\boldsymbol{\theta}_1'\mathbf{B}_k(y)$ and $s_2(y)=\boldsymbol{\theta}_2'\mathbf{B}_k(y)$. Then

$$|\boldsymbol{\theta}_1'L_k\boldsymbol{\theta}_2| = \left|\int s_1(y_1)\tilde{\mathbf{B}}_k'(\mathbf{x}_1)L_{\mathbf{XX}}^{-1}\tilde{\mathbf{B}}_k(\mathbf{x}_2)s_2(y_2)\,dF_n(y_1,\mathbf{x}_1)\,dF_n(y_2,\mathbf{x}_2)\right|.$$

Let us recall that $L_{\mathbf{XX}}^{-1}$ is a generalized inverse of $L_{\mathbf{XX}}$. The last expression is no larger than (by the Cauchy–Schwarz inequality)

$$\begin{aligned}&(\boldsymbol{\theta}_1'\hat{A}_{00}\boldsymbol{\theta}_1)^{1/2}(\boldsymbol{\theta}_2'\hat{A}_{00}\boldsymbol{\theta}_2)^{1/2}\left[\int\{\tilde{\mathbf{B}}_k'(\mathbf{x}_1)L_{\mathbf{XX}}^{-1}\tilde{\mathbf{B}}_k(\mathbf{x}_2)\}^2\,dF_n(y_1,\mathbf{x}_1)\,dF_n(y_2,\mathbf{x}_2)\right]^{1/2}\\ &\quad= (\boldsymbol{\theta}_1'\hat{A}_{00}\boldsymbol{\theta}_1)^{1/2}(\boldsymbol{\theta}_2'\hat{A}_{00}\boldsymbol{\theta}_2)^{1/2}\{\operatorname{rank}(L_{\mathbf{XX}}^{-1})\}^{1/2}.\end{aligned}$$

So, if either $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$ (or both) is in the space which is orthogonal to Ω, then $\boldsymbol{\theta}_1'L_k\boldsymbol{\theta}_2=0$. This tells us

$$e_{kn}^{*2} = 1-\sup\left\{\boldsymbol{\theta}'L_k\boldsymbol{\theta}: \boldsymbol{\theta}\in\Omega, \boldsymbol{\theta}'\hat{A}_{00}\boldsymbol{\theta}=1\right\}.$$

Clearly, this optimization problem has a solution and hence we have proved the existence of the estimates of the optimal transformations.

(ii) Let us note that if $\hat{A}_{\mathbf{XX}}$ and $\hat{A}_{00}$ are nonsingular, then by (4.3c) and (6.5),

$$e_{kn}^{*2} = 1 - \sup\{\boldsymbol{\theta}'\hat{A}_{0\mathbf{X}}\hat{A}_{\mathbf{XX}}^{-1}\hat{A}_{\mathbf{X}0}\boldsymbol{\theta}: \boldsymbol{\theta} \in R^{k+q}, \boldsymbol{\theta}'\hat{A}_{00}\boldsymbol{\theta} = 1\}.$$

By the result of Okamoto (1973), this maximization problem has a unique solution with probability 1 if $\hat{A}_{00}$ and $\hat{A}_{\mathbf{XX}}$ are nonsingular. Hence our result is proved by noting that by Lemma 6.1, $\hat{A}_{00}$ and $\hat{A}_{\mathbf{XX}}$ are nonsingular except on an event whose probability goes to zero as $n \to \infty$.

(b) As we have noted in the proof of Lemma 6.2, the eigenvalues of U_k and W_k are the same and the eigenvalues of $\hat{U}_k$ and $\hat{W}_k$ are the same. Let $\mathbf{a}_{kj}$, $j = 1, \dots, k + q$, be the eigenvectors of the matrix W_k and $\hat{\mathbf{a}}_{kj}$, $j = 1, \dots, k + q$, be the eigenvectors of the matrix $\hat{W}_k$, with $\|\mathbf{a}_{kj}\| = \|\hat{\mathbf{a}}_{kj}\| = 1$ for all j, where $\|\cdot\|$ is the usual Euclidean norm. Then it is easy to see that $h_{kj}(y) = \mathbf{a}'_{kj}A_{00}^{-1/2}\mathbf{B}_k(y)$ and $\hat{h}_{kj}(y) = \hat{\mathbf{a}}'_{kj}\hat{A}_{00}^{-1/2}\mathbf{B}_k(y)$, $j = 1, \dots, k + q$, are the eigenfunctions of the operators U_k and $\hat{U}_k$, respectively. To prove our result we will use arguments similar to those used in proving Theorem 4.1, except now we will use them for matrices.

As in the proof of Theorem 4.1, let Γ be a positively oriented closed curve in the complex plane which encloses $\lambda_1, \lambda_{k1}, \dots, \lambda_{km}$ for $k > k_0$ with λ_2 and $\lambda_{k,m+1}$ outside Γ. Let

$$T_k = \sum_1^m \mathbf{a}_{kj}\mathbf{a}'_{kj} \quad \text{and} \quad \hat{T}_k = \sum_1^m \hat{\mathbf{a}}_{kj}\hat{\mathbf{a}}'_{kj}.$$

Then T_k and $\hat{T}_k$ are projection matrices. Let $r_k(\lambda) = (W_k - \lambda)^{-1}$ and $\hat{r}_k(\lambda) = (\hat{W}_k - \lambda)^{-1}$. It is clear that $r_k(\lambda)$ exists and is bounded for every $\lambda \in \Gamma$. By Lemma 6.2, $\hat{r}_k(\lambda)$ exists and $\hat{r}_k(\lambda)$, $\lambda \in \Gamma$, are uniformly bounded except on an event whose probability goes to zero as $n \to \infty$. Arguing the same way as in the proof of Theorem 4.1, we obtain:

$$(6.13) \qquad \hat{T}_k - T_k = -(2\pi i)^{-1}\int_\Gamma r_k(\lambda)(\hat{W}_k - W_k)\hat{r}_k(\lambda)\,d\lambda.$$

Multiplying both sides of (6.13) by the vector $\hat{\mathbf{a}}_{k1}$, we get

$$(6.14) \qquad \begin{aligned}\hat{\mathbf{a}}_{k1} &= T_k\hat{\mathbf{a}}_{k1} - (2\pi i)^{-1}\int_\Gamma r_k(\lambda)(\hat{W}_k - W_k)(\hat{\lambda}_{k1} - \lambda)^{-1}\,d\lambda\,\hat{\mathbf{a}}_{k1}\\ &= \sum_1^m \alpha_j\mathbf{a}_{kj} - F_k(\hat{W}_k - W_k)\hat{\mathbf{a}}_{k1}, \quad \text{say,}\end{aligned}$$

where $F_k = (2\pi i)^{-1}\int_\Gamma r_k(\lambda)(\hat{\lambda}_{k1} - \lambda)^{-1}\,d\lambda$ and $\alpha_j = \hat{\mathbf{a}}'_{k1}\mathbf{a}_{kj}$. Let us note that $\|F_k\| = O_P(1)$. We have already shown in the proof of Lemma 6.2 that $\|\hat{W}_k - W_k\| = o_P(k\delta_{nk})$. So

$$(6.15) \qquad \left\|\hat{\mathbf{a}}_{k1} - \sum_1^m \alpha_j\mathbf{a}_{kj}\right\| \le \|F_k\|\,\|(\hat{W}_k - W_k)\hat{\mathbf{a}}_{k1}\| = o_P(k\delta_{nk}).$$

Multiplying both sides of equation (6.14) by $\hat{A}_{00}^{-1/2}\mathbf{B}_k(y)$ and by noting that

$$h_{kj}(y) = \mathbf{a}'_{kj}A_{00}^{-1/2}\mathbf{B}_k(y) \quad \text{and} \quad \hat{h}_{kj}(y) = \hat{\mathbf{a}}'_{kj}\hat{A}_{00}^{-1/2}\mathbf{B}_k(y),$$

we get

$$\begin{aligned}\hat{h}_{k1}(y) &= \sum_1^m \alpha_j \mathbf{a}'_{kj}\hat{A}_{00}^{-1/2}\mathbf{B}_k(y) - \hat{\mathbf{a}}'_{k1}(\hat{W}_k - W_k)F_k\hat{A}_{00}^{-1/2}\mathbf{B}_k(y) \\ &= \sum_1^m \alpha_j h_{kj}(y) + \sum_1^m \alpha_j \mathbf{a}'_{kj}(\hat{A}_{00}^{-1/2} - A_{00}^{-1/2})\mathbf{B}_k(y) \\ &\quad - \hat{\mathbf{a}}'_{k1}(\hat{W}_k - W_k)F_k\hat{A}_{00}^{-1/2}\mathbf{B}_k(y).\end{aligned} \tag{6.16}$$

Equation (6.16) gives us

$$\begin{aligned}\left\|\hat{h}_{k1} - \sum_1^m \alpha_j h_{kj}\right\| &\le \sum_1^m |\alpha_j| \left\|\mathbf{a}'_{kj}(\hat{A}_{00}^{-1/2} - A_{00}^{-1/2})\right\| \|A_{00}\|^{1/2} \\ &\quad + \left\|\hat{\mathbf{a}}'_{k1}(\hat{W}_k - W_k)\right\| \|F_k\| \left\|\hat{A}_{00}^{-1/2}\right\| \|A_{00}\|^{1/2}.\end{aligned}$$

Since $\|A_{00}\| = O(k^{-1})$ and $\|\hat{A}_{00}^{-1/2}\| = O_P(k^{1/2})$ by Lemma 6.1, the last expression tells us

$$\begin{aligned}\left\|\hat{h}_{k1} - \sum_1^m \alpha_j h_{kj}\right\| &\le \sum_1^m |\alpha_j| \left\|\mathbf{a}'_{kj}(\hat{A}_{00}^{-1/2} - A_{00}^{-1/2})\right\| O(k^{-1/2}) \\ &\quad + \left\|\hat{\mathbf{a}}'_{k1}(\hat{W}_k - W_k)\right\| O_P(1).\end{aligned}$$

We will first show that

$$\left\|\hat{h}_{k1} - \sum_1^m \alpha_j h_{kj}\right\| = O_P\big((k/n)^{1/2}\big). \tag{6.17}$$

Since $|\alpha_j| \le 1$ for $j = 1, \dots, m$, (6.17) is proved if we can show that

$$\left\|\mathbf{a}'_{kj}(\hat{A}_{00}^{-1/2} - A_{00}^{-1/2})\right\| = O_P(kn^{-1/2}), \qquad j = 1, \dots, m, \tag{6.18a}$$

$$\left\|\hat{\mathbf{a}}'_{k1}(\hat{W}_k - W_k)\right\| = O_P\big((k/n)^{1/2}\big). \tag{6.18b}$$

Let us first prove (6.18a). By part (b) of Lemma 6.3,

$$\begin{aligned}&\left\|\mathbf{a}'_{kj}(\hat{A}_{00}^{-1/2} - A_{00}^{-1/2}) + \mathbf{a}'_{kj}\left\{\pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)(\hat{A}_{00} - A_{00})\rho(\lambda)\,d\lambda\right\}\right\| \\ &\quad = o_P(k^{5/2}\delta_{nk}^2) = o_P(kn^{-1/2}).\end{aligned}$$

For the moment let us assume that for any $\mathbf{u} \in R^{k+q}$, with $\|\mathbf{u}\| = 1$,

$$E\left\|\mathbf{u}'(\hat{A}_{00} - A_{00})\right\|^2 \le c_{20}(kn)^{-1} \quad \text{for some } c_{20} > 0. \tag{6.19}$$

We will now show that (6.18a) follows from (6.19). Let $\bar{\gamma}$ be the smallest eigenvalue of A_{00}. By part (a) of Lemma (6.1), $\bar{\gamma}^{-1} = O_P(k)$.

$$
\begin{aligned}
E\left\| \mathbf{a}'_{kj}\left\{\pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)(\hat{A}_{00}-A_{00})\rho(\lambda)\,d\lambda\right\}\right\|
&\le \pi^{-1}\int_0^\infty \lambda^{-1/2}E\left\|\mathbf{a}'_{kj}\rho(\lambda)(\hat{A}_{00}-A_{00})\right\|\,\|\rho(\lambda)\|\,d\lambda\\
&\le c_{20}(kn)^{-1/2}\pi^{-1}\int_0^\infty \lambda^{-1/2}\left\|\mathbf{a}'_{kj}\rho(\lambda)\right\|\,\|\rho(\lambda)\|\,d\lambda\\
&\le c_{20}(kn)^{-1/2}\pi^{-1}\int_0^\infty \lambda^{-1/2}\|\rho(\lambda)\|^2\,d\lambda\\
&\le c_{20}(kn)^{-1/2}\pi^{-1}\int_0^\infty \lambda^{-1/2}(\lambda+\bar{\gamma})^{-2}\,d\lambda\\
&\le c_{20}c_{19}(kn)^{-1/2}\bar{\gamma}^{-3/2} \qquad [\text{by } (6.11)]\\
&= O(1)(kn)^{-1/2}k^{3/2} = O(kn^{-1/2}).
\end{aligned}
$$

So the proof of (6.18a) would be complete if we prove (6.19). Let $\mathbf{u} \in R^{k+q}$, with $\|\mathbf{u}\| = 1$. Now, $\|\mathbf{u}'(\hat{A}_{00} - A_{00})\|^2$ can be written as

$$
l = \int g(y_1, y_2)\,d(F_n - F)(y_1)\,d(F_n - F)(y_2),
$$

where $g(y_1, y_2) = \mathbf{u}'\mathbf{B}_k(y_1)\mathbf{B}'_k(y_1)\mathbf{B}_k(y_2)\mathbf{B}'_k(y_2)\mathbf{u}$. Since the B-splines we consider here are normalized, $\sup\{\|\mathbf{B}_k(y)\|: 0 \le y \le 1\} \le 1$, and hence $g(y, y) \le \mathbf{u}'\mathbf{B}_k(y)\mathbf{B}'_k(y)\mathbf{u}$. Noting that $\int g(y_1, y_2)\,dF(y_1)\,dF(y_2) \ge 0$, we get

$$
\begin{aligned}
E(l) &= n^{-1}\int g(y,y)\,dF(y) - n^{-1}\int g(y_1,y_2)\,dF(y_1)\,dF(y_2)\\
&\le n^{-1}\int g(y,y)\,dF(y) \le n^{-1}\mathbf{u}'A_{00}\mathbf{u} \le n^{-1}\|A_{00}\| = O(n^{-1}k^{-1}).
\end{aligned}
$$

This proves (6.18a). Now let us prove (6.18b). Inequality (6.15) tells us that

$$
\begin{aligned}
\left\|\hat{\mathbf{a}}'_{k1}(\hat{W}_k - W_k) - \sum_1^m \alpha_j \mathbf{a}'_{kj}(\hat{W}_k - W_k)\right\| &\le \|\hat{W}_k - W_k\|\,\left\|\hat{\mathbf{a}}_{k1} - \sum_1^m \alpha_j \mathbf{a}_{kj}\right\|\\
&= o_P(k^2\delta_{nk}^2) = o_P\left((k/n)^{1/2}\right).
\end{aligned}
$$

So, (6.18b) is proved if we can show that $\|\mathbf{a}'_{kj}(\hat{W}_k - W_k)\| = O_P((k/n)^{1/2})$ for $j = 1, \dots, m$, and this can be done by using arguments very similar to the ones used in proving (6.18a). Let us just very briefly outline these arguments. In the proof of (6.18a), the main step has been to expand $\hat{A}_{00}^{-1/2} - A_{00}^{-1/2}$ as the sum of two matrices $Z_{11} + Z_{12}$, where Z_{11} is linear in $\hat{A}_{00} - A_{00}$ [more precisely $Z_{11} = -\pi^{-1}\int_0^\infty \lambda^{-1/2}\rho(\lambda)(\hat{A}_{00} - A_{00})\rho(\lambda)\,d\lambda$] and Z_{12} is the remainder. Then we showed that $\|\mathbf{a}'_{kj}Z_{11}\|^2 = O_P(k/n)$ and $\|\mathbf{a}'_{kj}Z_{12}\|^2 = o_P(k/n)$. Similarly, we can expand $\hat{W}_k - W_k$ as $Z_{21} + Z_{22}$, where Z_{21} is the collection of terms which are linear in $\hat{A}_{00} - A_{00}$, $\hat{A}_{0X} - A_{0X}$ and $\hat{A}_{XX} - A_{XX}$ and Z_{22} is the remainder. It can be shown that $\|\mathbf{a}'_{kj}Z_{21}\|^2 = O_P(k/n)$ and $\|\mathbf{a}'_{kj}Z'_{22}\|^2 = o_P(k/n)$.

Let us note that the first part of part (b) of this theorem follows from (6.17) once we note that $\Sigma_1^m|\alpha_j|^2 = 1 + O_P((k/n)^{1/2})$. The rest of the theorem can now be easily proved. □

Acknowledgments. The author would like to thank Professor David Tyler for introducing him to the subject of perturbation theory, Professor Leo Breiman for correcting a mistake, Professor Charles Stone for suggesting some improvements in the readability of this paper and an Associate Editor for many valuable suggestions which have led to a substantial improvement in the presentation of this paper.

REFERENCES

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations, *J. Roy. Statist. Soc. Ser. B* **26** 211–252.

BREIMAN, L. and FRIEDMAN, J. (1985). Estimating optimal transformations for multiple regression and correlation, *J. Amer. Statist. Assoc.* **80** 580–619.

BURMAN, P. (1990a). Estimation of generalized additive models. *J. Multivariate Anal.* **32** 230–255.

BURMAN, P. (1990b). Estimation of optimal transformations of variables using v-fold cross validation and repeated learning-testing methods. *Sankhyā Ser. A.*

CHATELIN, F. (1983). *Spectral Approximation of Linear Operators*. Academic, New York.

DE BOOR, C. (1978). *A Practical Guide to Splines*. Springer, New York.

GREENACRE, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic, New York.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KATO, T. (1976). *Perturbation Theory of Linear Operators*, 2nd ed. Springer, New York.

KOYAK, R. (1990). Consistency for ACE-type methods. *Ann. Statist.* **18** 742–757.

OKAMOTO, M. (1973). Discreteness of the eigenvalues of a quadratic form in a multivariate sample. *Ann. Statist.* **1** 763–765.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. Wiley, New York.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606.

WEINBERGER, H. F. (1974). *Variational Methods for Eigenvalue Approximations*. SIAM, Philadelphia.

DIVISION OF STATISTICS
469 KERR HALL
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616