

SOME ASYMPTOTIC THEORY FOR FUNCTIONAL REGRESSION AND CLASSIFICATION

BY FRITS RUYMGAART, JING WANG, SHIH-HSUAN WEI, AND LI YU

Texas Tech University

Exploiting an expansion for analytic functions of operators, the asymptotic distribution of an estimator of the functional regression parameter is obtained in a rather simple way; the result is applied to testing linear hypotheses. The expansion is also used to obtain a quick proof for the asymptotic optimality of a functional classification rule, given Gaussian populations.

1. Introduction

Certain functions of the covariance operator (such as the square root of a regularized inverse) are important components of many statistics employed for functional data analysis. If Σ is a covariances operator on a Hilbert space, $\widehat{\Sigma}$ a sample analogue of this operator, and φ a function on the complex plane which is analytic on a domain containing a contour around the spectrum of Σ , a tool of generic importance is the comparison of $\varphi(\widehat{\Sigma})$ and $\varphi(\Sigma)$ by means of a Taylor expansion

$$(1.1) \quad \varphi(\widehat{\Sigma}) = \varphi(\Sigma) + \dot{\varphi}_{\Sigma}(\widehat{\Sigma} - \Sigma) + \text{remainder}.$$

(It should be noted that $\dot{\varphi}_{\Sigma}$ is not in general equal to $\varphi'(\Sigma)$, where φ' is the numerical derivative of φ ; see also Section 3.) In this paper two further applications of the approximation in (1.1) will be given, both related to functional regression.

The first application (Section 4) concerns the functional regression estimator itself. Hall & Horowitz (2007) have shown that the IMSE of their estimator, based on a Tikhonov type regularized inverse, is rate optimal. In this paper, as a complementary result, the general asymptotic distribution is obtained, with application to testing linear hypothesis of arbitrary finite dimension, mentioned in Cardot *et al.* (2003) as an open problem: these authors concentrate on testing a simple null hypotheses. Cardot *et al.* (1999) establish convergence in probability and almost sure convergence of their estimator which is based on spectral cutoff regularization of the inverse of

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: Functional regression, Functional classification;, Expansions for analytic functions of operators, Delta-method, Testing linear hypotheses

the sample covariance operator. In the present paper the covariance structure of the Gaussian limit will be completely specified. The proof turns out to be routine thanks to a “delta-method” for $\varphi(\widehat{\Sigma}) - \varphi(\Sigma)$, which is almost immediate from (1.1).

The second application (Section 5) concerns functional classification, according to a slight modification of a method by Hastie *et al.* (1995), exploiting penalized functional regression. It will be shown that this method is asymptotically optimal (Bayes) when the two populations are represented by equivalent Gaussian distributions with the same covariance operator. The simple proof is based on an upper bound for the norm of $\varphi(\widehat{\Sigma}) - \varphi(\Sigma)$ which follows at once from (1.1).

Let us conclude this section with some comments and further references. The expansion in (1.1) can be found in Gilliam *et al.* (2009), and the ensuing delta-method is derived and applied to regularized canonical correlation in Cupidon *et al.* (2007). For functional canonical correlation see also Eubank & Hsing (2008), He *et al.* (2002), and Leurgans *et al.* (1993). When the perturbation ($\widehat{\Sigma} - \Sigma$ in the present case) commutes with Σ the expansion (1.1) can already be found in Dunford & Schwartz (1957, Chapter VII), and the derivative does indeed reduce to the numerical derivative. This condition is fulfilled only in very special cases, for instance when the random function, whose covariance operator is Σ , is a second order stationary process on the unit interval. In this situation the eigenfunctions are known and only the eigenvalues are to be estimated. This special case, that will not be considered here, is discussed in Johannes (2008) who in particular deals with regression function estimators and their IMSE is Sobolev norms, when the regression is such a stationary process. General information about functional data analysis can be found in the monographs by Ramsay & Silverman (1997, 2002) and Ferraty & Vieu (2006). Functional time series are considered in Bosq (2000); see also Mas (2000).

2. Preliminaries and introduction of the models

2.1. Preliminaries

As will be seen in the examples below it is expedient to consider functional data as elements in an abstract Hilbert space \mathbb{H} of infinite dimension, separable, and over the real numbers. Inner product and norm in \mathbb{H} will be denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ respectively. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X : \Omega \rightarrow \mathbb{H}$ a Hilbert space valued random variable (i.e. measurable with respect to the σ -field of Borel sets $\mathcal{B}_{\mathbb{H}}$ in \mathbb{H}), and $\eta : \Omega \rightarrow \mathbb{R}$ a real valued

random variable. For all that follows it will be sufficient to assume that

$$(2.1) \quad \mathbb{E}\|X\|^4 < \infty, \quad \mathbb{E}\eta^2 < \infty.$$

The mean and covariance operator of X will be denoted by

$$(2.2) \quad \mathbb{E}X = \mu_X, \quad \mathbb{E}(X - \mu_X) \otimes (X - \mu_X) = \Sigma_{X,X},$$

respectively, where $a \otimes b$ is the tensor product in \mathbb{H} . The Riesz representation theorem guarantees that these quantities are uniquely determined by the relations

$$(2.3) \quad \mathbb{E}\langle a, X \rangle = \langle a, \mu_X \rangle \quad \forall a \in \mathbb{H},$$

$$(2.4) \quad \mathbb{E}\langle a, X - \mu_X \rangle \langle X - \mu_X, b \rangle = \langle a, \Sigma_{X,X} b \rangle \quad \forall a \in \mathbb{H} \quad \forall b \in \mathbb{H},$$

see Laha & Rohatgi (1979). Throughout $\Sigma_{X,X}$ is assumed to be one-to-one.

Let \mathcal{L} denote the Banach space of all bounded linear operators $T : \mathbb{H} \rightarrow \mathbb{H}$, equipped with the norm $\|\cdot\|_{\mathcal{L}}$. An operator $U \in \mathcal{L}$ is called Hilbert-Schmidt if

$$(2.5) \quad \sum_{k=1}^{\infty} \|Ue_k\|^2 < \infty,$$

for any orthonormal basis e_1, e_2, \dots of \mathbb{H} . (The number in (2.5) is in fact independent of the choice of basis.) The subspace $\mathcal{L}_{HS} \subset \mathcal{L}$ of all Hilbert-Schmidt operators is a Hilbert space in its own right with the inner product

$$(2.6) \quad \langle U, V \rangle_{HS} = \sum_{k=1}^{\infty} \langle Ue_k, Ve_k \rangle,$$

again independent of the choice of basis. This inner product yields the norm

$$(2.7) \quad \|U\|_{HS}^2 = \sum_{k=1}^{\infty} \|Ue_k\|^2,$$

which is the number in (2.5). The tensor product for elements $a, b \in \mathbb{H}$ will be denoted by $a \otimes b$, and that for elements $U, V \in \mathcal{L}_{HS}$ by $U \otimes_{HS} V$.

The two problems to be considered in this paper both deal with cases where the best linear predictor of η in terms of X is linear:

$$(2.8) \quad \mathbb{E}(\eta|X) = \alpha + \langle X, f \rangle, \quad \alpha \in \mathbb{R}, \quad f \in \mathbb{H}.$$

Just as in the univariate case (Rao (1965, Section 4g)), we have the relation

$$(2.9) \quad \Sigma_{X,X}f = \mathbb{E}(\eta - \mu_\eta)(X - \eta_X) = \Sigma_{X,\eta}.$$

It should be noted that if $\Sigma_{X,X}$ is one-to-one and $\Sigma_{X,\eta}$ in its range, we can solve (2.9) and obtain

$$(2.10) \quad f = \Sigma_{X,X}^{-1} \Sigma_{X,\eta}.$$

Since the underlying distribution is arbitrary, the empirical distribution, given a sample $(X_1, \eta_1) \dots, (X_n, \eta_n)$ of independent copies of (X, η) , can be substituted for it. The minimization property is now the least squares property, the same formulas are obtained with $\mu_X, \Sigma_{X,X}, \mu_\eta$, and $\Sigma_{X,\eta}$ replaced with their estimators

$$(2.11) \quad \hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X},$$

$$(2.12) \quad \hat{\Sigma}_{X,X} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) \otimes (X_i - \bar{X}),$$

$$(2.13) \quad \hat{\mu}_\eta = \frac{1}{n} \sum_{i=1}^n \eta_i = \bar{\eta},$$

$$(2.14) \quad \hat{\Sigma}_{X,\eta} = \frac{1}{n} \sum_{i=1}^n (\eta_i - \bar{\eta}) \times (X_i - \bar{X}),$$

Let us next specify the two problems.

2.2. Functional regression estimation

The model here is

$$(2.15) \quad \eta = \alpha + \langle X, f \rangle + \varepsilon,$$

where ε is a real valued error variable and the following assumption is satisfied.

Assumption 2.1. The error variable has a finite second moment, and

$$(2.16) \quad \varepsilon \perp\!\!\!\perp X, \quad \mathbb{E} \varepsilon = 0, \quad \text{Var } \varepsilon = v^2 < \infty.$$

Example 2.1. The functional regression model in Hall & Horowitz (2007) is essentially obtained by choosing $\mathbb{H} = L^2(0, 1)$, so that the generic observation is given by

$$(2.17) \quad \eta = \alpha + \int_0^1 X(t)f(t) dt + \varepsilon.$$

Example 2.2. Mas & Pumo (2007) argue that in the situation of Example 2.1 the derivative X' of X may contain important information and should therefore be included. Hence these authors suggest to choose for \mathbb{H} the Sobolev space $W^{2,1}(0, 1)$ in which case the generic observation satisfies

$$(2.18) \quad \eta = \alpha + \int_0^1 X(t)f(t) dt + \int_0^1 X'(t)f'(t) dt + \varepsilon,$$

Example 2.3. Just as in the univariate case we have that the model

$$(2.19) \quad \eta = \alpha + \langle X, f \rangle^2 + \varepsilon,$$

quadratic in the inner product of \mathbb{H} , is in fact linear in the inner product of \mathcal{L}_{HS} , because

$$(2.20) \quad \langle X, f \rangle^2 = \langle X \otimes X, f \otimes f \rangle_{HS}.$$

We will not pursue this example here.

In the infinite dimensional case $\widehat{\Sigma}_{X,X}$ cannot be one-to-one, and in order to estimate f from the sample version of (2.10) a regularized inverse of Tikhonov type will be used, as in Hall & Horowitz (2007). Thus we arrive at the estimator (see also (2.12) and (2.14))

$$(2.21) \quad \begin{aligned} \widehat{f}_\delta &= (\delta I + \widehat{\Sigma}_{X,X})^{-1} \left(\frac{1}{n} \sum_{i=1}^n \eta_i (X_i - \overline{X}) \right) \\ &= (\delta I + \widehat{\Sigma}_{X,X})^{-1} \widehat{\Sigma}_{X,\eta}, \quad \text{for some } \delta > 0. \end{aligned}$$

Let us also introduce

$$(2.22) \quad f_\delta = (\delta I + \Sigma)^{-1} \Sigma f, \quad f \in \mathbb{H}.$$

In Section 4 the asymptotic distribution of this estimator will be obtained and the result will be applied to testing.

2.3. Functional classification

The method discussed here is essentially the one in Hastie *et al.* (1995) and Hastie *et al.* (2001, Sections 4.2 and 4.3). Let P_1 and P_2 be two probability distributions on $(\mathbb{H}, \mathcal{B}_{\mathbb{H}})$ with means μ_1 and μ_2 and common covariance operator Σ . Consider a random element $(I, X) : \Omega \rightarrow \{1, 2\} \times \mathbb{H}$ with distribution determined by

$$(2.23) \quad \begin{cases} \mathbb{P}\{X \in B | I = j\} = P_j(B), & B \in \mathcal{B}_{\mathbb{H}}, \\ \mathbb{P}\{I = j\} = \pi_j \geq 0, & \pi_1 + \pi_2 = 1. \end{cases}$$

In this case the distribution of X is $\pi_1 P_1 + \pi_2 P_2$, with mean

$$(2.24) \quad \mu_X = \pi_1 \mu_1 + \pi_2 \mu_2,$$

and covariance operator

$$(2.25) \quad \Sigma_{X,X} = \Sigma + \pi_1 \pi_2 (\mu_1 - \mu_2) \otimes (\mu_1 - \mu_2).$$

Hastie *et al.* (2001) now introduce the indicator response variables $\eta_j = 1_{\{j\}}(I)$, $j = 1, 2$, and assume that the η_j satisfy (2.8) for $\alpha_j \in \mathbb{R}$ and $f_j \in \mathbb{H}$. Note that

$$(2.26) \quad \mu_{\eta_j} = \mathbb{E}\eta_j = \pi_j,$$

$$(2.27) \quad \Sigma_{X,\eta_j} = \mathbb{E}\eta_j(X - \mu_X) = \begin{cases} \pi_1 \pi_2 (\mu_1 - \mu_2), & j = 1, \\ \pi_1 \pi_2 (\mu_2 - \mu_1), & j = 2. \end{cases}$$

Since η_j is Bernoulli we have, of course, $\mathbb{E}(\eta_j | X) = \mathbb{P}\{I = j | X\}$. Precisely as for matrices (Rao & Toutenburg (1995, Theorem A.18)) the inverse of the operator in (2.25) equals

$$(2.28) \quad \Sigma_{X,X}^{-1} = \Sigma^{-1} - \gamma \cdot \Sigma^{-1}((\mu_1 - \mu_2) \otimes (\mu_1 - \mu_2))\Sigma^{-1},$$

where $\gamma = \pi_1 \pi_2 / (1 + \pi_1 \pi_2 \langle \mu_1 - \mu_2, \Sigma^{-1}(\mu_1 - \mu_2) \rangle)$, provided that the following assumption is satisfied.

Assumption 2.2. The vector $\mu_1 - \mu_2$ lies in the range of Σ , i.e.

$$(2.29) \quad \Sigma^{-1}(\mu_1 - \mu_2) \text{ is well defined.}$$

It will also be assumed that

$$(2.30) \quad \pi_1 = \pi_2 = \frac{1}{2}.$$

Assuming (2.29), equation (2.10) can be solved, and yields after some algebra

$$(2.31) \quad f_j = \Sigma_{X,X}^{-1} \Sigma_{X,\eta_j} = \begin{cases} \gamma \Sigma^{-1}(\mu_1 - \mu_2), & j = 1, \\ \gamma \Sigma^{-1}(\mu_2 - \mu_1), & j = 2. \end{cases}$$

If only X , and not I , is observed the rule in Hastie *et al.* (2001) assigns X to P_1 if and only if

$$(2.32) \quad \mathbb{E}(\eta_1|X) > \mathbb{E}(\eta_2|X),$$

$$\Longleftrightarrow$$

$$(2.33) \quad \langle X - \mu_X, \Sigma^{-1}(\mu_1 - \mu_2) \rangle > (\pi_2 - \pi_1)/2\gamma.$$

Because of assumption (2.30) the rule reduces to

$$(2.34) \quad \langle X - \frac{1}{2}(\mu_1 + \mu_2), \Sigma^{-1}(\mu_1 - \mu_2) \rangle > 0.$$

Hastie *et al.* (2001) claim that in the finite dimensional case their rule reduces to Fisher's linear discriminant rule and to the usual rule when the distributions are normal. This remains in fact true in the present infinite dimensional case. Let us assume that

$$(2.35) \quad P_j = \mathcal{G}(\mu_j, \Sigma), \quad j = 1, 2,$$

where $\mathcal{G}(\mu, \Sigma)$ denotes a Gaussian distribution with mean μ and covariance operator Σ . It is well known (Feldman (1958), Hájek (1958), Grenander (1981)) that under Assumption 2.2 these Gaussian distributions are equivalent. This is important since there is no “Lebesgue measure” on \mathbb{H} (Skorokhod (1974)). However, now the densities of P_1 and P_2 with respect to P_1 can be considered; it is well known that

$$(2.36) \quad \frac{dP_1}{dP_2}(x) = e^{-\langle x - \frac{1}{2}(\mu_1 + \mu_2), \Sigma^{-1}(\mu_1 - \mu_2) \rangle}, \quad x \in \mathbb{H}.$$

This leads at once to (2.34) as an optimal (Bayes) rule, equal in appearance to the one for the finite dimensional case.

In most practical situations μ_1, μ_2 , and Σ are not known, but a training sample $(I_1, X_1), \dots, (I_n, X_n)$ of independent copies of (I, X) is given. Let

$$(2.37) \quad \bar{X}_j = \frac{1}{n_j} \sum_{i \in \mathcal{J}_j} X_i, \quad \mathcal{J}_j = \{i : I_i = j\}, \quad \#\mathcal{J}_j = n_j,$$

$$(2.38) \quad \widehat{\Sigma} = \frac{1}{n} \sum_{j=1}^2 \sum_{i \in \mathcal{J}_j} (X_i - \overline{X}_j) \otimes (X_i - \overline{X}_j),$$

and we have (cf. (2.25))

$$(2.39) \quad \widehat{\Sigma}_{X,X} = \widehat{\Sigma} + \frac{n_1 n_2}{n} (\overline{X}_1 - \overline{X}_2) \otimes (\overline{X}_1 - \overline{X}_2).$$

Once again the operator $\widehat{\Sigma}$ (and $\widehat{\Sigma}_{X,X}$ for that matter) cannot be one-to-one. In order to obtain an empirical analogue of the rule (2.33), Hastie *et al.* (1995) employ penalized regression, and Hastie *et al.* (2001) also suggest to use a regularized inverse. (The methods are related.) Here the latter method will be used and X will be assigned to P_1 if and only if

$$(2.40) \quad \langle X - \frac{1}{2}(\overline{X}_1 + \overline{X}_2), (\delta I + \widehat{\Sigma})^{-1}(\overline{X}_1 - \overline{X}_2) \rangle > 0.$$

Section 5 is devoted to showing that this rule is asymptotically optimal when Assumption (2.2) is fulfilled.

3. A review of some relevant operator theory

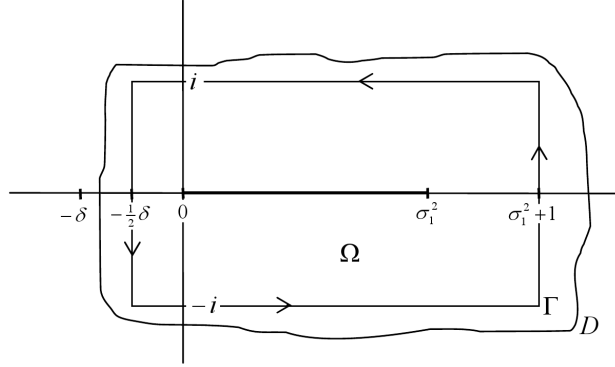
It is well known (Laha & Rohatgi (1979)) that the covariance operator Σ is nonnegative, Hermitian, of finite trace, and hence Hilbert-Schmidt and therefore compact. The assumption that Σ is one-to-one is equivalent to assuming that Σ is strictly positive. Consequently Σ has eigenvalues $\sigma_1^2 > \sigma_2^2 > \dots \downarrow 0$, all of finite multiplicity. If we let P_1, P_2, \dots be the corresponding eigenprojections, so that $\Sigma P_k = \sigma_k^2 P_k$, we have the spectral representation.

$$(3.1) \quad \Sigma = \sum_{k=1}^{\infty} \sigma_k^2 P_k, \text{ with } \sum_{k=1}^{\infty} P_k = I.$$

The spectrum of Σ equals $\sigma(\Sigma) = \{0, \sigma_1^2, \sigma_2^2, \dots\} \subset [0, \sigma_1^2]$. Let us introduce a rectangular contour Γ around the spectrum as in the picture above, where $\delta > 0$ is the regularization parameter in (2.21), and let Ω be the open region enclosed by Γ . Furthermore, let $D \supset (\Omega \cup \Gamma) = \overline{\Omega}$ be an open neighborhood of $\overline{\Omega}$, and suppose that

$$(3.2) \quad \varphi : D \rightarrow \mathbb{C} \text{ is analytic.}$$

We are interested in approximations of $\varphi(\widetilde{\Sigma}) = \varphi(\Sigma + \Pi)$, where $\Pi \in \mathcal{L}$ is a perturbation. The application we have in mind arises for $\Pi = \widehat{\Pi} = \widehat{\Sigma} - \Sigma$, and yields an approximation of $\varphi(\widehat{\Sigma})$; see also Watson (1983) for the matrix



case. Therefore we will *not* in general assume that Π and Σ will commute. In the special case where X is stationary, as considered in Johannes (2008), there exists a simpler estimator $\hat{\Sigma}$ of Σ , such that $\hat{\Sigma}$ and Π do commute, which results in a simpler theory; see also Remark (4.1).

The resolvent of Σ ,

$$(3.3) \quad R(z) = (zI - \Sigma)^{-1}, \quad z \in \rho(\Sigma) = [\sigma(\Sigma)]^c,$$

is analytic on the resolvent set $\rho(\Sigma)$, and the operator

$$(3.4) \quad \varphi(\Sigma) = \frac{1}{2\pi i} \oint_{\Gamma} \varphi(z) R(z) dz,$$

is well-defined. For the present operator Σ , as given in (3.1), the resolvent equals more explicitly

$$(3.5) \quad R(z) = \sum_{k=1}^{\infty} \frac{1}{z - \sigma_k^2} P_k.$$

Substitution of (3.5) in (3.4) and application of the Cauchy integral formula yields

$$(3.6) \quad \varphi(\Sigma) = \sum_{k=1}^{\infty} \varphi(\sigma_k^2) P_k.$$

Example 3.1. The two functions

$$(3.7) \quad \varphi_1(z) = \frac{1}{\delta + z}, \quad \varphi_2(z) = \frac{z}{\delta + z}, \quad z \in \mathbb{C} \setminus \{-\delta\},$$

are analytic on their domain that satisfies the conditions. With the help of these functions we may write (cf. (2.21) and (2.22))

$$(3.8) \quad \widehat{f}_\delta - f_\delta = (\varphi_1(\Sigma))\left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i(X_i - \overline{X})\right) + (\varphi_2(\widehat{\Sigma}) - \varphi_2(\Sigma))f, \quad f \in \mathbb{H},$$

and (cf. (2.40))

$$(3.9) \quad (\delta I + \widehat{\Sigma})^{-1}(\overline{X}_1 - \overline{X}_2) = \varphi_1(\widehat{\Sigma})(\overline{X}_1 - \overline{X}_2).$$

Regarding the following brief summary and slight extension of some of the results in Gilliam *et al.* (2008) we also refer to Dunford & Schwartz (1957), Kato (1966), and Watson (1983). Henceforth we will assume that

$$(3.10) \quad \|\Pi\|_{\mathcal{L}} \leq \frac{\delta}{4}.$$

For such perturbations we have $\sigma(\widetilde{\Sigma}) = \sigma(\Sigma + \Pi) \subset \Omega$, so that the resolvent set of $\widetilde{\Sigma}$ satisfies

$$(3.11) \quad \rho(\widetilde{\Sigma}) = \rho(\Sigma + \Pi) \supset \Omega^c \supset \Gamma.$$

It should also be noted that

$$(3.12) \quad \|R(z)\|_{\mathcal{L}} = \sup_{k \in \mathbb{N}} \frac{1}{|z - \sigma_k^2|} \leq \frac{2}{\delta} \quad \forall z \in \Omega^c.$$

The basic expansion (similar to Watson (1983))

$$(3.13) \quad \widetilde{R}(z) = (zI - \widetilde{\Sigma})^{-1} = R(z) + \sum_{k=1}^{\infty} R(z)(\Pi R(z))^k, \quad z \in \Omega^c,$$

can be written as

$$(3.14) \quad \widetilde{R}(z) = R(z) + R(z)\Pi R(z)(I - \Pi R(z))^{-1},$$

useful for analyzing the error probability for $\delta \rightarrow 0$, as $n \rightarrow \infty$, and also as

$$(3.15) \quad \widetilde{R}(z) = R(z) + R(z)\Pi R(z) + R(z)(\Pi R(z))^2(I - \Pi R(z))^{-1},$$

useful for analyzing the convergence in distribution of the estimators.

Let us decompose the contour Γ into two parts $\Gamma_0 = \{-\frac{1}{2}\delta + iy : -1 \leq y \leq 1\}$ and $\Gamma_1 = \Gamma \setminus \Gamma_0$, write $M_\varphi = \max_{z \in \Gamma} |\varphi(z)|$, and observe that (3.10) and (3.12) entail that

$$(3.16) \quad \|(I - \Pi R(z))^{-1}\|_{\mathcal{L}} \leq 2, \quad z \in \Omega^c.$$

We now have

$$\begin{aligned}
(3.17) \quad & \left\| \frac{1}{2\pi i} \oint_{\Gamma} \varphi(z) R(z) (\Pi R(z))^k (I - \Pi R(z))^{-1} dz \right\|_{\mathcal{L}} \\
& \leq \frac{1}{\pi} M_{\varphi} \|\Pi\|_{\mathcal{L}}^k \oint_{\Gamma} \|R(z)\|_{\mathcal{L}}^{k+1} dz \\
& \leq \frac{1}{\pi} M_{\varphi} \|\Pi\|_{\mathcal{L}}^k \left\{ \int_{-1}^{+1} \left(\frac{1}{4} \delta^2 + y^2 \right)^{-\frac{1}{2}(k+1)} dy + \left| \int_{\Gamma_1} 1 \, dz \right| \right\} \\
& \leq \frac{2}{\pi} M_{\varphi} \|\Pi\|_{\mathcal{L}}^k \left\{ \left(\frac{4}{\delta} \right)^k + 5 + 2 \|\Sigma\|_{\mathcal{L}} \right\}, \quad k \in \mathbb{N}.
\end{aligned}$$

Multiplying both sides by $\varphi(z)$, taking $(\frac{1}{2\pi i}) \oint_{\Gamma}$, and using $0 < C < \infty$ as a generic constant that does not depend on Π or δ , (3.14) and (3.15) yield the following.

Theorem 3.1. Provided that (3.10) is fulfilled, we have

$$(3.18) \quad \|\varphi(\Sigma + \Pi) - \varphi(\Sigma)\|_{\mathcal{L}} \leq C M_{\varphi} \frac{\|\Pi\|_{\mathcal{L}}}{\delta},$$

$$(3.19) \quad \|\varphi(\Sigma + \Pi) - \varphi(\Sigma) - \dot{\varphi}_{\Sigma} \Pi\|_{\mathcal{L}} \leq C M_{\varphi} \left(\frac{\|\Pi\|_{\mathcal{L}}}{\delta} \right)^2,$$

where $\dot{\varphi}_{\Sigma} : \mathcal{L} \rightarrow \mathcal{L}$ is a bounded operator, given by

$$(3.20) \quad \dot{\varphi}_{\Sigma} \Pi = \sum_k \varphi'(\sigma_k^2) P_k \Pi P_k + \sum_{j \neq k} \sum \frac{\varphi(\sigma_k^2) - \varphi(\sigma_j^2)}{\sigma_k^2 - \sigma_j^2} P_j \Pi P_k.$$

Remark 3.1. If Σ and Π commute, so will P_k and Π , and $R(z)$ and Π , and the expressions simplify considerably. In particular (3.20) reduces to

$$(3.21) \quad \dot{\varphi}_{\Sigma} \Pi = \left(\sum_k \varphi'(\sigma_k^2) P_k \right) \Pi,$$

i.e. $\dot{\varphi}_{\Sigma} = \varphi'(\Sigma)$, where φ' is the numerical derivative of φ , and $\varphi'(\Sigma)$ should be understood in the sense of “functional calculus” as in (3.6). For the commuting case see Dunford & Schwartz (1957).

Let us now present some basic facts that are useful to subsequent statistical applications. Dauxois *et al.* (1982) have shown that there exists a Gaussian random element $\mathcal{G}_{\Sigma} : \Omega \rightarrow \mathcal{L}_{HS}$, such that $\sqrt{n}(\hat{\Sigma} - \Sigma) \xrightarrow{d} \mathcal{G}_{\Sigma}$, as

$n \rightarrow \infty$, in \mathcal{L}_{HS} . Because the identity map from \mathcal{L}_{HS} in \mathcal{L} is continuous, we may state

$$(3.22) \quad \sqrt{n}(\widehat{\Sigma} - \Sigma) \xrightarrow{d} \mathcal{G}_{\Sigma}, \text{ as } n \rightarrow \infty, \text{ in } \mathcal{L}_{HS} \Rightarrow \text{in } \mathcal{L},$$

by the continuous mapping theorem. This entails that

$$(3.23) \quad \|\widehat{\Pi}\|_{\mathcal{L}} = \|\widehat{\Sigma} - \Sigma\|_{\mathcal{L}} = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right), \text{ as } n \rightarrow \infty,$$

so that condition (3.10) is fulfilled with arbitrary high probability for n sufficiently large. Expansions (3.14) and (3.15) and the resulting inequalities hold true for $\widehat{\Sigma}$ replaced with $\widehat{\Sigma}(\omega) = \Sigma + \widehat{\Pi}(\omega)$ for $\omega \in \{\|\widehat{\Pi}\|_{\mathcal{L}} \leq \frac{\delta}{4}\}$.

Example 3.2. : application to asymptotic distribution theory. In this application $\delta > 0$ will be kept fixed: see also Section 4.1. It is based on the delta method for functions of operators (Cupidon *et al.* (2007)) which follows easily from (3.19). In conjunction with (3.22) this yields

$$(3.24) \quad \sqrt{n}(\varphi_2(\widehat{\Sigma}) - \varphi_2(\Sigma)) \xrightarrow{d} \dot{\varphi}_{2,\Sigma} \mathcal{G}_{\Sigma}, \text{ as } n \rightarrow \infty, \text{ in } \mathcal{L}.$$

In turn this yields

$$(3.25) \quad \sqrt{n}(\varphi_2(\widehat{\Sigma}) - \varphi_2(\Sigma))f \xrightarrow{d} (\dot{\varphi}_{2,\Sigma} \mathcal{G}_{\Sigma})f, \text{ as } n \rightarrow \infty, \text{ in } \mathbb{H},$$

for any $f \in \mathbb{H}$, by the continuous mapping theorem. This result will be used in Section 4.

Example 3.3. : application to classification. Here we will let $\delta = \delta(n) \downarrow 0$, as $n \rightarrow \infty$, and write $\varphi_{1,n}(z) = 1/(\delta(n) + z)$ to stress this dependence on sample sizes. Since

$$(3.26) \quad \max_{z \in \Gamma} |\varphi_{1,n}(z)| \leq \frac{1}{\delta(n)},$$

it is immediate from (3.18) that

$$(3.27) \quad \|\varphi_{1,n}(\widehat{\Sigma}) - \varphi_{1,n}(\Sigma)\|_{\mathcal{L}} = \mathcal{O}_p\left(\frac{1}{\delta^2(n)\sqrt{n}}\right), \text{ as } n \rightarrow \infty,$$

a result that will be used in Section 5.

4. Asymptotics for the regression estimator

4.1. The asymptotic distribution

The central limit theorem in Hilbert spaces entails at once

$$(4.1) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i (X_i - \mu) \xrightarrow{d} \mathcal{G}_0, \text{ as } n \rightarrow \infty, \text{ in } \mathbb{H},$$

where \mathcal{G}_0 is a zero mean Gaussian random variable in \mathbb{H} , and

$$(4.2) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \{(X_i - \mu) \otimes (X_i - \mu) - \Sigma\} \xrightarrow{d} \mathcal{G}_\Sigma, \text{ as } n \rightarrow \infty, \text{ in } \mathcal{L}_{HS},$$

where \mathcal{G}_Σ is a zero mean random variable in \mathcal{L}_{HS} . These convergence results remain true with μ replaced by \bar{X} and, because $\varepsilon \perp\!\!\!\perp X$ by assumption (2.16), we also have that jointly

$$(4.3) \quad \left[\begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i (X_i - \bar{X}) \\ \sqrt{n}(\hat{\Sigma} - \Sigma) \end{array} \right] \xrightarrow{d} \left[\begin{array}{c} \mathcal{G}_0 \\ \mathcal{G}_\Sigma \end{array} \right], \text{ as } n \rightarrow \infty, \text{ in } \mathbb{H} \times \mathcal{L}_{HS},$$

where \mathcal{G}_Σ is the same in (3.22), and

$$(4.4) \quad \mathcal{G}_0 \perp\!\!\!\perp \mathcal{G}_\Sigma.$$

Because the limiting variables are generated by the sums of iid variables on the left in (4.1) and (4.2) we have

$$(4.5) \quad \begin{aligned} \mathbb{E}\mathcal{G}_0 \otimes \mathcal{G}_0 &= \mathbb{E}\{\varepsilon(X - \mu)\} \otimes \{\varepsilon(X - \mu)\} \\ &= (\mathbb{E}\varepsilon^2) \mathbb{E}(X - \mu) \otimes (X - \mu) \\ &= v^2 \Sigma, \end{aligned}$$

$$(4.6) \quad \mathbb{E}\mathcal{G}_\Sigma \otimes_{HS} \mathcal{G}_\Sigma = \mathbb{E}\{(X - \mu) \otimes (X - \mu) - \Sigma\} \otimes_{HS} \{(X - \mu) \otimes (X - \mu) - \Sigma\},$$

for the respective covariance structures. These are important to further specify the limiting distribution of the regression estimator as will be seen in Section 4.2.

Let us write, for brevity,

$$(4.7) \quad \hat{f}_\delta - f_\delta = U_n + V_n,$$

where, according to (2.21) and (2.22),

$$(4.8) \quad U_n = (\varphi_1(\widehat{\Sigma})) \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i (X_i - \overline{X}) \right),$$

$$(4.9) \quad V_n = (\varphi_2(\widehat{\Sigma}) - \varphi_2(\Sigma))f.$$

Note that φ_1 and φ_2 depend on δ .

With statistical applications in mind, it would be interesting if there would exist numbers $a_n \uparrow \infty$ and $\delta(n) \downarrow 0$, as $n \rightarrow \infty$, such that

$$(4.10) \quad a_n(\widehat{f}_{\delta(n)} - f) \xrightarrow{d} \mathcal{H}, \text{ as } n \rightarrow \infty, \text{ in } \mathbb{H},$$

where \mathcal{H} is a nondegenerate random vector. First we will show that when $f = 0$, such numbers do not exist. In this case $\widehat{f}_\delta - f_\delta = \widehat{f}_\delta = U_n$, so let us assume that

$$(4.11) \quad a_n U_n = a_n(\delta(n)I + \Sigma)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i (X_i - \mu) \right) \xrightarrow{d} \mathcal{H}, \text{ as } n \rightarrow \infty,$$

for some nondegenerate (i.e. not concentrated on a submanifold) random element \mathcal{H} in \mathbb{H} . If this were true we'd have, on the one hand,

$$(4.12) \quad \langle a_n U_n, p_k \rangle \xrightarrow{d} \langle \mathcal{H}, p_k \rangle, \text{ as } n \rightarrow \infty,$$

where the limit is nondegenerate in \mathbb{R} , and on the other hand

$$(4.13) \quad \langle a_n U_n, p_k \rangle = \frac{a_n/\sqrt{n}}{\delta(n) + \sigma_k^2} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \langle X_i - \mu, p_k \rangle,$$

where $n^{-1/2} \sum_{i=1}^n \varepsilon_i \langle X_i - \mu, p_k \rangle \xrightarrow{d} Z_k \stackrel{d}{=} N(0, v^2 \sigma_k^2)$, cf. (2.16). Therefore, and because $\delta(n) \rightarrow 0$, (4.12) and (4.13) can only be compatible if $a_n \sim c\sqrt{n}$, for some $0 < c < \infty$. Taking $c = 1$ we henceforth may assume that $a_n = \sqrt{n}$.

Relation (4.11) now entails the existence of a number x_0 , such that

$$(4.14) \quad \mathbb{P}\{n\|U_n\|^2 \geq x_0\} \rightarrow \mathbb{P}\{\|\mathcal{H}\|^2 \geq x_0\} < \frac{1}{4}.$$

According to (4.12) we have, however,

$$\begin{aligned}
(4.15) \quad \mathbb{P}\{n\|U_n\|^2 \geq x_0\} &\geq \mathbb{P}\{n\langle U_n, p_k \rangle^2 \geq x_0\} \rightarrow \mathbb{P}\left\{\left(\frac{v}{\sigma_k}\right)^2 Z^2 \geq x_0\right\} \\
&= \mathbb{P}\left\{Z^2 \geq x_0 \left(\frac{\sigma_k}{v}\right)^2\right\}, \text{ as } n \rightarrow \infty,
\end{aligned}$$

where $Z \stackrel{d}{=} N(0, 1)$. This holds true for every $k \in \mathbb{N}$, and since $\sigma_k^2 \downarrow 0$, as $k \uparrow \infty$, this entails that the l.h.s in (4.15) must be larger than $\frac{1}{2}$, for n sufficiently large, which yields a contradiction with (4.14).

For $f \neq 0$ we are inclined to decompose $\widehat{f}_{\delta(n)} - f$ according to (4.8) and (4.9) and write $a_n(\widehat{f}_{\delta(n)} - f) = a_n U_n + a_n V_n + a_n(f_{\delta(n)} - f)$, where the third term is deterministic. As before, $a_n U_n$ cannot converge. The sum of the three terms together, however, could possibly still converge due to some cancellation. We will not pursue this further and focus on the asymptotics of $\widehat{f}_\delta - f_\delta$, for fixed but arbitrary $\delta > 0$. The following basic result is immediate from (4.3), (4.7)-(4.9), the continuous mapping theorem, and the delta-method as in (3.25).

Theorem 4.1. For fixed $\delta > 0$ we have

$$(4.16) \quad \sqrt{n}(\widehat{f}_\delta - f_\delta) \xrightarrow{d} \mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2, \text{ as } n \rightarrow \infty, \text{ in } \mathbb{H},$$

where $\mathcal{H}_1 = \varphi_1(\Sigma)\mathcal{G}_0$ and $\mathcal{H}_2 = (\dot{\varphi}_{2,\Sigma}\mathcal{G}_\Sigma)f$ are zero mean Gaussian random elements, and $\mathcal{H}_1 \perp\!\!\!\perp \mathcal{H}_2$.

Further information about the structure of the covariance operator of the random vector \mathcal{H} on the right in (4.16) will be needed in order to be able to exploit the theorem for statistical inference. This will be addressed in the next section.

4.2. Further specification of the limiting distribution

It follows from (4.5) that \mathcal{G}_0 has a Karhunen-Loève expansion

$$(4.17) \quad \mathcal{G}_0 = \sum_{i=1}^{\infty} v\sigma_j Z_j p_j,$$

where the real valued random variables

$$(4.18) \quad Z_j (j \in \mathbb{N}) \text{ are iid } N(0, 1).$$

Accordingly \mathcal{H}_1 in (4.16) can be further specified as

$$(4.19) \quad \mathcal{H}_1 = \varphi_1(\Sigma)\mathcal{G}_0 = v \sum_{j=1}^{\infty} \frac{\sigma_j}{\delta + \sigma_j^2} Z_j p_j.$$

The Gaussian operator in (4.2) has been investigated in Dauxois *et al.* (1982) and here we will briefly summarize some of their results in our notation. By evaluating the inner product in \mathcal{L}_{HS} in the basis p_1, p_2, \dots it follows from (4.6) that

$$(4.20) \quad \begin{aligned} & \mathbb{E}\langle p_j \otimes p_k, \mathcal{G}_\Sigma \rangle_{HS} \langle \mathcal{G}_\Sigma, p_\alpha \otimes p_\beta \rangle_{HS} \\ &= \mathbb{E}\langle p_j, \langle X - \mu, p_k \rangle (X - \mu) - \sigma_k^2 p_k \rangle \times \langle p_\alpha, \langle X - \mu, p_\beta \rangle (X - \mu) - \sigma_\beta^2 p_\beta \rangle \\ &= \mathbb{E}\langle X - \mu, p_j \rangle \langle X - \mu, p_k \rangle \langle X - \mu, p_\alpha \rangle \langle X - \mu, p_\beta \rangle - \delta_{j,k} \delta_{\alpha,\beta} \sigma_k^2 \sigma_\beta^2. \end{aligned}$$

This last expression does not in general further simplify. However, if we assume that the regressor X satisfies

$$(4.21) \quad X \stackrel{d}{=} \text{Gaussian}(\mu, \Sigma),$$

it can be easily seen that the

$$(4.22) \quad \langle X - \mu, p_j \rangle \stackrel{d}{=} N(0, \sigma_j^2) \text{ are independent,}$$

so that the expression in (4.20) equals zero if $(j, k) \neq (\alpha, \beta)$. As in Dauxois *et al.* (1982) we obtain in this case

$$(4.23) \quad \mathbb{E}\langle p_j \otimes p_k, \mathcal{G}_\Sigma \rangle_{HS} \langle \mathcal{G}_\Sigma, p_\alpha \otimes p_\beta \rangle_{HS} = \begin{cases} 0, & (j, k) \neq (\alpha, \beta), \\ v_{j,k}^2, & (j, k) = (\alpha, \beta), \end{cases}$$

where

$$(4.24) \quad v_{j,k}^2 = \begin{cases} 2\sigma_j^4, & j = k, \\ \sigma_j^2 \cdot \sigma_k^2, & j \neq k. \end{cases}$$

Consequently the $p_j \otimes p_k$ ($j \in \mathbb{N}, k \in \mathbb{N}$) are an orthonormal basis of eigenvectors of the covariance operator of \mathcal{G}_Σ with eigenvalues $v_{j,k}^2$. Hence \mathcal{G}_Σ has the Karhunen-Loève expansion (in \mathcal{L}_{HS})

$$(4.25) \quad \mathcal{G}_\Sigma = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{j,k} Z_{j,k} p_j \otimes p_k,$$

where the random variables

$$(4.26) \quad Z_{j,k} (j \in \mathbb{N}, k \in \mathbb{N}) \text{ are iid } N(0, 1).$$

Let us, for brevity write (see (3.7) for φ_2)

$$(4.27) \quad w_{j,k} = \begin{cases} \frac{\varphi_2(\sigma_k^2) - \varphi_2(\sigma_j^2)}{\sigma_k^2 - \sigma_j^2}, & j \neq k \\ \varphi_2'(\sigma_j^2), & j = k \end{cases} \\ = \frac{\delta}{(\delta + \sigma_j^2)(\delta + \sigma_k^2)}, \text{ for all } j, k \in \mathbb{N}.$$

Then we have, exploiting (4.25)

$$(4.28) \quad \mathcal{H}_2 = (\dot{\varphi}_{2,\Sigma} \mathcal{G})f \\ = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} P_j \mathcal{G}_{\Sigma} P_k f \\ = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} P_j \left(\sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} v_{\alpha,\beta} Z_{\alpha,\beta} p_{\alpha} \otimes p_{\beta} \right) P_k f \\ = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} v_{j,k} Z_{j,k} \langle f, p_k \rangle p_j.$$

Summarizing we have the following result.

Corollary 4.1. The random element \mathcal{H}_1 , on the right in (4.16) can be represented by (4.19). If we assume that the regressor $X \stackrel{d}{=} \text{Gaussian}(\mu, \Sigma)$, the random element \mathcal{H}_2 on the right in (4.16) can be represented by (4.28), where the $Z_{j,k}$ in (4.26) are stochastically independent of the Z_j in (4.18).

4.3. Asymptotics under the null hypothesis

Let us recall that f_{δ} is related to f according to (2.22) so that the equivalence

$$(4.29) \quad H_0 : f_{\delta} = 0 \iff f = 0,$$

where again $\delta > 0$ is fixed, holds true. Under the null hypothesis in (4.29) it is immediate from Theorem 4.1 that

$$(4.30) \quad n \|\widehat{f_{\delta}}\|^2 \xrightarrow{d} \|\mathcal{H}\|^2 = \|\mathcal{H}_1\|^2, \text{ as } n \rightarrow \infty,$$

and from (4.19) that

$$(4.31) \quad \|\mathcal{H}_1\|^2 \stackrel{d}{=} v^2 \sum_{j=1}^{\infty} \frac{\sigma_j^2}{(\delta + \sigma_j^2)^2} Z_j^2.$$

Hence the test statistic on the left in (4.30) has a complicated chi-square type distribution with unknown parameters $v^2, \sigma_1^2, \sigma_2^2, \dots$. Let $F(v^2, \sigma_1^2, \sigma_2^2, \dots; x)$, $x \geq 0$, be the c.d.f of the random variable on the right in (4.31). For any collection of consistent estimators $\hat{v}^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots$ for the above mentioned parameters let us write $\hat{F}(x) = F(\hat{v}^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots; x)$, $x \geq 0$. The estimators $\hat{\sigma}_k^2$ can be obtained as the eigenvalues of $\hat{\Sigma}$; these are necessarily finite in number. To estimate v^2 we may use

$$(4.32) \quad \hat{v}^2 = \frac{1}{n} \sum_{i=1}^n \{\eta_i - \langle X_i - \bar{X}, \hat{f}_{\delta(n)} \rangle\}^2,$$

where $\hat{f}_{\delta(n)}$ is as in (2.6) with δ replaced by $\delta(n) \downarrow 0$, as $n \rightarrow \infty$. Note that we may choose $\delta(n)$ different from δ . We now have the following result.

Theorem 4.2. An approximate level $\alpha \in (0, 1)$ test for testing H_0 in (4.29) rejects the null hypothesis for $n\|\hat{f}_\delta\|^2 > \hat{F}^{-1}(1 - \alpha)$.

An immediate generalization of the hypothesis in (4.29) is

$$(4.33) \quad H_0 : f_\delta = f_{\delta,0} = (\varphi_2(\Sigma))f_0 \iff f = f_0,$$

for some given $f_0 \in \mathbb{H}$. This hypothesis is in principle of interest for confidence sets. Of course, testing (4.33) can be reduced to testing (4.29) by replacing the η_i with

$$(4.34) \quad \eta'_i = \eta_i - \langle X_i, f_{\delta,0} \rangle = \alpha + \langle X_i, f - f_{\delta,0} \rangle + \varepsilon_i,$$

and then using the estimator

$$(4.35) \quad \hat{f}'_\delta = (\delta I + \hat{\Sigma})^{-1} \left(\frac{1}{n} \sum_{i=1}^n \eta_{i,0} (X_i - \bar{X}) \right).$$

Under (4.33) we have $f - f_{\delta,0} = 0$ and consequently

$$(4.36) \quad n\|\hat{f}'_\delta\|^2 \xrightarrow{d} \|\mathcal{H}_1\|^2, \text{ as } n \rightarrow \infty,$$

where $\|\mathcal{H}_1\|^2$ has the same distribution as in (4.31). Related results can be found in Cardot *et al.* (2003).

Another generalization of the hypothesis in (4.29) is

$$(4.37) \quad H_0 : f_\delta \in \mathbb{M} = [q_1, \dots, q_M],$$

where q_1, \dots, q_M are orthonormal vectors in \mathbb{H} . Let Q and Q^\perp denote the orthogonal projection onto \mathbb{M} and \mathbb{M}^\perp respectively, and note that $f_\delta \in \mathbb{M}$ if and only if $\|Q^\perp f_\delta\| = 0$. A natural test statistic in this case appears to be

$$(4.38) \quad n\|Q^\perp \widehat{f}_\delta\|^2 \xrightarrow{d} \|Q^\perp \mathcal{H}\|^2, \text{ as } n \rightarrow \infty,$$

where the limiting result in (4.38) is under H_0 in (4.37).

The distribution on the right in (4.38) is rather complicated if q_1, \dots, q_M remain arbitrary. But if we are willing to assume (4.21) it follows from (4.28) that

$$(4.39) \quad \begin{aligned} \|Q^\perp \mathcal{H}\|^2 &= \|Q^\perp \mathcal{H}_1\|^2 + \|Q^\perp \mathcal{H}_2\|^2 + 2\langle Q^\perp \mathcal{H}_1, Q^\perp \mathcal{H}_2 \rangle \\ &= v^2 \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{\sigma_j}{\delta + \sigma_j^2} \frac{\sigma_k}{\delta + \sigma_k^2} Z_j Z_k \langle Q^\perp p_j, Q^\perp p_k \rangle \\ &\quad + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} w_{j,k} w_{\alpha,\beta} v_{j,k} v_{\alpha,\beta} Z_{j,k} Z_{\alpha,\beta} \\ &\quad \quad \times \langle f, p_k \rangle \langle f, p_\beta \rangle \langle Q^\perp p_j, Q^\perp p_\alpha \rangle \\ &\quad + 2v \sum_{j=1}^{\infty} \frac{\sigma_j}{\delta + \sigma_j^2} Z_j \sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} w_{\alpha,\beta} v_{\alpha,\beta} Z_{\alpha,\beta} \\ &\quad \quad \times \langle f, p_\beta \rangle \langle Q^\perp p_j, Q^\perp p_\alpha \rangle. \end{aligned}$$

The numbers $w_{j,k}$ are known, the eigenvalues σ_j^2 and eigenvectors p_j can in principle be estimated by the corresponding quantities of the sample covariance operator $\widehat{\Sigma}$, the unknown regression function f should be estimated by \widehat{f}_{δ_n} as in (4.32), an estimator of v^2 is given by \widehat{v}^2 in (4.32). Since we are in the Gaussian case a natural estimator of $\sigma_{j,k}$ will be

$$(4.40) \quad \widehat{v}_{j,k}^2 = \begin{cases} 2\widehat{\sigma}_j^4, & j = k, \\ \widehat{\sigma}_j^2 \cdot \widehat{\sigma}_k^2, & j \neq k, \end{cases}$$

as can be seen from (4.24). In principle it is now possible to obtain an approximate level α test in a similar manner as in Theorem 4.1, but it is clear that the procedure is extremely complicated.

A simplification is possible if we are willing to modify the hypothesis in (4.37) and use a so-called neighborhood hypothesis. This notion has a rather long history, and has been investigated by Hodges & Lehmann (1954) for certain parametric models. Dette & Munk (1998) have rekindled the interest in it by an application in nonparametric regression. In the present context we might replace (4.37) with the neighborhood hypothesis

$$(4.41) \quad H_{0,\varepsilon} : \|Q^\perp f_\delta\|^2 \leq \varepsilon^2, \text{ for some } \varepsilon > 0.$$

It is known from the literature that the advantage of using a neighborhood hypothesis is not only that such a hypothesis might be more realistic and that the asymptotics are much simpler, but also that without extra complication we might interchange null hypothesis and alternative. This means in the current situation that we might as well test the null hypothesis

$$(4.42) \quad H'_{0,\varepsilon} : \|Q^\perp f_\delta\|^2 \geq \varepsilon^2, \text{ for some } \varepsilon > 0,$$

which could be more suitable, in particular in goodness-of-fit problems.

The functional $g \mapsto \|Q^\perp g\|^2$, $g \in \mathbb{H}$, has a Fréchet derivative at f_δ given by the functional $g \mapsto 2\langle g, Q^\perp f_\delta \rangle$, $g \in \mathbb{H}$. Therefore the delta-method in conjunction with Theorem 4.1 entails the following

Corollary 4.2. Under the hypothesis in (4.37) we have

$$(4.43) \quad \sqrt{n}\{\|Q^\perp \hat{f}_\delta\|^2 - \|Q^\perp f_\delta\|^2\} \xrightarrow{d} 2\langle \mathcal{H}, Q^\perp f_\delta \rangle, \text{ as } n \rightarrow \infty.$$

The limiting distribution on the right in (4.43) is normal with mean zero and complicated variance.

$$(4.44) \quad \begin{aligned} \Delta^2 &= 4\mathbb{E}\langle \mathcal{H}, Q^\perp f_\delta \rangle^2 \\ &= 4\{\mathbb{E}\langle \mathcal{H}_1, Q^\perp f_\delta \rangle^2 + \mathbb{E}\langle \mathcal{H}_2, Q^\perp f_\delta \rangle^2\} \\ &= 4v^2 \sum_{j=1}^{\infty} \frac{\sigma_j^2}{(\delta + \sigma_j^2)^2} \langle p_j, Q^\perp f_\delta \rangle^2 \\ &\quad + 4\mathbb{E}\left\langle \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} P_j \left(\sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} v_{\alpha,\beta} Z_{\alpha,\beta} p_\alpha \otimes p_\beta \right) P_k f, Q^\perp f_\delta \right\rangle^2 \\ &= 4v^2 \sum_{j=1}^{\infty} \frac{\sigma_j^2}{(\delta + \sigma_j^2)^2} \langle p_j, Q^\perp f_\delta \rangle^2 \\ &\quad + 4 \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k}^2 v_{j,k}^2 \langle f, p_k \rangle^2 \langle p_j, Q^\perp f_\delta \rangle^2. \end{aligned}$$

Using the same estimators as suggested for the previous problem, including $\hat{f}_{\delta(n)}$ for f and \hat{f}_δ for f_δ , we arrive at a consistent estimator $\hat{\Delta}^2$ of Δ^2 . However, assuming (4.21) again we have the the following result which is immediate from (4.42) and (4.43).

Corollary 4.3. Assume that $\Delta^2 > 0$. Under $H_{0,\varepsilon}$ in (4.41) we have

$$(4.45) \quad T_n = \frac{1}{\hat{\Delta}} \sqrt{n}(\|Q^\perp \hat{f}_\delta\|^2 - \varepsilon^2) \xrightarrow{d} N(0, 1), \text{ as } n \rightarrow \infty,$$

and hence an asymptotic level $\alpha \in (0, 1)$ test is obtained when we reject the null hypothesis if $T_n > \Phi^{-1}(1 - \alpha)$, where Φ is the standard normal c.d.f.

Remark 4.1. In this paper we have dealt with the situation where Σ is entirely unknown. It has been observed in Johannes (2008) that if X is a stationary process on the unit interval, the eigenfunctions p_j of the covariance operator are always the same, known system of trigonometric functions, and only its eigenvalues σ_j^2 are unknown. Knowing the p_j leads to several simplifications.

In the first place, Σ can now be estimated by the expression on the right in (3.1) with only the σ_k^2 replaced with estimators. If $\tilde{\Sigma}$ is this estimator, it is clear that Σ and $\tilde{\Pi} = \tilde{\Sigma} - \Sigma$ commute, so that the derivative $\dot{\varphi}_{2,\Sigma}$ now simplifies considerably (see Remark 3.1). Secondly, we might consider the special case of H_0 in (4.37), where $q_j = p_j, j = 1, \dots, M$. We now have

$$(4.46) \quad f_\delta \in \mathbb{M} = [p_1, \dots, p_M] \iff f \in \mathbb{M},$$

so that even for fixed δ we can test the actual regression function.

In the third place, under the null hypothesis in (4.46), the number of unknown parameters in (4.39) reduces considerably because now $Q^\perp p_j = 0$ for $j = 1, \dots, M$.

When the p_j are known, in addition to all the changes mentioned above, also the limiting distribution of Σ differs from that of $\hat{\Sigma}$. Considering all these modifications that are needed, it seems better not to include this important special case in this paper.

4.4. Asymptotics under local alternatives

Again we assume that X is Gaussian. Suppose that

$$(4.47) \quad f = f_n = \tilde{f} + \frac{1}{\sqrt{n}}g, \text{ for } \tilde{f}, g \in \mathbb{H}.$$

For such f_n only minor changes in the asymptotics are required, because the conditions on the X_i and ε_i are still the same and don't change with n . Let us write

$$(4.48) \quad f_\delta = f_{n,\delta} = \tilde{f}_\delta + \frac{1}{\sqrt{n}}g_\delta,$$

where $\tilde{f}_\delta = (\delta I + \Sigma)^{-1}\tilde{f}$, $g_\delta = (\delta I + \Sigma)^{-1}\Sigma g$. The following is immediate from a minor modification of Theorem 4.1.

Theorem 4.3. For $f_\delta = f_{n,\delta}$ as in (4.48) we have

$$(4.49) \quad \sqrt{n}(\widehat{f}_\delta - \widetilde{f}_\delta) \xrightarrow{d} g_\delta + \mathcal{H}_1 + \widetilde{\mathcal{H}}_2,$$

where $\mathcal{H}_1 = \varphi_1(\Sigma)\mathcal{G}_0$ is the same as in (4.19), $\widetilde{\mathcal{H}}_2$ is obtained from \mathcal{H}_2 in (4.28) by replacing f with \widetilde{f} , and $\mathcal{H}_1 \perp\!\!\!\perp \widetilde{\mathcal{H}}_2$.

By way of an example let us apply this result to testing the neighborhood hypothesis and consider the asymptotics of the test statistics in (4.43) under the local alternatives $f_{n,\delta}$ in (4.48) with

$$(4.50) \quad \|Q^\perp \widetilde{f}_\delta\|^2 = \varepsilon^2, \quad g_\delta \perp \mathbb{M}, \quad \langle \widetilde{f}_\delta, g_\delta \rangle > 0.$$

Note that under such alternatives $\widehat{f}_{\delta(n)}$ is still a consistent estimator of \widetilde{f} for any $\delta(n) \downarrow 0$, as $n \rightarrow \infty$, and so is \widehat{f}_δ for f_δ . The following is immediate from Theorem 4.3.

Corollary 4.4. For $f_\delta = f_{n,\delta}$ as in (4.49) and (4.50), assuming (4.21) we have

$$(4.51) \quad T_n = \frac{\sqrt{n}}{\widehat{\Delta}} (\|Q^\perp \widehat{f}_\delta\|^2 - \varepsilon^2) \xrightarrow{d} N\left(\frac{2\langle g_\delta, Q^\perp \widetilde{f}_\delta \rangle}{\Delta}, 1\right), \text{ as } n \rightarrow \infty,$$

where $\widehat{\Delta}$ is the same as in (4.45).

5. Asymptotic optimality of the classification rule

In addition to Assumption 2.2 and (2.35) it will be assumed that the smoothness parameter $\delta = \delta(n)$ in (2.40) satisfies

$$(5.1) \quad \delta(n) \rightarrow 0 \quad \& \quad \delta(n) \gg n^{-1/4}, \text{ as } n \rightarrow \infty.$$

We will also assume that the sizes of the training samples n_1 and n_2 (see (2.37)) are deterministic and satisfy ($n = n_1 + n_2$)

$$(5.2) \quad 0 < \liminf_{n \rightarrow \infty} \frac{n_j}{n} \leq \limsup_{n \rightarrow \infty} \frac{n_j}{n} < 1.$$

Let us recall from (3.7) that $\varphi_1(z) = \varphi_{1,n}(z) = 1/\{\delta(n) + z\}$, $z \neq -\delta(n)$. Under these conditions the probability of misclassification equals $\mathbb{P}\{\langle X - \frac{1}{2}(\overline{X}_1 + \overline{X}_2), \varphi_{1,n}(\widehat{\Sigma})(\overline{X}_1 - \overline{X}_2) \rangle > 0 \mid X \stackrel{d}{=} \mathcal{G}(\mu_2, \Sigma)\}$. Let us note that

$$(5.3) \quad |\langle X - \frac{1}{2}(\overline{X}_1 + \overline{X}_2), \varphi_{1,n}(\widehat{\Sigma})(\overline{X}_1 - \overline{X}_2) \rangle$$

$$\begin{aligned}
& -\langle X - \frac{1}{2}(\mu_1 + \mu_2), \varphi_{1,n}(\Sigma)(\mu_1 - \mu_2) \rangle | \\
\leq & \frac{1}{2} \{ \|\bar{X}_1 - \mu_1\| + \|\bar{X}_2 - \mu_2\| \} \|\varphi_{1,n}(\Sigma)\|_{\mathcal{L}} \|\mu_1 - \mu_2\| \\
& + \|X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)\| \times \left[\|\varphi_{1,n}(\hat{\Sigma}) - \varphi_{1,n}(\Sigma)\|_{\mathcal{L}} \|\bar{X}_1 - \bar{X}_2\| \right. \\
& \left. + \{ \|\bar{X}_1 - \mu_1\| + \|\bar{X}_2 - \mu_2\| \} \|\varphi_{1,n}(\Sigma)\|_{\mathcal{L}} \right].
\end{aligned}$$

Since $\|\bar{X}_j - \mu_j\| = \mathcal{O}_p(n^{-1/2})$, $\|\varphi_{1,n}(\Sigma)\|_{\mathcal{L}} = \mathcal{O}(\delta^{-1}(n))$, and, according to (3.21),

$$(5.4) \quad \|\varphi_{1,n}(\hat{\Sigma}) - \varphi_{1,n}(\Sigma)\|_{\mathcal{L}} = \mathcal{O}_p\left(\frac{1}{\delta^2(n)\sqrt{n}}\right),$$

it follows from (5.1) that the limit of the misclassification probability equals

$$\begin{aligned}
(5.5) \quad & \lim_{n \rightarrow \infty} \mathbb{P}\left\{\langle X - \mu_2 - \frac{1}{2}(\mu_1 - \mu_2), \varphi_{1,n}(\Sigma)(\mu_1 - \mu_2) \rangle > 0\right\} \\
& = 1 - \Phi\left(\frac{1}{2}\langle \mu_1 - \mu_2, \Sigma^{-1}(\mu_1 - \mu_2) \rangle\right),
\end{aligned}$$

where Φ is the standard normal cdf.

For (5.5) we have used the well-known property of regularized inverses that $\|(\delta + \Sigma)^{-1}\Sigma f - f\| \rightarrow 0$, as $\delta \rightarrow 0$, for all $f \in \mathbb{H}$, and the fact that we may choose $f = \Sigma^{-1}(\mu_1 - \mu_2)$ by Assumption 2.2. Since rule (2.34) is optimal when parameters are known, we have obtained the following result.

Theorem 5.1. Let $P_j \stackrel{d}{=} \mathcal{G}(\mu_j, \Sigma)$ and let Assumption 2.2 and (5.1) be satisfied. Then the classification rule (2.40) is asymptotically optimal.

References

- [1] BOSQ, D. (2000). Linear processes in function spaces. *Lecture Notes in Statistics*. Springer, New York.
- [2] CARDOT, H., FERRATY, F., & SARDA, P. (1999). Functional linear model. *Statist. Probab. Letters* **45**, 11-22.
- [3] CARDOT, H., FERRATY, F., MAS, A. & SARDA, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.* **30**, 241-255.
- [4] CUPIDON, J., EUBANK, R., GILLIAM, D. & RUYMGAART, F. (2007). The delta-method for analytic functions of random operators with application to functional data. *Bernoulli*. **13**, 1179-1194.
- [5] DAUXOIS, J., POUSSE, A. & ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivar. Anal.* **12**, 136-154.
- [6] DETTE, H. & MUNK, A. (1998). Validation of linear regression models. *Ann. Statist.* **26**, 778-800.
- [7] DUNFORD, N. & SCHWARTZ, J. T. (1957). *Linear Operators, Part I: General Theory*. Interscience, New York.
- [8] EUBANK, R. L. & HSING, T. (2008). Canonical correlation for stochastic processes. *Stoch. Process. Appl.* **118**, 1634-1661.
- [9] FELDMAN, J. (1958). Equivalence and perpendicularity of Gaussian processes. *Pac. J. Math.* **8**, 699-708.
- [10] FERRATY, F. & VIEU, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York.
- [11] GILLIAM, D. S., HOHAGE, T., JI, X. & RUYMGAART, F. (2009). The Fréchet derivative of an analytic function of a bounded operator with some applications. *Int. J. Math. Sci.*, Article ID 239025.
- [12] GRENNANDER, U. (1981). *Abstract Inference*. Wiley, New York.
- [13] HALL, P. & HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70-91.
- [14] HÁJEK, J. (1958). On a property of normal distributions of any stochastic process. *Cz. Math. J.* **8**, 610-617.
- [15] HASTIE, T., BUJA, A. & TIBSHIRANI, R. (1995). Penalized discriminant analysis. *Ann. Statist.* **23**, 73-102.
- [16] HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (2001). *The Elements of Statistical Learning*. Springer, New York.
- [17] HE, G., MÜLLER, H.-G. & WANG, J. L. (2002) Functional canonical analysis for square integrable stochastic processes. *J. Multivar. Anal.* **85**, 54-77.
- [18] HODGES, J. L. & LEHMANN, E. L. (1954). Testing the approximate validity of statistical hypotheses. *J. Roy. Statist. Soc. B* **16**, 261-268.
- [19] JOHANNES, J. (2008). Privileged communication.
- [20] KATO, T. (1966). *Perturbation Theory for Linear Operators*. Springer, New York.
- [21] LAHA, R. G. & ROHATGI, V. K. (1979). *Probability Theory*. Wiley, New York.
- [22] LAX, P. D. (2002). *Functional Analysis*. Wiley-Interscience, New York.
- [23] LEURGANS, S. E., MOYEED R., A. & SLIVERMAN, B. W. (1993). Canonical correlation analysis when the data are curves. *J. Roy. Statist. Soc. B* **55**, 725-740.
- [24] MAS, A. (2000). Estimation d'Opérateurs de Corrélation de Processus Fonctionnels: Lois Limites, Tests, Déviations Modérées. Thèse de Doctorat, Université Paris VI.
- [25] MAS, A. & PUMO, B. (2006). Functional linear regression with derivatives. Tech. Report. Institut de Modélisation Mathématique de Montpellier.

- [26] RAMSAY, J. & SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed., Springer, New York.
- [27] RAO, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.
- [28] RAO, C. R. & TOUTENBURG, H. (1995). *Linear Models*. Springer, New York.
- [29] SKOROKHOD, A. V. (1974). *Integration in Hilbert Space*. Springer, New York.
- [30] WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York.

TEXAS TECH UNIVERSITY
LUBBOCK, TEXAS 79409, USA
E-MAIL: h.ruymgaart@ttu.edu
eliza.wang@ttu.edu

TEXAS TECH UNIVERSITY
LUBBOCK, TEXAS 79409, USA
E-MAIL: scottie.wei@ttu.edu