Journal of Multivariate Analysis **81**, 100–119 (2002)
doi:10.1006/jmva.2001.1994, available online at http://www.idealibrary.com on IDEAL®

Spatially Adaptive Splines for Statistical Linear Inverse Problems

Hervé Cardot

Unité Biométrie et Intelligence Artificielle, INRA Toulouse, 31326 Castanet-Tolosan cedex, France
E-mail: cardot@toulouse.inra.fr

Received November 24, 1999; published online October 17, 2001

This paper introduces a new nonparametric estimator based on penalized regression splines for linear operator equations when the data are noisy. A local roughness penalty that relies on local support properties of B-splines is introduced in order to deal with spatial heterogeneity of the function to be estimated. This estimator is shown to be consistent under weak conditions on the asymptotic behaviour of the singular values of the linear operator. Furthermore, in the usual nonparametric settings, it is shown to attain optimal rates of convergence. Then its good performances are confirmed by means of a simulation study. © 2001 Elsevier Science (USA)

AMS subject classifications: 45A05; 62G07; 65D07.

Key words and phrases: linear inverse problems; integral equations; deconvolution; regularization; local roughness penalties; spatially adaptive estimators; regression splines; convergence.

1. INTRODUCTION

Statistical linear inverse problems consist of indirect noisy observations of a parameter (a function generally) of interest. Such problems occur in many areas of science such as genetics with DNA sequences (Mendelsohn and Rice, 1982), optics and astronomy with image restoration (Craig and Brown, 1986), biology and natural sciences (Tikhonov and Goncharsky, 1987). Then, the data are a linear transform of an original signal f corrupted by noise, so that we have

$$Y_i = Kf(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where K is some known compact linear operator defined on a separable Hilbert space H (supposed in the following to be $L^2[0, 1]$, the space of square integrable functions defined on $[0,1]$) and ε_i is a white noise with unknown variance σ^2. These problems are also called ill-posed problems because the operator K is compact and consequently equation (1) can not be inverted directly since K^{-1} is not a bounded operator. The reader is

0047-259X/01 $35.00

referred to Tikhonov and Arsenin (1977) for a seminal book on ill-posed operator equations and O'Sullivan (1986) for a review of the statistical perspective on ill-posed problems. In the following, we will restrict ourself to integral equations with kernel $k(s, t)$,

$$Kf(t) = \int_0^1 k(s, t)\, f(s)\, ds, \qquad t \in [0, 1], \tag{2}$$

which include deconvolution

$$Kf(t) = \int_0^1 k(t-s)\, f(s)\, ds.$$

There is a vast literature in numerical analysis (Hansen 1998, Neumaier 1998 and references therein) and in statistics dealing with inverse problems (e.g., Wahba, 1977; Mendelsohn and Rice, 1982; Nychka and Cox, 1989; Abramovich and Silverman, 1998 among others). Actually, since model (1) can not be inverted directly, even if the data are not corrupted by noise, one has to regularize the estimator by adding a constraint in the estimation procedure (Tikhonov and Arsenin, 1977). The regularization can be linear and is generally based on a windowed singular value decomposition (SVD) of K. Indeed, since K is compact, it admits the following decomposition $K(.) = \sum_j \lambda_j \langle \phi_j, . \rangle\, \psi_j$, where $\{\phi_j\}_j$ and $\{\psi_j\}_j$ are orthonormal bases of H and the singular values are sorted by decreasing order, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant 0$. Most estimators of f proposed in the literature are based either on a finite rank, depending on the sample size, approximation of K achieved by truncating the basis expansions obtained by means of the SVD or by adding a regularization (smoothing) parameter to the eigenvalues that makes K invertible:

$$\hat{f}(t) = \sum_{j \geqslant 0} f_j \, \frac{\langle y, \psi_j \rangle}{\lambda_j}\, \phi_j. \tag{3}$$

The sequence of filtering coefficients f_j controls the regularity of the solution: $f_j = \lambda_j^2 / (\lambda_j^2 + \rho^2)$ for the Tikhonov method and $f_j = I_{[j \leqslant k]}$ for the truncated SVD method (see Hansen, 1998, for an exhaustive review of these methods). The rate of decay of the singular values indicates the degree of ill-posedness of the problem: the more the singular values decrease rapidly the more ill-posed the problem is.

Nevertheless, estimators based on the SVD have two main defaults. On the one hand, the basis functions depend explicitly on the operator K and not on the function of interest. For instance, it is well known that Fourier basis are the singular functions of K for deconvolution problems and that they can not provide a parsimonious approximation of the true function if it is smooth in some regions and rapidly oscillates in other regions. On the other hand, the usual regularization procedures do not allow to deal with spatial heterogeneity of the function to be recovered. Several authors have proposed spatially adaptive estimators based on wavelet decomposition (Donoho, 1995; Abramovich and Silverman, 1998) that attain minimax rates of convergence for particular operators K such as homogeneous operators. Our approach is quite different and relies on spline fitting with local roughness penalties.

Until now, spatially adaptive splines were computed by means of knots selection procedures that require sophisticated algorithms (Friedman, 1991; Stone *et al.*, 1997; Denison *et al.*, 1998). This paper does not address the topic of knots selection and the estimator proposed below is a penalized regression splines whose original idea traces back to O'Sullivan (1986) and Ruppert and Carroll (1999). Actually, Ruppert and Carroll's method consists in penalizing the jumps of the function at the interior knots, each being controlled by a smoothing parameter, in order to manage both the highly variable part and the smooth part of the estimator. In this article, a similar approach is proposed. Using the fact that B-spline functions have local supports and that the derivative of a B-spline of order q is the combination of two B-splines of order $q-1$ we are able to define local measures of the squared norm of a given order derivative of the function of interest. Thus the curvature of the estimator can be controlled locally by means of smoothing parameters associated to these local measures of roughness. Some asymptotic properties of the estimator are given. These local penalties are controlled by local smoothing parameters whose values must be chosen very carefully in practical situations in order to get accurate estimates. The generalized cross validation (GCV) criterion is widely used for nonparametric regression and generally allows to select "good" values of the smoothing parameter (Green and Silverman, 1994). Unfortunately, GCV seems to fail to select effective smoothing parameter values in the framework of adaptive splines for inverse problems by giving too often undersmoothed estimates. Further investigation is needed to cope with this important practical topic but that is beyond the scope of this paper. Nevertheless a small Monte Carlo experiment has been performed to show the potential of this new approach.

The organization of the paper is as follows. In Section 2, the spatial adaptive regression splines estimator is defined. In Section 3, upper bounds for the rates of convergence are given. The particular case where $Kf(t) = f(t)$ (the usual nonparametric framework) is also tackled and the spatially adaptive estimator is shown to attain optimal rates of convergence. Then, in Section 4, a simulation study compares the behavior of this estimator to the penalized regression splines proposed by O'Sullivan (1986). Finally, Section 5 gathers the proofs.

Splus programs for carrying out the estimation are available on request.

2. SPATIALLY ADAPTIVE SPLINE ESTIMATES

The estimator proposed below is based on spline functions. Let's now briefly recall the definition and some known properties of these functions. Suppose that q and k are integers and let S_{qk} be the space of spline functions defined on $[0, 1]$, of order q $(q \geqslant 2)$, with k equispaced interior knots. The set S_{qk} is then the set of functions s defined as :

- s is a polynomial of degree $q-1$ on each interval $[\frac{t-1}{k+1}, \frac{t}{k+1}]$, $t = 1, ..., k+1$;
- s is $q-2$ times continuously differentiable on $[0, 1]$.

The space S_{qk} is known to be of dimension $q+k$ and one can derive a basis by means of normalized B-splines $\{B^q_{kj}, j = 1, ..., q+k\}$ (see de Boor, 1978 or Dierckx, 1993). These functions are non negative and have local support

$$B^q_{kj}(x) = 0 \qquad \text{if} \quad x \notin [\delta_j, \delta_{j+q}], \tag{4}$$

where

$$\begin{aligned} &\delta_1 = \delta_2 = \cdots = \delta_q = 0, \\ &\delta_{q+j} = j/(k+1), \ j = 1, ..., k \\ &\delta_{q+k+1} = \cdots = \delta_{2q+k} = 1. \end{aligned} \tag{5}$$

Furthermore, a remarkable property of B-splines is that the derivative of a B-spline of order q can be expressed as a linear combination of two B-splines of order $q-1$. More precisely, if $s = \sum_{j=1}^{k+q} \theta_j B^q_{kj} = \mathbf{B}'_{qk}\boldsymbol{\theta}$ then

$$\begin{aligned} s' &= (q-1) \sum_{j=1}^{k+q-1} \frac{\theta_{j+1} - \theta_j}{\delta_{j+q} - \delta_{j+1}} B^{q-1}_{kj}, \\ &= \mathbf{B}'_{(q-1)\,k}\boldsymbol{\theta}^{(1)}, \end{aligned} \tag{6}$$

where B^{q-1}_{kj} is the jth normalized B-spline of $S_{(q-1)\,k}$ and $\mathbf{B}_{qk}$ is the vector of all the B-splines of S_{qk}. Let's define by $\mathbf{D}_{qk}$ the weighted differentiation $(k+q-1) \times (k+q)$ matrix which gives the coordinates in $S_{(q-1)\,k}$ of the derivative of a function of S_{qk}:

$$\boldsymbol{\theta}^{(1)} = \mathbf{D}_{qk}\boldsymbol{\theta}. \tag{7}$$

Then, by iterating this process, one can easily obtain the coordinates of a given order derivative of a function of S_{qk} by applying the $(k+q-m)\times(k+q)$ matrix $\boldsymbol{\Delta}^{(m)}$ defined as follows:

$$\begin{aligned}\boldsymbol{\theta}^{(m)} &= \mathbf{D}_{(q-m+1)\,k} \cdots \mathbf{D}_{qk}\boldsymbol{\theta} \\ &= \boldsymbol{\Delta}^{(m)}\boldsymbol{\theta} \end{aligned} \tag{8}$$

We consider a penalized least squares regression estimator with penalty proportional to the weighted squared norm of a given order m $(m<q-1)$ derivative of the functional coefficient, the effect of which being to give preference for a certain local degree of smoothness. Using the local support properties of B-splines, this adaptive roughness penalty is controlled by $k+q-m$ local positive smoothing parameters $\rho_1, \ldots, \rho_{k+q-m}$ that may take spatial heterogeneity into account. Our penalized B-splines estimate of f is thus defined as

$$\begin{aligned}\hat{f} &= \sum_{j=1}^{q+k} \hat{\theta}_j B_{kj}^q \\ &= \mathbf{B}'_{qk}\hat{\theta}, \end{aligned} \tag{9}$$

where $\hat{\theta}$ is a solution of the following minimization problem

$$\min_{\theta \in \mathscr{R}^{q+k}} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{q+k} \theta_j K B_{kj}^q(t_i) \right)^2 + \left\| \sum_{j=1}^{k+q-m} \rho_j \theta_j^{(m)} B_{kj}^{q-m} \right\|^2. \tag{10}$$

$\theta_j^{(m)}$ is the jth element of $\boldsymbol{\theta}^{(m)}$ and $\|.\|$ denotes the usual $L^2[0,1]$ norm.

Let $\mathbf{A}_n$ be the $n\times(q+k)$ design matrix with elements $KB_{kj}^q(t_i)$ and $\mathbf{C}_{qk}$ the $(k+q)\times(k+q)$ matrix whose generic element is the inner product between two B-splines:

$$[\mathbf{C}_{qk}]_{ij} = \int_0^1 B_{ki}^q(t)\, B_{kj}^q(t)\, dt.$$

Let us define

$$\mathbf{G}_{n,\rho} = \left(\frac{1}{n} \mathbf{A}'_n \mathbf{A}_n + \boldsymbol{\Delta}^{(m)'} \mathbf{I}_\rho \mathbf{C}_{(q-m)\,k} \mathbf{I}_\rho \boldsymbol{\Delta}^{(m)} \right),$$

where $\mathbf{I}_\rho$ is the diagonal matrix with diagonal elements ρ_j. Then, the solution $\hat{\boldsymbol{\theta}}$ of the minimization problem (10) is given by

$$\hat{\theta} = \mathbf{G}_{n,\rho}^{-1} \frac{1}{n} \mathbf{A}'_n \mathbf{Y}, \tag{11}$$

where $\mathbf{Y}$ is the vector of $\mathscr{R}^n$ with elements Y_i.

Remark 2.1. If $\rho_j = \rho$, $j = 1, ..., k+q-m$, then the estimator defined by (9) is the same as the estimator proposed by O'Sullivan (1986):

$$\min_{\theta \in \mathcal{R}^{q+k}} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{q+k} \theta_j K B_{kj}^q(t_i) \right)^2 + \rho^2 \left\| \sum_{j=1}^{k+q-m} \theta_j^{(m)} B_{kj}^{q-m} \right\|^2. \tag{12}$$

Furthermore, in the usual nonparametric settings (i.e $Kf(t) = f(t)$), these kind of penalized regression splines have already been used for different purposes. Kelly and Rice (1991) have used them to nonparametrically estimate dose-response curves and Cardot and Diack (1998) have demonstrated they could attain optimal rates of convergence. Besse *et al.* (1997) have performed the principal components analysis of unbalanced longitudinal data and Cardot (2000) has studied the asymptotic convergence of the principal components analysis of sampled noisy functional data.

Remark 2.2. The local penalty defined in (10) may be viewed as a kind of discrete version of the continuous penalty defined as

$$\int_0^1 \rho^2(t) \left(\sum_{j=1}^{k+q-m} \theta_j^{(m)} B_{kj}^{q-m}(t) \right)^2 dt,$$

the local roughness being continously controlled by function $\rho(t)$.

3. ASYMPTOTIC RESULTS

Our convergence results are derived according to the semi-norm induced by the linear operator K (named K-norm in the following)

$$\begin{aligned} \|f\|_K^2 &= \langle Kf, Kf \rangle, \qquad f \in L^2[0, 1], \\ &= \int_0^1 (Kf(t))^2 \, dF(t) \end{aligned} \tag{13}$$

and the empirical norm

$$\|f\|_{K,n}^2 = \frac{1}{n} \sum_{i=1}^{n} (Kf(t_i))^2, \qquad f \in L^2[0, 1]. \tag{14}$$

Then we have $\|f\|_K = 0$ if f belongs to the null space of K and thus can not be estimated. This norm allows us to measure the distance of the estimate from the recoverable part of function f and has been considered by Wahba

(1977). Let's define $\mathscr{K}(m)=\{\mathbf{B}'_{qk}\boldsymbol{\theta}\mid\boldsymbol{\Delta}^{(m)}\boldsymbol{\theta}=0\}$. It is easily seen that $\mathscr{K}(m)$ is the space of polynomial functions defined on $[0, 1]$ with degree less than m.

To ensure the existence and the convergence of the estimator we need the following assumptions on the regularity of f, on the repartition of the design points, the moments of the noise and on the operator K

(H.1) $f \in C^p[0, 1]$ and $0 < p < q$.

(H.2) The ε_i's are independent and distributed as ε where $\mathbb{E}\varepsilon = 0$, $\mathbb{E}\varepsilon^2 = \sigma^2 < \infty$.

(H.3) Let's denote by F_n the empirical distribution of the design sequence, $\{t_{j,n}, 1 \leqslant j \leqslant n\} \subset [0, 1]$ and suppose it converges to a design measure F that has a continuous, bounded, and strictly positive density h on $[0, 1]$. Furthermore, let's suppose that there exists a sequence $\{d_n\}$ of positive numbers tending to zero such that

$$\sup_{t\in[0,1]} |F(t)-F_n(t)| = O(d_n).$$

(H.4) The kernel $k(s, t)$ belongs to $L^2([0, 1]\times[0, 1])$ and, for fixed s, the function $t \mapsto k(s, t)$ is a continous function whose derivative belongs to $L^2([0, 1])$.

(H.5) It exists $C > 0$ such that $\forall g \in \mathscr{K}(m)$, $\|Kg\| \geqslant C\,\|g\|$.

In other words, assumption (H.5) means that the null space of K should not contain a (non null) polynomial whose degree is less than m. This condition is rather weak when dealing with deconvolution problems but excludes some operator equations such as differentiation. By assumption (H.3), the norm of $L^2([0, 1], dF(t))$ is equivalent to the $L^2([0, 1], dt)$ norm with respect to the Lebesgue measure. Assumptions (H.5) and (H.3) ensure the invertibility of $\mathbf{G}_{n,\rho}$ and hence the unicity of $\hat{f}$ provided that n is sufficiently large. Finally assumption (H.4) is a technical assumption that ensures a certain amount of regularity for operator K but that can be relaxed for particular operator equations. More precisely, it implies that K is a Hilbert–Schmidt operator, i.e $\sum_j \lambda_j^2 < +\infty$, where $\{\lambda_j\}$ is the sequence of singular values of K.

Let's define $\bar{\rho}=\sup_j\rho_j$, $\underline{\rho}=\inf_j\rho_j$ and suppose $\underline{\rho}>0$. We can state now the two main theorems of this article:

THEOREM 3.1. *Suppose that n tends to infinity and k=o(n), $\bar{\rho}k^m=o(1)$, then under hypotheses* (H.1)–(H.5) *we have:*

$$\mathbb{E}\,\|f-\hat{f}\|^2_{K,n} = O\left(\frac{1}{k^{2p}}\right)+O(\bar{\rho}^2k^{2m})+O\left(\frac{k}{n}\right).$$

The best upper bound is

$$\mathbb{E}\,\|f-\hat{f}\|_{K,n}^2 = O\left(n^{\frac{-2p}{2p+1}}\right).$$

It is obtained when $k = O(n^{1/(2p+1)})$ and $\bar{\rho} = O(n^{-(p+m)/(2p+1)})$. There is no strong assumption on the decay of $\bar{\rho}$ as n goes to infinity and actually $\bar{\rho}$ can be as small as we want. However, it is well known that in practical situations a too small value of $\bar{\rho}$ leads to very bad estimates having undesirable oscillations. Thus the empirical K-norm should not be considered as an effective criterion to evaluate the asymptotic performance of this estimator.

THEOREM 3.2. *Suppose that n tends to infinity and* $k=o(n)$, $d_n = o(\underline{\rho}^3)$, $\bar{\rho}k^m = o(1)$, *then under hypotheses* (H.1)–(H.5) *one has the following upper bound:*

$$\mathbb{E}\,\|f-\hat{f}\|_K^2 = O\left(\frac{1}{k^{2p}}\right) + O(\bar{\rho}^2 k^{2m}) + O\left(\frac{d_n^2}{\underline{\rho}^6}\right) + O\left(\frac{k}{n}\right).$$

Remark 3.1. Upper bounds for the empirical and the K-norm are different and surprinsigly, that difference is entirely caused by the bias term whereas one should expect it would be the result of variance. The bounds obtained in the K-norm error depend directly on how accurately the empirical measure F_n of the design points approximates the true measure F. Furthermore, a larger amount of regularization is needed for the estimator to be convergent. For instance, if the sequence d_n decreases at the usual rate $d_n = n^{-1}$ and if one choose $\bar{\rho} \approx n^{-(p+m)/(2p+1)}$ as before then the estimator $\hat{f}$ is not consistent since $\underline{\rho} \leqslant \bar{\rho}$ and then $d_n^2/\underline{\rho}^6$ goes to infinity.

If we choose $\bar{\rho} \approx \underline{\rho} \approx n^{-(p+m)/(4p+3m)}$ and $k \approx n^{1/(4p+3m)}$, then the asymptotic error is

$$\mathbb{E}\,\|f-\hat{f}\|_K^2 = O(n^{-2p/(4p+3m)}). \tag{15}$$

Remark 3.2. This bound may not be optimal for particular operator equations since the demonstration relies on general arguments without assuming any particular decay of the singular values of K (excepted the implicit conditions imposed by H.4). Thus it must be interpreted as an upper bound for the rates of convergence: under assumptions (H.4) and (H.5) on operator K, the rate of convergence is at least the one given in (15).

Remark 3.3. The consistency of the estimator (Eq. 12) proposed by O'Sullivan (1986) is a direct consequence of Theorem 3.2. Upper bounds for the rates of convergence are those obtained in (15).

Remark 3.4. We have supposed that the interior knots were equispaced but Theorem 3.2 remains true provided that the distance between two successive knots satisfies the asymptotic condition :

$$\max_j |\delta_{j+1}-\delta_j| = O(k^{-1}) \qquad \text{and} \qquad \frac{1}{\min_{j=q,\,\ldots,\,q+k} |\delta_{j+1}-\delta_j|} = O(k).$$

Remark 3.5. In the usual nonparametric framework,

$$Y_i = f(t_i) + \varepsilon_i, \qquad i=1, \ldots, n$$

the estimator $\hat{f} = \mathbf{B}'_{qk}\hat{\boldsymbol{\theta}}$ is defined as

$$\min_{\theta \in \mathscr{R}^{q+k}} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{q+k} \theta_j B_{kj}^q(t_i) \right)^2 + \left\| \sum_{j=1}^{k+q-m} \rho_j \theta_j^{(m)} B_{kj}^{q-m} \right\|^2. \tag{16}$$

Writing

$$\mathbf{G}_{n,\,\rho} = \mathbf{C}_n + \boldsymbol{\Delta}^{(m)'} \mathbf{I}_\rho \mathbf{C}_{(q-m)\,k} \mathbf{I}_\rho \boldsymbol{\Delta}^{(m)},$$

where $[\mathbf{C}_n]_{ij} = \frac{1}{n}\sum_{\ell=1}^{n} B_{ki}^q(t_\ell)\, B_{kj}^q(t_\ell)$, $\hat{\boldsymbol{\theta}}$ is defined as in (11). The demonstration of the convergence of this estimator is an immediate consequence of the convergence of (9) since it only remains to study the asymptotic behaviour of $\mathbf{C}_n$. This has already been done by Agarwall and Studden (1980) who have shown that if $k = o(d_n^{-1})$ then $\|\mathbf{C}_n^{-1}\| = O(k)$. If $d_n = o(n^{-1/2})$, we get under (H1), (H2) and (H3)

$$\mathbb{E}\, \|f - \hat{f}\|^2 = O\left(\frac{1}{k^{2p}}\right) + O(\bar{\rho}^2 k^{2m}) + O\left(\frac{k}{n}\right) \tag{17}$$

and the usual optimal rates of convergence are attained if $k \approx n^{1/(2p+1)}$ and $\bar{\rho} = O(n^{-(p+m)/(2p+1)})$. Note that there are no conditions on $\underline{\rho}$ since $\mathbf{C}_n$ is a well conditioned matrix.

4. A SIMULATION STUDY

In this section, a small Monte Carlo experiment has been performed in order to compare the behaviour of the two estimators defined in Section 2.

We have simulated $ns = 100$ samples, each being composed of $n = 400$ noisy measurements of the convoluted function at equidistant design points in $[0, 1]$,

$$y_i = \int_0^1 k\left(\frac{i}{n} - s\right) f(s)\, ds + \varepsilon_i, \qquad i = 1, \ldots, n = 400 \tag{18}$$

with

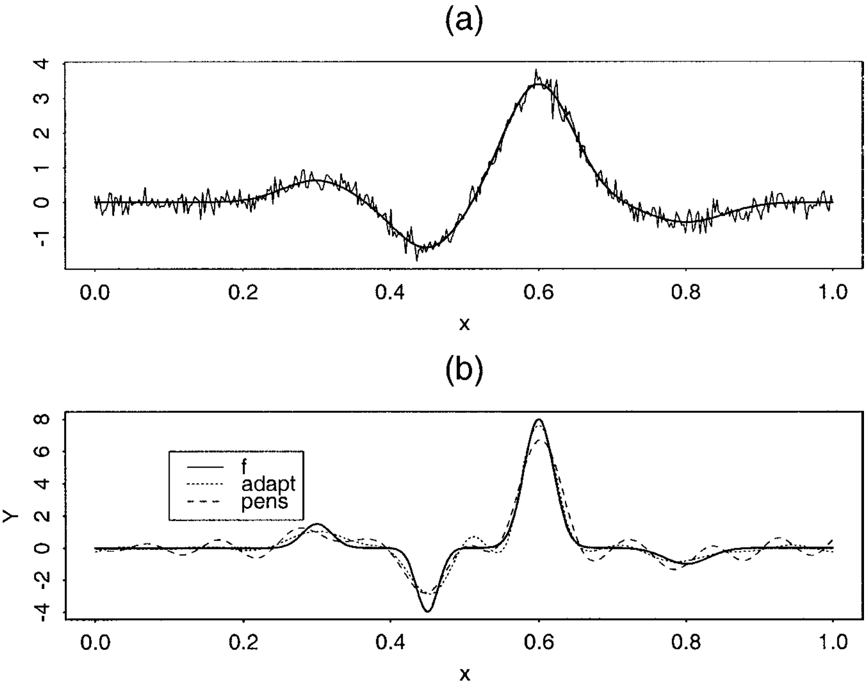

FIG. 1. (a) Convoluted function Kf and its noisy observation. (b) True function f, adaptive spline and O'Sullivan's penalized spline estimates with median fit.

$$f(t) = 1.5\exp\left(-0.5\,\frac{(x-0.3)^2}{0.02^2}\right) - 4\exp\left(-0.5\,\frac{(x-0.45)^2}{0.015^2}\right) + 8\exp\left(-0.5\,\frac{(x-0.6)^2}{0.02^2}\right) - \exp\left(-0.5\,\frac{(x-0.8)^2}{0.03^2}\right)$$

and

$$k(t-s) = \exp\left(-\frac{(t-s)^2}{\delta^2}\right), \qquad \delta = 0.08.$$

The noise ε has gaussian distribution with standard deviation 0.2 so that the signal-to-noise-ratio is 1 to 8. The integral equation (18) is practically approximated by means of a quadrature rule. Function f is drawn in Fig. 1, it is flat in some regions and oscillates in others.

We need to choose the smoothing parameter values to compute the estimates. These tuning parameters which control the regularity of the estimators are numerous: the number of knots, the order q of the splines, the order m of derivation involved in the roughness penalty and the vector $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{q+k-m})$ of smoothing parameters. Fortunately, all these parameters have not the same importance to control the behaviour of the estimators. Indeed, it appears in the usual nonparametric settings that the most crucial parameters are the elements of $\boldsymbol{\rho}$ which are regularization parameters. The number of knots and their locations are of minor importance (Eilers and Marx, 1996; Besse *et al.*, 1997), provided they are numerous enough to capture the variability of the true function f. The number of derivatives used in (10) controls the roughness penalty is rather important since two different values of m lead to two different estimators. It may have mechanical interpretation and its value can be chosen by the practitioners. Here, it was fixed to $m = 2$ and the order of the splines to $q = 4$ as it is the case in a lot of applications in the literature. We consider a set $k = 40$ equispaced knots in $[0, 1]$ to build the estimator and thus we deal with 44×44 square matrices. Nevertheless, the number of smoothing parameters remains very large: $\boldsymbol{\rho} \in \mathscr{R}_+^{42}$. To face this problem, we used the method proposed by Ruppert and Carrol (1999) which consists to select a subset of N_k, $N_k < k$, smoothing parameters $\boldsymbol{\rho}^* = (\rho^*_{N_1}, \rho^*_{N_2}, \ldots, \rho^*_{N_k})$ including the "edges" $\rho^*_{N_1} = \rho_1$ and $\rho^*_{N_k} = \rho_{q+k-m}$. The criterion (GCV, AIC, ...) used to select the smoothing parameter values is then optimized according to this subset of variables, the values of the other smoothing parameters being determined by linear interpolation: if $N_\ell < j < N_{\ell+1}$, then $\rho_j = ((\rho^*_{N_{\ell+1}} - \rho^*_{N_\ell})/(N_{\ell+1} - N_\ell)(j - N_\ell) + \rho^*_{N_\ell}$. This subset of smoothing parameters may be chosen *a priori* if one has some *a priori* knowledge of the spatial variability of the true function but in the following we will consider $N_k = 6$ equispaced "quantile" smoothing parameters.

We first consider a generalized cross validation criterion in order to choose the values of $\boldsymbol{\rho}^*$ because it is computationally fast, widely used as an automatic procedure and has been proved to be efficient in many statistical settings (Green and Silverman, 1994). Unfortunately, it seems to fail here (if there is more than one smoothing parameter) and systematically gives too small smoothing parameter values that lead to undersmoothed estimates. Actually, we think it would be better to consider a penalized version of the GCV that takes into account the number of smoothing parameters and our future work will go in that direction.

Thus, we have defined the exact empirical risk

$$R_n(\hat{f}) = \frac{1}{n} \sum_{i=1}^{n} (f(t_i) - \hat{f}(t_i))^2 \tag{19}$$

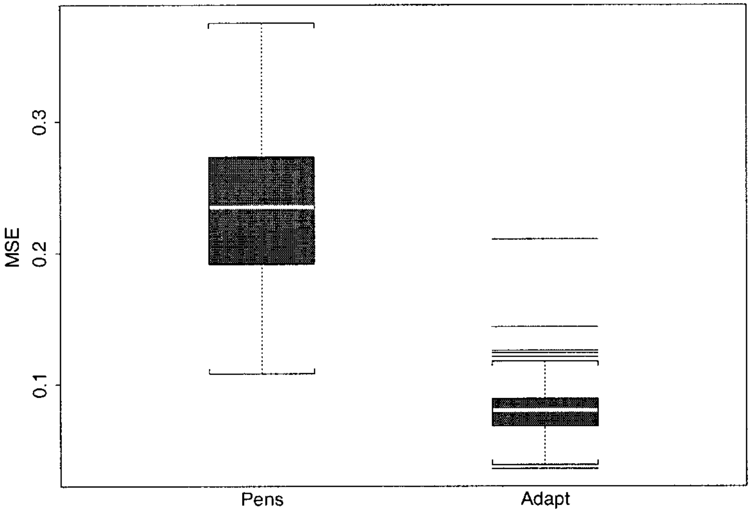

FIG. 2. Boxplot of the mean square error of estimation for each estimate. The median error is 0.23 for the penalized spline and 0.08 for the adaptive spline.

in order to evaluate the accuracy of estimate $\hat{f}$. Smoothing parameter values are chosen by minimizing the above risk so that we compare the best attainable penalized splines defined in (12) and adaptive splines estimators from samples y_i.

Boxplots of this empirical risk are drawn in Fig. 2 and show that the use of local penalties may lead to substantial improvements of the estimate: the median error is 0.08 for the adaptative spline whereas it is 0.23 for the penalized spline. The penalized spline estimates whose curvature is only controlled by one parameter can not manage both the flat regions and the oscillatories regions of function f. That's why undesirable oscillations of the penalized spline estimate appear in the intervals $[0, 0.2]$ and $[0.7, 1]$ whereas the use of local smoothing parameters allows to cope efficiently this problem (see Fig. 1).

5. PROOFS

Let us decompose the mean square error into a squared bias and a variance term according to the x norm which is successively $\{K, n\}$ and $\{K\}$:

$$\mathbb{E}\,\|f-\hat{f}\|_x^2 = \|f-\mathbb{E}\hat{f}\|_x^2 + \mathbb{E}\,\|\hat{f}-\mathbb{E}\hat{f}\|_x^2.$$

and let us study each term separately. Technical Lemmas are gathered at the end of the section.

5.1. *Bias Term*

5.1.1. *Empirical bias.* Define by $\bar{f}_k = \mathbb{E}\hat{f}$, then $\bar{f}_k = \mathbf{B}'_{qk}\bar{\boldsymbol{\theta}}_k$ where $\bar{\boldsymbol{\theta}}_k = \mathbf{G}_{n,\rho}^{-1}\frac{1}{n}\mathbf{A}'_n\mathbb{E}\mathbf{Y}$. Furthermore, it is easy to show that $\bar{\boldsymbol{\theta}}_k$ is the solution of the minimization problem

$$\min_{\boldsymbol{\theta}\in\mathcal{R}^{q+k}} \frac{1}{n}\sum_{i=1}^{n}\left(Kf(t_i)-\sum_{j=1}^{q+k}\theta_j KB^q_{kj}(t_i)\right)^2+\left\|\sum_{j=1}^{k+q-m}\rho_j\theta_j^{(m)}B_{kj}^{q-m}\right\|^2. \quad (20)$$

Criterion (20) can be written equivalently with the empirical K-norm :

$$\min_{\boldsymbol{\theta}\in\mathcal{R}^{q+k}} \|f-\mathbf{B}'_{qk}\boldsymbol{\theta}\|^2_{K,n}+\left\|\sum_{j=1}^{k+q-m}\rho_j\theta_j^{(m)}B_{kj}^{q-m}\right\|^2.$$

From Theorem XII.1 in De Boor (1978) and regularity assumption (H.1), there exists $s=\mathbf{B}'_{qk}\boldsymbol{\theta}_s\in S_{qk}$ such that

$$\sup_{t\in[0,1]} |f(t)-s(t)| = Ck^{-p}, \quad (21)$$

where constant C does not depend on k.

Furthermore we have:

$$\begin{aligned}\|f-\mathbf{B}'_{qk}\boldsymbol{\theta}_s\|^2_{K,n} &= \frac{1}{n}\sum_{i=1}^{n}\left(\int k(x,t_i)(f(x)-s(x))\,dx\right)^2\\ &\leqslant \frac{1}{n}\sum_{i=1}^{n}\left(\int |k(x,t_i)|\,|f(x)-s(x)|\,dx\right)^2\\ &\leqslant Ck^{-2p}\,\frac{1}{n}\sum_{i=1}^{n}\|k(.,t_i)\|^2\\ &= O(k^{-2p}),\end{aligned}$$

because $\lim_n (1/n)\sum_{i=1}^{n}\|k(.,t_i)\|^2 = \int_0^1\int_0^1 (k(s,t))^2\,ds\,dF(t) < +\infty$ by assumptions (H.3) and (H.4).

On the other hand, since $\bar{f}_k$ is the solution of (20), we have:

$$\|f-\bar{f}_k\|^2_{K,n}\leqslant\|f-\mathbf{B}'_{qk}\boldsymbol{\theta}_s\|^2_{K,n}+\left\|\sum_{j=1}^{k+q-m}\rho_j\theta_{s,j}^{(m)}B_{kj}^{q-m}\right\|^2.$$

From Lemma 5.2 and $\|\boldsymbol{\theta}_s\|^2 = O(k)$ one gets

$$\left\| \sum_{j=1}^{k+q-m} \rho_j \theta_{s,j}^{(m)} B_{kj}^{q-m} \right\|^2 = O(\bar{\rho}^2 k^{2m}).$$

Finally, the empirical bias is bounded as follows:

$$\|f - \mathbb{E}\hat{f}\|_{K,n}^2 = O(k^{-2p}) + O(\bar{\rho}^2 k^{2m}). \tag{22}$$

5.1.2. *K-norm bias.* Let us denote by $f_\rho = \mathbf{B}'_{qk}\boldsymbol{\theta}_\rho$ the solution of the following optimization problem:

$$\min_{\boldsymbol{\theta} \in \mathscr{R}^{q+k}} \|f - \mathbf{B}'_{qk}\boldsymbol{\theta}\|_K^2 + \left\| \sum_{j=1}^{k+q-m} \rho_j \theta_j^{(m)} B_{kj}^{q-m} \right\|^2. \tag{23}$$

Since f_ρ is the solution of (23), using the continuity of K, one gets with Lemma 5.2

$$\begin{aligned} \|f - f_\rho\|_K^2 &\leqslant \|f - s\|_K^2 + \left\| \sum_{j=1}^{k+q-m} \rho_j \theta_{s,j}^{(m)} B_{kj}^{q-m} \right\|^2 \\ &= O(k^{-2p}) + O(\bar{\rho}^2 k^{2m}), \end{aligned} \tag{24}$$

where function s is defined in (21).

Writing now

$$\|f - \bar{f}_k\|_K^2 \leqslant 2(\|f - f_\rho\|_K^2 + \|f_\rho - \bar{f}_k\|_K^2),$$

it remains to study the last term of the right side of this inequality to complete the proof. Let's denote by $\mathbf{K}_k$ the $(q+k)\times(q+k)$ matrix with elements $\langle KB_{ki}, KB_{kj}\rangle$ and define $\mathbf{G}_\rho = \mathbf{K}_k + \boldsymbol{\Delta}^{(m)\prime}\mathbf{I}_\rho \mathbf{C}_{(q-m)\,k}\mathbf{I}_\rho\boldsymbol{\Delta}^{(m)}$. We have

$$\begin{aligned} \|f_\rho - \bar{f}_k\|_K^2 &= \|\mathbf{B}'_{qk}(\boldsymbol{\theta}_\rho - \bar{\boldsymbol{\theta}}_k)\|_K^2 \\ &= (\boldsymbol{\theta}_\rho - \bar{\boldsymbol{\theta}}_k)'\, \mathbf{K}_k(\boldsymbol{\theta}_\rho - \bar{\boldsymbol{\theta}}_k) \\ &\leqslant (\boldsymbol{\theta}_\rho - \bar{\boldsymbol{\theta}}_k)'\, \mathbf{G}_\rho(\boldsymbol{\theta}_\rho - \bar{\boldsymbol{\theta}}_k) \end{aligned} \tag{25}$$

with $\boldsymbol{\theta}_\rho = \mathbf{G}_\rho^{-1}\mathbf{b}_k$, $\mathbf{b}_k = \langle K\mathbf{B}_{qk}, Kf\rangle \in \mathscr{R}^{q+k}$ and $\bar{\boldsymbol{\theta}}_k = \mathbf{G}_{n,\rho}^{-1}\bar{\mathbf{b}}_k$, $\bar{\mathbf{b}}_k = \frac{1}{n}\mathbf{A}'_n\mathbb{E}(\mathbf{Y})$. It is easy to check that $\|\bar{\mathbf{b}}_k\|^2 = O(k^{-1})$ and by classical interpolation theory that $\|\mathbf{b}_k - \bar{\mathbf{b}}_k\|^2 = O(d_n^2)$.

Pursuing the calculus begun in (25) and appealing to Lemmas 5.1 and 5.3, one gets

$$\begin{aligned}
\|f_\rho - \bar f_k\|_K^2 &\leqslant 2\,\|\mathbf{G}_\rho^{-1}\|\,(\|\mathbf{b}_k - \bar{\mathbf{b}}_k\|^2 + \|\mathbf{I}_{q+k} - \mathbf{G}_\rho \mathbf{G}_{n,\rho}^{-1}\|^2\,\|\bar{\mathbf{b}}_k\|^2) \\
&\leqslant 2\,\|\mathbf{G}_\rho^{-1}\|\,(\|\mathbf{b}_k - \bar{\mathbf{b}}_k\|^2 + \|\mathbf{G}_{n,\rho}^{-1}\|^2\,\|\mathbf{G}_\rho - \mathbf{G}_{n,\rho}\|^2\,\|\bar{\mathbf{b}}_k\|^2) \\
&= O\left(\frac{k}{\underline{\rho}^2}\right)\left(O(d_n^2) + O\left(\frac{k^2}{\underline{\rho}^4}\,\frac{d_n^2}{k^2}\,\frac{1}{k}\right)\right) \\
&= O\left(\frac{k}{\underline{\rho}^2}\right)\left(O(d_n^2) + O\left(\frac{d_n^2}{k\underline{\rho}^4}\right)\right) \\
&= O\left(\frac{d_n^2}{\underline{\rho}^6}\right);
\end{aligned}$$

that completes the proof.

5.2. *Variance*

5.2.1. *Empirical variance.* Let us denote by $\mathbf{K}_{n,k}$ the $(q+k)\times(q+k)$ matrix with elements $\frac{1}{n}\sum_{\ell=1}^n KB_{ki}(t_\ell)\,KB_{kj}(t_\ell)$. It is exactly the matrix $n^{-1}\mathbf{A}_n'\mathbf{A}_n$. Before embarking on the calculus, let's notice that $\mathbf{I}_{q+k} - \mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k} = \mathbf{\Delta}^{(m)\prime}\mathbf{I}_\rho\mathbf{C}_{(q-m)\,k}\mathbf{I}_\rho\mathbf{\Delta}^{(m)}$ is a nonnegative matrix and hence

$$\operatorname{tr}(\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}) \leqslant \operatorname{tr}(\mathbf{I}_{q+k}). \tag{26}$$

Furthermore, the largest eigenvalues of $\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}$ is less than one and thus for any $(q+k)\times(q+k)$ nonnegative matrix $\mathbf{A}$, one has $\operatorname{tr}(\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}\mathbf{A}) \leqslant \operatorname{tr}(\mathbf{A})$ (Zhou *et al.*, 1998, Lemma 6.5). Thus, under (H.2), the empirical variance term is bounded as follows :

$$\begin{aligned}
\mathbb{E}\,\|\hat f - \mathbb{E}\hat f\|_{K,n}^2 &= \mathbb{E}\left\|\mathbf{B}_k'\mathbf{G}_{n,\rho}^{-1}\,\frac{1}{n}\,\mathbf{A}_n'\boldsymbol{\varepsilon}\right\|_{K,n}^2 \\
&= \frac{1}{n^2}\,\mathbb{E}(\boldsymbol{\varepsilon}'\mathbf{A}_n\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}\mathbf{G}_{n,\rho}^{-1}\mathbf{A}_n'\boldsymbol{\varepsilon}) \\
&= \frac{\sigma^2}{n}\,\operatorname{tr}(\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}) \\
&\leqslant \frac{\sigma^2}{n}\,\operatorname{tr}(\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}) \\
&\leqslant \frac{\sigma^2(q+k)}{n}.
\end{aligned} \tag{27}$$

5.2.2. *K-norm variance.* Using the same decomposition as in (27), one obtains readily:

$$\mathbb{E}\,\|\hat{f}-\mathbb{E}\hat{f}\|_K^2 = \mathbb{E}\left\|\mathbf{B}'_{qk}\mathbf{G}_{n,\rho}^{-1}\,\frac{1}{n}\,\mathbf{A}'_n\boldsymbol{\varepsilon}\right\|_K^2$$

$$\leqslant \frac{\sigma^2}{n}\,\operatorname{tr}(\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_k). \tag{28}$$

On the other hand, one can easily check that

$$\begin{aligned}\|\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_k\| &\leqslant \|\mathbf{G}_{n,\rho}^{-1}\|\,\|\mathbf{K}_k-\mathbf{K}_{n,k}\|+\|\mathbf{G}_{n,\rho}^{-1}\mathbf{K}_{n,k}\|\\ &= O\left(\frac{k}{\underline{\rho}^2}\right)O\left(\frac{d_n}{k}\right)+O(1)\\ &= O\left(\frac{d_n}{\underline{\rho}^2}\right)+O(1)\end{aligned} \tag{29}$$

by Lemma 5.1.

Using equations (28), (29) and the condition $d_n = o(\underline{\rho}^3)$, one finally gets

$$\mathbb{E}\,\|\hat{f}-\mathbb{E}\hat{f}\|_K^2 = O\left(\frac{k}{n}\right). \tag{30}$$

5.3. *Technical Lemmas*

LEMMA 5.1. *Under assumptions* (H.3) *and* (H.4) *one has*

$$\left\|\mathbf{K}_k - \frac{1}{n}\,\mathbf{A}'_n\mathbf{A}_n\right\| = O\left(\frac{d_n}{k}\right).$$

Proof. Let g be a function of S_{qk}, then $g = \mathbf{B}'_{qk}\boldsymbol{\theta}_g$. By integration by parts followed by the Hölder inequality and invoking assumption (H.3), we have:

$$\begin{aligned}|\boldsymbol{\theta}'_g(\mathbf{K}_k - n^{-1}\mathbf{A}'_n\mathbf{A}_n)\,\boldsymbol{\theta}_g| &= |\,\|g\|_K^2 - \|g\|_{K,n}^2\,|\\ &= \left|\int_0^1 Kg(t)^2\,d(F-F_n)(t)\right|\\ &= O(d_n)\int_0^1 |Kg(t)|\,|DKg(t)|\,dt\\ &= O(d_n)\,\|Kg\|\,\|DKg\|,\end{aligned}$$

where $DKg(t) = \int_0^1 \frac{\partial k}{\partial t}(s,t)\,g(s)\,ds$.

One the other hand, from Lemma 5.3 we have $\|Kg\|^2 = \boldsymbol{\theta}_g' \mathbf{K}_k \boldsymbol{\theta}_g = O(k^{-1}) \|\boldsymbol{\theta}_g\|^2$ and, with similar arguments, since by assumption (H.4) DK is a bounded operator, we also get $\|D\,Kg\|^2 = O(k^{-1}) \|\boldsymbol{\theta}_g\|^2$.

Matching previous remarks, one finally obtains the desired result :

$$\left|\boldsymbol{\theta}_g'\left(\mathbf{K}_k - \frac{1}{n}\mathbf{A}_n'\mathbf{A}_n\right)\boldsymbol{\theta}_g\right| = O\left(\frac{d_n}{k}\right)\|\boldsymbol{\theta}_g\|^2. \quad \blacksquare$$

Let us denote by $\mathcal{N}(\boldsymbol{\Delta}^{(m)}) = \{\mathbf{u} \in \mathcal{R}^{q+k} \mid \boldsymbol{\Delta}^{(m)}\mathbf{u} = 0\}$. It is the null space of $\boldsymbol{\Delta}^{(m)}$.

LEMMA 5.2. *There are two positive constants c_1 and c_2 such that:*

- $\forall \mathbf{u} \in \mathcal{R}^{q+k}, \quad \mathbf{u}'\boldsymbol{\Delta}^{(m)\prime}\mathbf{I}_\rho \mathbf{C}_{(q-m)\,k}\mathbf{I}_\rho\boldsymbol{\Delta}^{(m)}\mathbf{u} \leqslant c_1 k^{2m-1}\bar{\rho}^2 \|\mathbf{u}\|^2$
- $\forall \mathbf{u} \in \mathcal{N}(\boldsymbol{\Delta}^{(m)})^{\perp}, \quad c_2 k^{-1}\underline{\rho}^2 \|\mathbf{u}\|^2 \leqslant \mathbf{u}'\boldsymbol{\Delta}^{(m)\prime}\mathbf{I}_\rho \mathbf{C}_{(q-m)\,k}\mathbf{I}_\rho\boldsymbol{\Delta}^{(m)}\mathbf{u}.$

Proof. The Grammian matrix $\mathbf{C}_{(q-m)\,k}$ is positive and from Agarwall and Studden (1980) it exists two positive constants c_3 and c_4 such that:

$$\forall \mathbf{u} \in \mathcal{R}^{q+k}, \quad c_3 k \|\mathbf{u}\|^2 \leqslant \mathbf{u}'\mathbf{C}_{(q-m)\,k}\mathbf{u} \leqslant c_4 k \|\mathbf{u}\|^2. \tag{31}$$

Then, it easy to check that the matrix $\mathbf{I}_\rho \mathbf{C}_{(q-m)\,k}\mathbf{I}_\rho$ is positive, its largest eigenvalue is proportional to $\bar{\rho}^2 k^{-1}$ and its smallest eigenvalue is proportional to $\underline{\rho}^2 k^{-1}$. Thus, it remains to study $\|\boldsymbol{\Delta}^{(m)}\mathbf{u}\|^2$ to complete the proof. Let's begin with the first point.

If $m=1$, then $\boldsymbol{\Delta}^{(1)}\mathbf{u} = \mathbf{D}_{qk}\mathbf{u}$ where $\mathbf{D}_{qk}$ is defined in (7). Furthermore, writing this weighted difference as

$$\begin{aligned}
\|\mathbf{D}_{qk}\mathbf{u}\|^2 &= (q-1)^2 \sum_{j=1}^{k+q-1} \left(\frac{u_{j+1}-u_j}{\delta_{j+q}-\delta_{j+1}}\right)^2 \\
&\leqslant (q-1)^2\,(k+1)^2 \sum_{j=1}^{k+q-1} (u_{j+1}-u_j)^2 \\
&\leqslant 4(q-1)^2\,(k+1)^2\,\|\mathbf{u}\|^2
\end{aligned} \tag{32}$$

and remembering that $\boldsymbol{\Delta}^{(m)}$ is obtained by iterations of the differentiation process, one gets

$$\begin{aligned}
\|\boldsymbol{\Delta}^{(m)}\mathbf{u}\|^2 &= \|\mathbf{D}_{(q-m)\,k}\cdots\mathbf{D}_{qk}\mathbf{u}\|^2 \\
&\leqslant 4^m(k+1)^2\,k^2\cdots(k+1-m)^2\,(q-1)^2\cdots(q-m)^2\,\|\mathbf{u}\|^2 \\
&= O(k^{2m})\,\|\mathbf{u}\|^2;
\end{aligned} \tag{33}$$

that completes the proof of the first point.

Let us suppose now that $\mathbf{u} \in \mathcal{N}(\mathbf{\Delta}^{(m)})^{\perp}$ and begin with $m=1$. Then,

$$\|\mathbf{D}_{qk}\mathbf{u}\|^2 \geqslant \frac{(q-1)^2\,(k+1)^2}{q^2} \sum_{j=1}^{k+q-1} (u_{j+1}-u_j)^2 \tag{34}$$

since $(\delta_{j+q}-\delta_{j+1})^{-2} \geqslant (k+1)^2\, q^{-2}$. Furthermore, the sum of $(u_{j+1}-u_j)^2$ can be expressed in a matrix way

$$\sum_{j=1}^{k+q-1} (u_{j+1}-u_j)^2 = \mathbf{u}'\mathbf{L}\mathbf{u},$$

where matrix $\mathbf{L}$ is a kind of discretized Laplacian matrix defined as follows

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & \cdots \\ 0 & -1 & 2 & -1 & \cdots & \cdots \\ \cdots & \ddots & \ddots & \ddots & \ddots & \cdots \\ \cdots & \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix},$$

Its eigenvalues are $2[1-\cos(\pi(j-1)/(k+q))]$, $j=1, \ldots, k+q$ (see Graybill, 1969). The null space of $\mathbf{L}$ is spanned by the constant vector and the smallest non null eigenvalue is proportional to $\pi^2(k+q)^{-2}$. Hence, we have

$$\forall \mathbf{u} \in \mathcal{N}(\mathbf{\Delta}^{(1)})^{\perp}, \qquad \|\mathbf{\Delta}^{(1)}\mathbf{u}\|^2 \geqslant C(k+1)^2\,\pi^2(k+q)^{-2}. \tag{35}$$

Since the smallest eigenvalue of $\mathbf{I}_\rho \mathbf{C}_{(q-m)\,k} \mathbf{I}_\rho$ is proportional to $\underline{\rho}^2 k^{-1}$, one gets the desired result for $m=1$. The proof is complete by iterating these calculus for $m=2$ and so on. ▮

LEMMA 5.3.

- $\|\mathbf{K}_k\| = O(k^{-1})$
- $\|\mathbf{G}_\rho^{-1}\| = O(k/\underline{\rho}^2)$
- *if* $d_n = o(\underline{\rho}^2)$ *then* $\|\mathbf{G}_{n,\rho}^{-1}\| = O(\frac{k}{\underline{\rho}^2})$

Proof. Let us denote by K^* the adjoint operator of K, then, for $\boldsymbol{\theta} \in \mathscr{R}^{q+k}$

$$\begin{aligned}
\boldsymbol{\theta}'\mathbf{K}_k\boldsymbol{\theta} &= \langle K\mathbf{B}'_{qk}\boldsymbol{\theta}, K\mathbf{B}'_{qk}\boldsymbol{\theta}\rangle \\
&= \langle \mathbf{B}'_{qk}\boldsymbol{\theta}, K^*K\mathbf{B}'_{qk}\boldsymbol{\theta}\rangle \\
&\leqslant \|K^*K\mathbf{B}'_{qk}\boldsymbol{\theta}\| \, \|\mathbf{B}'_{qk}\boldsymbol{\theta}\| \\
&\leqslant \|K^*K\| \, \|\mathbf{B}'_{qk}\boldsymbol{\theta}\|^2 \\
&\leqslant \|K^*K\| \, \|\mathbf{C}_{qk}\| \, \|\boldsymbol{\theta}\|^2.
\end{aligned}$$

By inequality (31), one gets $\|\mathbf{C}_{qk}\| = O(k^{-1})$. Furthermore, $\|K^*K\|$ is bounded since K is continuous and the first point is now complete.

Let us recall that $\mathbf{G}_\rho = \mathbf{K}_k + \boldsymbol{\Delta}^{(m)'}\mathbf{I}_\rho \mathbf{C}_{(q-m)\,k}\mathbf{I}_\rho\boldsymbol{\Delta}^{(m)}$ and decompose any function $g = \mathbf{B}'_{qk}\boldsymbol{\theta}_g \in \mathscr{S}_{qk}$ as

$$g = \mathbf{B}'_{qk}\boldsymbol{\theta}_{g1} + \mathbf{B}'_{qk}\boldsymbol{\theta}_{g2},$$

where $\boldsymbol{\theta}_{g1} \in \mathscr{N}(\boldsymbol{\Delta}^{(m)})$, $\boldsymbol{\theta}_{g1} + \boldsymbol{\theta}_{g2} = \boldsymbol{\theta}_g$ and $< \boldsymbol{\theta}_{g1}, \boldsymbol{\theta}_{g2} > \, = 0$.

From Lemma 5.2, one gets:

$$\boldsymbol{\theta}'_{g2}\boldsymbol{\Delta}^{(m)'}\mathbf{I}_\rho \mathbf{C}_{(q-m)\,k}\mathbf{I}_\rho\boldsymbol{\Delta}^{(m)}\boldsymbol{\theta}_{g2} \geqslant c\,\underline{\rho}^2 k^{-1}\,\|\boldsymbol{\theta}_{g2}\|^2.$$

Then, under (H.5), one has

$$\begin{aligned}
\boldsymbol{\theta}'_g\mathbf{G}_\rho\boldsymbol{\theta}_g &\geqslant \boldsymbol{\theta}'_{g1}\mathbf{K}_k\boldsymbol{\theta}_{g1} + c\,\underline{\rho}^2 k^{-1}\,\|\boldsymbol{\theta}_{g2}\|^2 \\
&\geqslant c'k^{-1}\,\|\boldsymbol{\theta}_{g1}\|^2 + c\,\underline{\rho}^2 k^{-1}\,\|\boldsymbol{\theta}_{g2}\|^2 \\
&\geqslant c'k^{-1}(\|\boldsymbol{\theta}_g\|^2 - \|\boldsymbol{\theta}_{g2}\|^2) + c\,\underline{\rho}^2 k^{-1}\,\|\boldsymbol{\theta}_{g2}\|^2 \\
&\geqslant \min(c'k^{-1}, c\,\underline{\rho}^2 k^{-1})\,\|\boldsymbol{\theta}_g\|^2,
\end{aligned} \tag{36}$$

and thus the smallest eigenvalue of $\mathbf{G}_\rho$ satisfies $\lambda_{\min}(\mathbf{G}_\rho) \approx \underline{\rho}^2 k^{-1}$. Writing now $\mathbf{G}_{n,\rho} - \mathbf{G}_\rho = \frac{1}{n}\mathbf{A}'_n\mathbf{A}_n - \mathbf{K}_k$, we get with Lemma 5.1:

$$\|\mathbf{G}_{n,\rho} - \mathbf{G}_\rho\| = O\left(\frac{d_n}{k}\right). \tag{37}$$

Consequently the smallest eigenvalue of $\mathbf{G}_{n,\rho}$ satisfies

$$|\lambda_{\min}(\mathbf{G}_{n,\rho}) - \lambda_{\min}(\mathbf{G}_\rho)| = O\left(\frac{d_n}{k}\right)$$

and then $\lambda_{\min}(\mathbf{G}_{n,\rho}) \approx \underline{\rho}^2 k^{-1}$ provided that $d_n = o(\underline{\rho}^2)$. ∎

REFERENCES

F. Abramovich and B. W. Silverman, Wavelet decomposition approaches to statistical inverse problems, *Biometrika* **85** (1998), 115–129.

G. Agarwal and W. Studden, Asymptotic integrated mean square error using least squares and bias minimizing splines, *Ann. Statist.* **8** (1980), 1307–1325.

P. Besse, H. Cardot, and F. Ferraty, Simultaneous nonparametric regressions of unbalanced longitudinal data, *Computat. Statist. Data Anal.* **24** (1997), 255–270.

H. Cardot and C. A. T. Diack, Convergence en moyenne quadratique de l'estimateur de la régression par splines hybrides. *C. R. Acad. Sci. Paris, Ser. I* **326** (1998), 615–618.

H. Cardot, Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Nonparametric Statist.* **12** (2000), 503–538.

I. J. Craig and J. C. Brown, "Inverse Problems in Astronomy," Hilger, Bristol, 1986.

C. de Boor, "A Practical Guide to Splines," Springer-Verlag, New York, 1978.

D. G. T. Denison, B. K. Mallick, and A. F. M. Smith, Automatic Bayesian curve fitting, *J. Roy. Statist. Soc. B* **60** (1998), 333–350.

P. Dierckx, "Curve and Surface Fitting with Splines," Clarendon Press, Oxford, 1993.

D. L. Donoho, Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comp. Harm. Anal.* **2** (1995), 101–126.

P. H. C. Eilers and B. D. Marx, Flexible smoothing with B-splines and penalties (with discussion), *Statistical Science* **11** (1996), 89–121.

J. H. Friedman, Multivariate adaptive regression splines (with discussion), *Annals of Stat.* **19** (1991), 1–141.

F. A. Graybill, "Introduction to Matrices with Applications in Statistics," Wadsworth, Belmont, CA, 1969.

P. Green and B. Silverman, "Nonparametric Regression and Generalized Linear Models," Chapman and Hall, London, 1994.

P. C. Hansen, "Rank-Deficient and Discrete Ill-posed Problems. Numerical Aspects of Linear Inversion," SIAM, Philadelphia, 1998.

C. Kelly and J. Rice, Monotone smoothing with application to dose-response curves and the assessment of synergism, *Biometrics* **46** (1990), 1071–1085.

J. Mendelsohn and J. Rice, Deconvolution of microfluometric histograms with B-splines, *J. Amer. Statist. Assoc.* **77** (1982), 748–753.

A. Neumaier, Solving Ill-Conditioned and Singular Linear Systems: a Tutorial on Regularization, *SIAM Review* **40** (1998), 636–666.

D. W. Nychka and D. Cox, Convergence rates for regularized solutions of integral equations from discrete noisy data, *Annals of Stat.* **17** (1989), 694–712.

F. O'Sullivan, A statistical perspective on ill-posed inverse problems, *Statist. Sci.* **4** (1986), 502–527.

D. Ruppert and R. J. Carroll, Spatially-adaptive penalties for spline smoothing, *Austral. New Zealand J. Statist.*, in press.

C. J. Stone, M. Hansen, C. Kooperberg, and Y. K. Truong, Polynomial splines and their tensor product in extended linear modeling, *Annals of Stat.* **25** (1997), 1371–1425.

A. N. Tikhonov and V. Y. Arsenin, "Solutions of Ill-Posed Problems," Wiley, New York, 1977.

A. N. Tikhonov and A. V. Goncharsky (Eds.), "Ill-Posed Problems in the Natural Sciences," MIR, Moscow, 1987.

G. Wahba, Practical approximate solutions to linear operator equations when the data are noisy, *SIAM J. Numer. Anal.* **14** (1977), 651–667.

S. Zhou, X. Shen, and D. A. Wolfe, Local asymptotics for regression splines and confidence regions. *Annals of Stat.* **26** (1998), 1760–1782.