Local roughness penalties for regression splines

Hervé Cardot[1]

[1] Unité Biométrie et Intelligence Artificielle, INRA Toulouse, BP 27, 31326, Castanet-Tolosan, cedex, France
Email: Herve.Cardot@toulouse.inra.fr

Summary

This paper introduces a new nonparametric estimator of the regression based on penalized regression splines. Local roughness penalties that rely on local support properties of B-splines are introduced in order to deal with the spatial heterogeneity of the function to be estimated. This estimator is shown to attain optimal rates of convergence. Then its good performances are confirmed on a simulation study.

Keywords: Local roughness penalties, Spatially adaptive estimators, Regression Splines, Convergence.

1 Introduction

Consider the nonparametric regression problem:

$$Y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, n. \tag{1}$$

where f is some unknown function that is supposed to be spatially heterogenous, t_i are design points -deterministic or random- and ϵ_i is a white noise independent of t_i with unknown variance σ^2. Without loss of generality we assume that $t_i \in [0, 1]$, $i = 1, \dots, n$.

In such a situation, as shown in Fig. 1, classical nonparametric estimators are not able to handle the varying regularity of f and generally give estimates

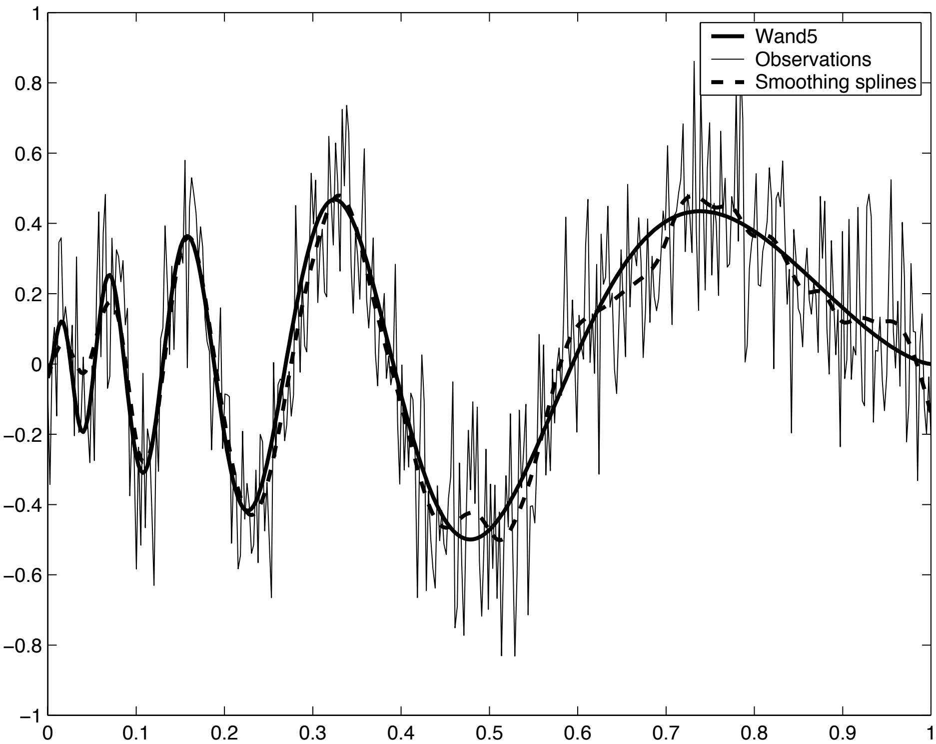

Figure 1: *True function f and its noisy observation and cubic smoothing spline estimate. The smoothing parameter value is chosen by GCV.*

that undersmooth the regular part of f and oversmooth its irregular part. Many adaptive nonparametric estimators have been proposed in the literature to cope with such a situation. They rely on local bandwith choice for kernels (Vieu 1991), adaptative choice of the knots location for regression splines (Friedman 1991, Denison *et al.* 1998) and require sophisticated algorithms (Wand 2000).

The estimator proposed below is a penalized regression splines whose original idea traces back to O'Sullivan (1986) and Kelly and Rice (1991). In these two papers a roughness penalty was introduced for the least squares regression splines estimator. More recently, Ruppert and Carroll (2000) have proposed an estimator based on truncated polynomials with penalizations 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. The main drawbacks are that the design matrix may be ill-conditioned and the penalty, based on the coefficients, may be difficult to interpret. In this article, a similar approach that relies on B-splines is proposed. It has two main advantages compared with the Ruppert and Carroll's procedure. At first, the design matrix is well conditionned because the estimation rely on B-splines which are functions with local support. On the other hand, the roughness penalty is based on the norm of a given order derivative of the estimator and thus may be interpreted more easily.

Indeed, using the local support property of B-spline functions and the fact 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. In practical situations, the local smoothing parameters values must be chosen very carefully in order to get accurate estimates. The generalized cross validation (GCV) criterion is widely used for nonparametric regression (Green and Silverman, 1994) and a small Monte Carlo experiment has shown that it allows to select "good" values of the smoothing parameters. Note that here the tuning parameters are too numerous to be selected directly with the GCV criterion. Indeed, following Ruppert and Carroll (2000), this criterion is optimized onto a subset of smoothing parameters.

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 integrated square error are briefly given for deterministic and random designs. Then, in section 4, a simulation study compares the behaviour of this estimator to the smoothing splines and the estimator proposed by Ruppert and Carroll. Finally, a sketch of the proof is given in section 5.

Matlab programs for carrying out the estimation are avalaible on request.

2 Spline estimates with local roughness penalties

Before presenting our estimator let's recall the definition of the P-spline approach proposed by Ruppert and Carroll (2000). The function of interest is expanded in a basis of truncated polynomials :

$$m(x,\boldsymbol{\beta}) = \sum_{j=1}^{q} \beta_j x^{j-1} + \sum_{j=1}^{k} \beta_{q+j} \left(x-\delta_{q+j}\right)_+^{q-1} \qquad (2)$$

where $0 < \delta_{q+1} < \cdots < \delta_{q+k} < 1$ are the interior knots. The estimator $m(.,\widehat{\boldsymbol{\beta}})$ is given by the solution of :

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{q+k}} \sum_{i=1}^{n} \left(Y_i - m(x_i,\boldsymbol{\beta})\right)^2 + \sum_{j=1}^{k} \rho_j \beta_{q+j}^2 . \qquad (3)$$

The penalization acts on the coefficients associated to the truncated polynomials. When all the value of the penalties are the same ($i.e$ $\rho_j = \rho$) then one gets the estimator proposed by Marx and Eilers (1996).

Let's denoted by S_{qk} the space spanned by the basis

$$\{1, x, \ldots, x^{q-1}, (x - \delta_{q+j})_+, \ldots, (x - \delta_{q+j})_+^{q-1}\}.$$

It can be shown (see e.g Dierckx 1993) that S_{qk} also admits a basis of normalized B-splines of order q with k interior knots $0 < \delta_{q+1} < \cdots < \delta_{q+k} < 1$. These functions which are denoted by $\{B_{kj}^q, j = 1, \ldots, q+k\}$, are non negative and have local support:

$$B_{kj}^q(x) = 0 \quad \text{if} \quad x \notin [\delta_j, \delta_{j+q}], \quad j = 1, \ldots, q+k, \tag{4}$$

where $\delta_1 = \delta_2 = \ldots = \delta_q = 0$ and $\delta_{q+k+1} = \cdots = \delta_{2q+k} = 1$.

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 $s = \sum_{j=1}^{k+q} \theta_j B_{kj}^q = \mathbf{B}_{qk}^T \boldsymbol{\theta}$ then its derivative

$$\begin{aligned} Ds &= (q-1) \sum_{j=1}^{k+q-1} \frac{\theta_{j+1} - \theta_j}{\delta_{j+q} - \delta_{j+1}} B_{kj}^{q-1}, \\ &= \mathbf{B}_{(q-1)k}^T \boldsymbol{\theta}^{(1)} \end{aligned} \tag{5}$$

where B_{kj}^{q-1} 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 difference $(k+q-1) \times (k+q)$ matrix which gives the coordinates in $S_{(q-1)k}$ of the derivative of a function of S_{qk}:

$$\boldsymbol{\theta}^{(1)} = \mathbf{D}_{qk} \boldsymbol{\theta}. \tag{6}$$

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:

$$\begin{aligned} \boldsymbol{\theta}^{(m)} &= \mathbf{D}_{(q-m+1)k} \cdots \mathbf{D}_{qk} \boldsymbol{\theta} \\ &= \boldsymbol{\Delta}_{(m)} \boldsymbol{\theta} \end{aligned} \tag{7}$$

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 estimator. 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

$$\begin{aligned} \hat{f} &= \sum_{j=1}^{q+k} \hat{\theta}_j B_{kj}^q \\ &= \mathbf{B}_{qk}^T \hat{\boldsymbol{\theta}}, \end{aligned} \tag{8}$$

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

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{q+k}} \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{q+k}\theta_j B_{kj}^q(t_i)\right)^2 + \left\|\sum_{j=1}^{k+q-m}\rho_j\theta_j^{(m)}B_{kj}^{q-m}\right\|^2. \qquad (9)$$

$\theta_j^{(m)}$ 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 $B_{kj}^q(t_i)$ 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's define

$$\mathbf{G}_{n,\rho} = \left(\frac{1}{n}\mathbf{A}_n^T\mathbf{A}_n + \boldsymbol{\Delta}_{(m)}^T\mathbf{I}_\rho\mathbf{C}_{(q-m)k}\mathbf{I}_\rho\boldsymbol{\Delta}_{(m)}\right), \qquad (10)$$

where $\mathbf{I}_\rho$ is the diagonal matrix with diagonal elements ρ_j. Then, the solution $\hat{\boldsymbol{\theta}}$ of the minimization problem (9) is given by:

$$\hat{\boldsymbol{\theta}} \ = \ \mathbf{G}_{n,\rho}^{-1}\frac{1}{n}\mathbf{A}_n^T\mathbf{Y} \qquad (11)$$

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

Remark 2.1 *If $\rho_j = \rho,\ j = 1,\ldots,k+q-m$, then the estimator defined by (8) is the same as the estimator proposed by Kelly and Rice (1991). These kind of penalized regression splines have already been used in different frameworks. Besse et al. (1997) have performed the principal components analysis of unbalanced longitudinal data, Diack and Thomas (1998) have built a test of convexity of a regression function. Slightly different estimators have also been proposed by Eilers and Marx (1996).*

Remark 2.2 *The local penalty defined in (9) may be viewed as a kind of discrete version of the following continuous penalty:*

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

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

3 Asymptotic results

Very few results of convergence have been published until now for penalized splines. We give here asymptotic results that show, under classical assumptions, that these estimators attain the optimal rates of convergence for well chosen smoothing parameter values.

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 and knots and the moments of the noise:

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

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

(H.3) deterministic design: Let's denote by F_n the empirical distribution of the design sequence, $\{t_{j,n},\ 1 \leq j \leq n\} \subset [0,1]$ 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.3b) random design: Suppose the design points are sampled from a distribution F that has a continuous, bounded, and strictly positive density h on $[0,1]$. Let's denote by F_n the empirical distribution of the design sequence, $\{t_{j,n},\ 1 \leq j \leq n\} \subset [0,1]$, it follows from the Glivenko-Cantelli theorem that

$$\sup_{t\in[0,1]} |F(t) - F_n(t)| = O_p(n^{-1/2}).$$

(H.4) The distance between two successive interior knots satisfies the asymptotic condition :

$$\max_{j=0,\ldots,k} |\delta_{q+j} - \delta_{q+j+1}| = O(k^{-1}) \quad \text{and} \quad \frac{1}{\min_{j=0,\ldots,k} |\delta_{q+j} - \delta_{q+j+1}|} = O(k).$$

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. It ensures the invertibility of $\mathbf{G}_{n,\rho}$ and hence the unicity of $\hat{f}$ provided that n is sufficiently large compared to k. Finally assumption (H.4) means that the interior knots have to fill the intervall $[0,1]$ homogeneously. It is fulfilled for instance when they are equispaced.

Let's define $\bar{\rho} = \sup_j \rho_j$. We can state now the main theorem of this article:

Theorem 3.1 *Suppose that n tends to infinity and $k{=}o(\min(n, d_n))$, $\bar{\rho}k^m = o(1)$, then under hypotheses (H.1), (H.2), (H.3) and (H.4) we have:*

$$\mathbb{E}\left\|f - \hat{f}\right\|^2 = O\left(\frac{1}{k^{2p}}\right) + O(\bar{\rho}^4 k^{4m}) + O\left(\frac{k}{n}\right).$$

If the design is random, if $k = o(n^{-1/2})$ and $\bar{\rho}k^m = o(1)$, we have under (H.1), (H.2), (H.3b) and (H.4):

$$\mathbb{E}\left(\left\|f - \hat{f}\right\|^2 |t_1, \ldots, t_n\right) = O_p\left(\frac{1}{k^{2p}}\right) + O_p(\bar{\rho}^4 k^{4m}) + O_p\left(\frac{k}{n}\right).$$

The optimal rates of convergence are attained when $k = O(n^{1/(2p+1)})$ and $\bar{\rho} = O(n^{-(p+2m)/(4p+2)})$.

4 A simulation study

In this section, a small Monte Carlo experiment has been performed in order to compare the behaviour of this estimators with the well known smoothing splines and the estimator proposed by Ruppert and Carroll (2000)[1].

We have simulated $ns = 100$ samples, each being composed of $n = 100$ (for moderate sample sizes) and $n = 400$ (for large sample sizes) noisy measurements of function f at equidistant design points in $[0, 1]$:

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{12}$$

with $t_i = \frac{i-1}{n-1}$,

$$f_1(t) = \sqrt{t(1-t)} \sin\left(\frac{2\pi(1 + 2^{(9-4j)/5})}{t + 2^{(9-4j)/5}}\right), \quad j = 5,$$

and $f_2(t) = \sin(2\pi t)$.

Function f_1 and its noisy observation are drawn in Fig 1. It has already been used by Wand (2000) to evaluate the ability of different types of regression splines estimators to adapt to spatial variability. Function f_2 allow us to evaluate the "robustness" of the local penalized estimators when the function varies regurlarly. The noise ϵ has gaussian distribution with standard deviation 0.2

4.1 Choice of the smoothing parameters

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 do not have the same importance to control the behaviour of

[1] Their Matlab programs are available at the address : `http://www.orie.cornell.edu/~davidr/matlab/`. The function `srslocal` was used to perform the estimation.

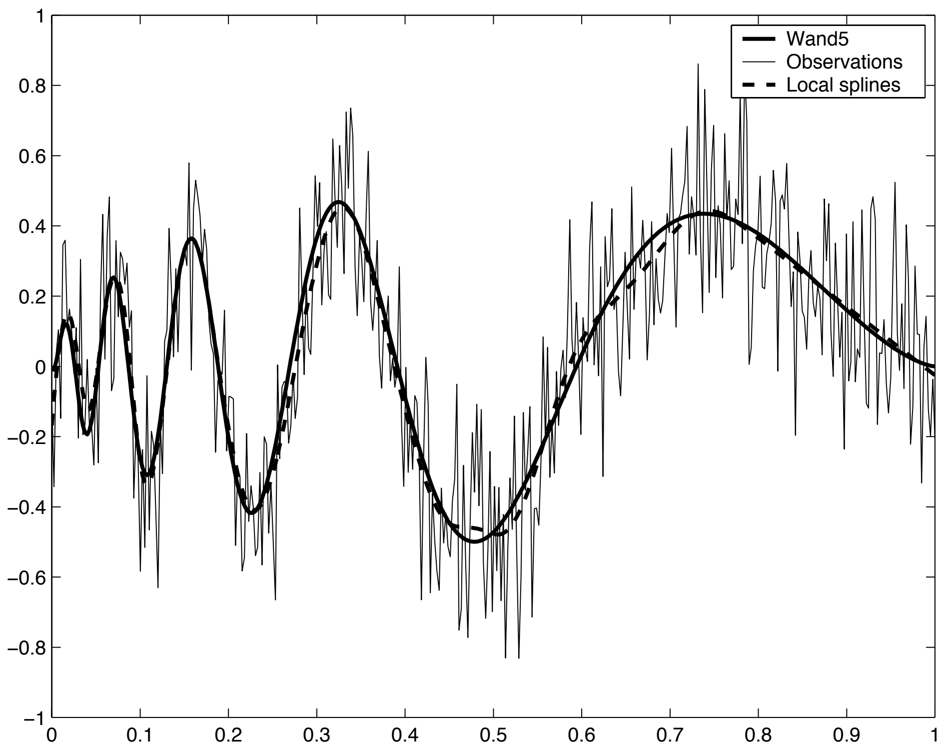

Figure 2: *True function f and its noisy observation and local penalty cubic regression spline estimator with median fit. The number of interior knots is 40 and there are 3 "leading" smoothing parameters whose values are chosen by GCV. This figure must be compared to Fig. 1.*

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.

We considered for both estimators cubic regression splines ($q = 4$) and a set of $k = 40$ (resp. $k = 20$) equispaced knots in $[0, 1]$ when $n = 400$ (resp. when $n = 100$) to build the estimators. Thus we deal with 44×44 and 24×24 square matrices. The penalty of the locally penalized B-splines is on the second order derivative ($m = 2$).

Nevertheless, the number of smoothing parameters remains very large: for instance, when $n = 400$, $\boldsymbol{\rho} \in \mathbb{R}_+^{42}$ for the penalized B-splines. To face this problem, we used the strategy proposed by Ruppert and Carrol (2000). It consists in selecting a subset of N_k, $N_k < k$, "leading" smoothing parameters $\boldsymbol{\rho}^* = (\rho^*_{p(1)}, \rho^*_{p(2)}, \ldots, \rho^*_{p(N_k)})$ including the "edges", $p(1) = 1$ and $p(N_k) = q+k-m$ and in interpolating between these leading parameters in order to get the values of the other smoothing parameters. Considering the penalization in equation (9), a natural way to achieve that is to locate the smoothing parameters at the maximum value of the associated B-spline function. More precisely, let's define $x_j = \arg\max_x B^{q-m}_{kj}(x)$, for $j = 1, \ldots, k+q-m$. Then, if $p(\ell) < j < p(\ell+1)$, the value of the jth smoothing parameter is obtained as follows:

$$\rho_j = \frac{\rho^*_{p(\ell+1)} - \rho^*_{p(\ell)}}{x_{p(\ell+1)} - x_{p(\ell)}} \left(x_j - x_{p(\ell)}\right) + \rho^*_{p(\ell)}$$

Then, the GCV criterion

$$GCV(\boldsymbol{\rho}^*) \quad = \quad \frac{\sum_{i=1}^{n} \left(Y_i - \widehat{f}_{\boldsymbol{\rho}}(t_i)\right)^2}{\left(n - \operatorname{tr}\left(\mathbf{G}^{-1}_{n,\boldsymbol{\rho}} \frac{1}{n} \mathbf{A}_n^T \mathbf{A}_n\right)\right)^2}, \qquad (13)$$

is minimized according to this subset $\boldsymbol{\rho}^*$ of parameters, the values of the other smoothing parameters being entirely determined by linear interpolation. Let's notice that the optimization of GCV, with multiple smoothing parameters does not require much more computations than for the smoothing splines. Indeed, the computation time is multiplied by a factor N_k (see Ruppert and Carroll, 2000) since the optimization is performed effectively with a line seach type algorithm. It consists in minimizing GCV over each ρ^*_j, the other leading ones being kept fixed, and iterate the process for each "leading" smoothing parameter. In practice, each ρ^*_j varies in a grid and the evaluation of the GCV does not require much computation since the only matrix that is affected is $\mathbf{I}_{\boldsymbol{\rho}}$ in $\mathbf{G}_{n,\boldsymbol{\rho}}$ (see equation 10). As noticed by Ruppert and Carroll, this procedure is really effective here since the penalties are defined locally and variations of the values of one ρ^*_j only affect the estimator in a neighbourhood of the knot associated to ρ^*_j.

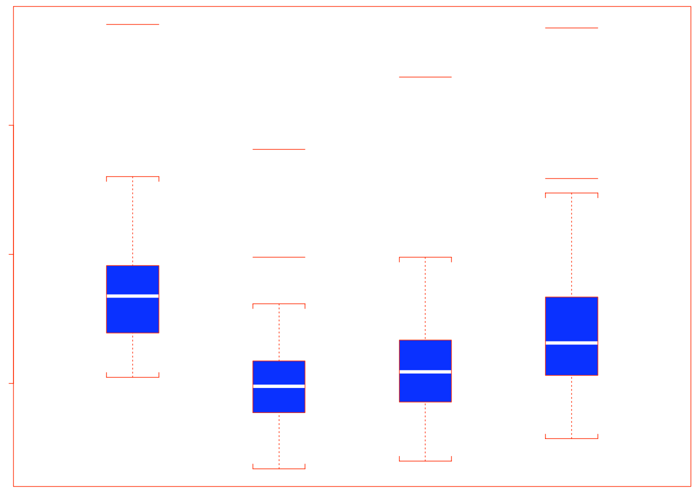

Figure 3: *Boxplot of the mean square error of estimation for each estimate. The first box represents error for smoothing splines and the others error for local penalties regression B-splines for different number of "leading" smoothing parameters.*

4.2 Comparison and discussion

Then, we have defined the exact empirical risk

$$R_n(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\Big(f(t_i) - \hat{f}(t_i)\Big)^2 \tag{14}$$

in order to evaluate empirically the accuracy of the estimates of f.

We have compared for function f_1 the local B-splines estimates with $N_k = 3$, $N_k = 6$ and $N_k = 10$ equispaced "quantile" smoothing parameters. Boxplots of the empirical risks R_n are drawn in Fig. 3 and show that the use local penalties may lead to substantial improvements of the estimator. The smoothing spline whose curvature is only controlled by one parameter can not manage both the flat regions and the oscillatory regions of function f. One the other hand if the "leading" smoothing parameters are too numerous, then the variance of the error seems to increase.

The subset of "leading" smoothing parameters may be chosen *a priori* if one has some knowledge of the spatial variability of the true function. If it is not the case, this subset is also chosen by GCV. It consists in computing the GCV for different numbers of leading smoothing parameters and choosing the estimator associated to the lowest value of the GCV.

Then a comparison was made with Ruppert and Carroll (2000) estimators. Median errors are given in table 1. One can notice that these two locally adaptative approaches do not have the same behaviors even if both penalties are quadratic. On the one hand, the Ruppert and Carroll estimator seems to perform better for smooth functions such as f_2. On the other hand, our estimator gives better results for the locally varying function f_1. Finally, let's remark that both estimator perform well for regular function f_2 when the sample size is large ($n = 400$).

	smoothing splines	local splines	R & C P-splines
$n = 100$, $f = f_1$	1.05	0.81	1.13
$n = 400$, $f = f_1$	0.38	0.22	0.26
$n = 100$, $f = f_2$	0.25	0.38	0.30
$n = 400$, $f = f_2$	0.09	0.11	0.09

Table 1: *Median errors of criterion R_n for the different estimators. The number of subknots of the adaptative estimators are chosen automatically by minimizing the GCV criterion.*

In this article we have proposed a nonparametric estimator based on regression splines that can adapt to locally varying signals through local roughness penalties. It has been shown to be convergent under classical conditions and we have noticed on a simulation study that it behaves well both for irregular and regular signals provided the sample size is large enough. Nevertheless, further work is needed to choose the number of leading smoothing parameters N_k and their locations. One possible strategy could be to adapt the bayesian approach of Denison and *al.* (1998) based on reversible jump MCMC in order to choose not the knots, as done in their work, but the number and positions of "leading" smoothing parameters. The main advantage of such an approach based on the smoothing parameters is that we reduce considerably dimension of the space and thus decrease the computation time.

5 Proofs

Suppose first that the design is deterministic. In order to prove the convergence of $\hat{f}$ it suffices to bound $\mathbb{E}||\hat{f} - \hat{f}_k||^2$ where $\hat{f}_k$ is the classical regression splines estimator. It has been shown that $\hat{f}_k$ attains the optimal rates of convergence (see Agarwall and Studden 1980). Actually, we have

$$\begin{aligned} \mathbb{E}||f - \hat{f}||^2 &\leq 2\left(\mathbb{E}||f - \hat{f}_k||^2 + \mathbb{E}||\hat{f} - \hat{f}_k||^2\right) \\ &= O\left(k^{-2p}\right) + O\left(k/n\right) + O\left(\mathbb{E}||\hat{f} - \hat{f}_k||^2\right). \end{aligned}$$

Let us write now

$$\begin{aligned}\mathbb{E}||\hat f - \hat f_k||^2 &= \mathbb{E}(\hat{\boldsymbol\theta}_k - \hat{\boldsymbol\theta})^T \mathbf{C}_{qk}(\hat{\boldsymbol\theta} - \hat{\boldsymbol\theta}) \\ &\le ||\mathbf{C}_{qk}||\, \mathbb{E}\left\|\hat{\boldsymbol\theta}_k - \hat{\boldsymbol\theta}\right\|^2 \\ &\le ||\mathbf{C}_{qk}||\, \mathbb{E}\left\|\left(\mathbf{C}_n^{-1} - \mathbf{G}_{n,\rho}^{-1}\right)\frac{1}{n}\mathbf{A}_n^T\mathbf{Y}\right\|^2 \\ &\le ||\mathbf{C}_{qk}||\, \mathbb{E}\left\|\mathbf{C}_n^{-1} - \mathbf{G}_{n,\rho}^{-1}\right\|^2 ||\frac{1}{n}\mathbf{A}_n^T\mathbf{Y}||^2\end{aligned}$$

where $[\mathbf{C}_n]_{ij} = \frac{1}{n}\sum_{\ell=1}^n B_{ki}^q(t_\ell)B_{kj}^q(t_\ell)$. It has been shown (Agarwall and Studden 1980) that $||\mathbf{C}_{qk}|| = O(k^{-1})$ and $||\mathbf{C}_n^{-1}|| = O(k)$. Thus, using the decomposition $\mathbf{G}_{n,\rho} = \mathbf{C}_n + \boldsymbol\Delta_{(m)}^T \mathbf{I}_\rho \mathbf{C}_{(q-m)k}\mathbf{I}_\rho\boldsymbol\Delta_{(m)}$ we have

$$\mathbf{C}_n^{-1} - \boldsymbol{G}_{n,\rho}^{-1} = \mathbf{C}_n^{-1}\left(\mathbf{I} - \left(\mathbf{I} + \boldsymbol\Delta_{(m)}^T\mathbf{I}_\rho\mathbf{C}_{(q-m)k}\mathbf{I}_\rho\boldsymbol\Delta_{(m)}\mathbf{C}_n^{-1}\right)^{-1}\right). \quad (15)$$

Writing now

$$\boldsymbol\Delta_{(m)}^T\mathbf{I}_\rho\mathbf{C}_{(q-m)k}\mathbf{I}_\rho\boldsymbol\Delta_{(m)} = \bar\rho^2\left(\boldsymbol\Delta_{(m)}^T\bar\rho^{-1}\mathbf{I}_\rho\mathbf{C}_{(q-m)k}\bar\rho^{-1}\mathbf{I}_\rho\boldsymbol\Delta_{(m)}\right)$$

it is easily seen (Cardot 2000) that

$$\begin{aligned}||\boldsymbol\Delta_{(m)}^T\bar\rho^{-1}\mathbf{I}_\rho\mathbf{C}_{(q-m)k}\bar\rho^{-1}\mathbf{I}_\rho\boldsymbol\Delta_{(m)}\mathbf{C}_n^{-1}|| &\le ||\boldsymbol\Delta_{(m)}^T\mathbf{C}_{(q-m)k}\boldsymbol\Delta_{(m)}\mathbf{C}_n^{-1}|| \\ &= O(k^{2m})\end{aligned}$$

since $||\bar\rho^{-1}\mathbf{I}_\rho|| = 1$ and $||\boldsymbol\Delta_{(m)}^T\boldsymbol\Delta_{(m)}|| = O(k^{2m})$. Assuming now that $\bar\rho^2 k^{2m} = o(1)$, we can expand equality (15) in order to get the approximation:

$$\mathbf{C}_n^{-1} - \mathbf{G}_{n,\rho}^{-1} = \bar\rho^2\mathbf{C}_n^{-1}\left(\boldsymbol\Delta_{(m)}^T\bar\rho^{-1}\mathbf{I}_\rho\mathbf{C}_{(q-m)k}\bar\rho^{-1}\mathbf{I}_\rho\boldsymbol\Delta_{(m)}\mathbf{C}_n^{-1} + o(k^{2m})\right). \quad (16)$$

Consequently, we have $||\mathbf{C}_n^{-1} - \mathbf{G}_{n,\rho}^{-1}|| = O(\bar\rho^2 k^{2m+1})$ and thus $\mathbb{E}||\hat f - \hat f_k||^2 = O(\bar\rho^4 k^{4m+1})\mathbb{E}||\frac{1}{n}\mathbf{A}_n^T\mathbf{Y}||^2$. Appealing again to Agarwall and Studden, it can be shown that $\mathbb{E}||\frac{1}{n}\mathbf{A}_n^T\mathbf{Y}||^2 = O(k^{-1})$, and

$$E||\hat f - \hat f_k||^2 = O(\bar\rho^4 k^{4m})$$

that completes the proof.

When the design is random, Zhou *et al.* (1998) have shown, under (H.3b):

$$\mathbb{E}\left(\left\|f - \hat f_k\right\|^2 |t_1,\ldots,t_n\right) = O_p\left(\frac{1}{k^{2p}}\right) + O_p\left(\frac{k}{n}\right).$$

Furthermore, we have from Burman and Chen (1989) that $||\mathbf{C}_n|| = O_p(k^{-1})$ and $||\mathbf{C}_n^{-1}|| = O_p(k)$. Consequently, it is easy to show $||\mathbf{C}_n^{-1} - \mathbf{G}_{n,\rho}^{-1}|| = O_p(\bar\rho^2 k^{2m+1})$ and $||\frac{1}{n}\mathbf{A}_n^T\mathbf{Y}||^2 = O_p(k^{-1})$ that completes the proof.

Acknowledgements. I'm indebted to M. Goulard, P. Sarda and A. Trubuil for helpful discussions which improved this manuscript in various ways.

References

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

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

Burman, P. and Chen, K.W. (1989). Nonparametric estimation of a regression function. *Ann. Statist.*, **17**, 1567-1596.

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

de Boor, C. (1978). *A Practical Guide to Splines.* Springer, New-York.

Dierckx, P. (1993). *Curve and Surface Fitting with Splines.* Clarendon Press, Oxford.

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

Diack, C.A.T. and Thomas, C. (1998). A nonparametric test of the non convexity of the regression. *Nonparametric Statistics*, **9**, 335-362.

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

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

Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall.

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

O'Sullivan, F. (1986). A Statistical Perspective on Ill-Posed Inverse Problems. *Statistical Science*, **4**, 502–527.

Ruppert, D. and Carrol, R.J. (2000). Spatially-adaptive penalties for spline smoothing. *Australian and New Zealand Journal of Statistics*, **42**, 205–223.

Vieu, P. (1991). Nonparametric Regression: Optimal Local Bandwith Choice *J. R. Statist. Soc. B*, **53**, 453-464.

Wand, M. P. (2000). A comparison of regression spline smoothing procedures. *Comput. Statist.*, **15**, 443-462.

Zhou, S., Shen, X. and Wolfe, D.A. (1998). Local Asymptotics for Regression Splines and Confidence Regions. *Annals of Stat.*, **26**, 1760-1782.