Available online at www.sciencedirect.com
ScienceDirect
Journal of the Korean Statistical Society 37 (2008) 259–267
JOURNAL OF THE KOREAN STATISTICAL SOCIETY
www.elsevier.com/locate/jkss

On knot placement for penalized spline regression

Fang Yao[a], Thomas C.M. Lee[b,c,*]

[a] *Department of Statistics, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada*
[b] *Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*
[c] *Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA*

Received 1 August 2007; accepted 1 January 2008
Available online 16 June 2008

Abstract

This paper studies the problem of knot placement in penalized regression spline fitting. Given a pre-specified number of knots, most existing knot placement methods allocate the knots in an equally spaced fashion. This paper proposes a simple knot placement scheme for improving such "equally spaced methods". This new scheme first identifies locations of local extrema in the target function, and then it places additional knots in such places. The rationale behind this is that quite often such local extrema coincide with the critical locations for placing knots. The proposed scheme is shown to be superior in a simulation study.

AMS 2000 subject classifications: 62G08; 65D10

Keywords: Knot placement; Local extrema; Nonparametric regression; Penalized splines; Semiparametric regression

1. Introduction

Due to its simplicity and effectiveness for handling different semiparametric smoothing problems, penalized spline regression has recently become a popular tool for solving various estimation problems, ranging from environmental modeling (Wood & Augustin, 2002) to longitudinal and functional data analysis (Harezlak, Ryan, Giedd, & Lange, 2005; Yao & Lee, 2006) to remote sensing imaging (Clarke et al., 2006). In addition, it has been applied to conduct sample surveys (Breidt, Claeskens, & Opsomer, 2005), to perform classification (Banerjee, Maiti, & Mukhopadhyay, 2006), and to construct confidence intervals for generalized additive models (Wood, 2006). For background details on penalized splines, see for example Eilers and Marx (1996) and Ruppert, Wand, and Carroll (2003), while some theoretical results on penalized splines can be found in Hall and Opsomer (2005). When comparing to smoothing splines, an attractive property of penalized spline regression is the ease for conducting statistical inference (e.g., see Ruppert et al. (2003)).

There are two important components in fitting penalized splines. The first is the selection of the smoothing parameter, while the second is the choice of the number of knots and their locations. This paper focuses on the

* Corresponding author at: Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.
E-mail addresses: fyao@utstat.toronto.edu (F. Yao), tlee@sta.cuhk.edu.hk (T.C.M. Lee).

1226-3192/$ - see front matter
doi:10.1016/j.jkss.2008.01.003

second component. For the first component, common methods include cross-validation, generalized cross-validation and Mallows' C_p.

For the second component, a common two-step knot selection strategy is to first select the total number of knots and then allocate the knots in an "equally spaced" fashion, in the sense that either the distance or the number of distinct design points between any two adjacent knots remains the same throughout the whole data domain. These two steps can further be iterated in the following manner to obtain better regression estimates. After a first regression estimate is obtained with a given number of equally spaced knots, a diagnostic procedure is applied to check whether a lack of fit exists for this estimate. If a lack of fit does exist, a second regression estimate will be obtained with a larger number of equally spaced knots; otherwise the first initial regression estimate will be taken as the final estimate. If a lack of fit still exists for the second regression estimate, then a third estimate will be obtained by further increasing the number of knots. This procedure continues until a satisfactory estimate is obtained.

The goal of this paper is to propose a simple knot placement scheme that improves the above "equally spaced strategy". This proposed scheme first detects those locations at which sharp changes are likely to be present in the underlying regression function. Then it places additional knots at these detected locations. This new knot placement scheme performed very well in the simulation study reported below.

The rest of this paper is organized as follows. Section 2 provides some background material while Section 3 presents the proposed procedure. Then in Section 4 the proposed procedure is numerically evaluated via a simulation study. Lastly concluding remarks are offered in Section 5.

2. Background

Suppose that n pairs of measurements are observed, $\{x_i, y_i\}_{i=1}^n$, satisfying the model $y_i = m(x_i) + \epsilon_i$, where $m(x)$ is an unknown regression function and the errors ϵ_i are independent with constant variance σ^2. It is assumed that $m(x)$ can be well modeled by a pth-degree penalized spline with truncated polynomial basis:

$$m(x) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{k=1}^{K} \beta_{pk}(x - \kappa_k)_+^p, \tag{1}$$

where $\{\kappa_1, \ldots, \kappa_K\}$ is a set of fixed knots and the function $(x)_+$ is defined as $(x)_+ = \max(x, 0)$. With this setup the estimation of $m(x)$ can be achieved via the estimation of the coefficients $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p, \beta_{p1}, \ldots, \beta_{pK})^{\mathrm{T}}$ in the following manner. Define $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, the diagonal matrix $\boldsymbol{D}$ as $\boldsymbol{D} = \mathrm{diag}\{0_{(p+1)\times(p+1)}, 1_{K\times K}\}$ and

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^p & (x_1-\kappa_1)_+^p & \cdots & (x_1-\kappa_K)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^p & (x_n-\kappa_1)_+^p & \cdots & (x_n-\kappa_K)_+^p \end{bmatrix}.$$

Then, for a given smoothing parameter λ, a penalized least-squares estimate $\hat{\boldsymbol{\beta}}_\lambda$ for $\boldsymbol{\beta}$ can be obtained as the minimizer of the following penalized least-squares criterion:

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda^{2p}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\beta}.$$

Direct algebra shows that $\hat{\boldsymbol{\beta}}_\lambda$ admits the following closed-form solution:

$$\hat{\boldsymbol{\beta}}_\lambda = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \lambda^{2p}\boldsymbol{D})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}. \tag{2}$$

The corresponding fitted values $\hat{\boldsymbol{m}}_\lambda = (\hat{m}_\lambda(x_1), \ldots, \hat{m}_\lambda(x_n))^{\mathrm{T}}$ are given by $\hat{\boldsymbol{m}}_\lambda = \boldsymbol{X}\hat{\boldsymbol{\beta}}_\lambda$, while estimate $\hat{m}_\lambda(x)$ of $m(x)$ for any value of x can be obtained by substituting $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_\lambda$ in (1). In practice the quality of $\hat{\boldsymbol{m}}_\lambda$ and $\hat{m}_\lambda(x)$ depends on the choices of λ and $\{\kappa_1, \ldots, \kappa_K\}$. The next section presents a new method for choosing $\{\kappa_1, \ldots, \kappa_K\}$.

3. The proposed knot placement scheme

3.1. Motivation

As mentioned before, most existing knot selection methods are "equally spaced", in the sense that either the distance or the number of distinct x_i between any two adjacent knots remains the same throughout the range of

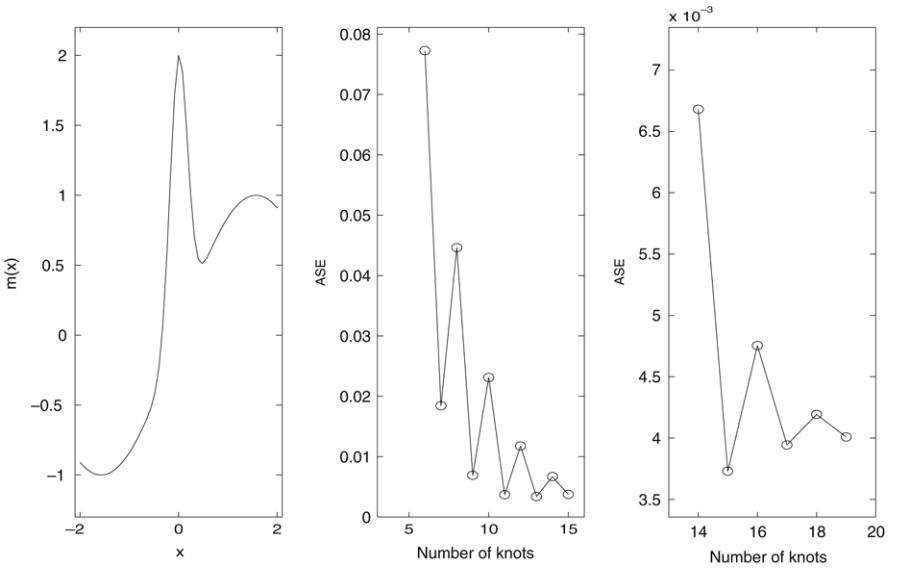

Fig. 1. Left: the test function $m(x)$ defined by (5); middle and right: the averages of ASE (6), taken over 400 samples, for different numbers of equally spaced knots.

the x_i's; e.g., see Ruppert (2002) and Ruppert et al. (2003). In particular Ruppert et al. (2003, Section 5.5.3) propose the following method for choosing the knots $\{\kappa_1, \ldots, \kappa_K\}$:

$$\kappa_k = \left(\frac{k+1}{K+2}\right) \text{th sample quantile of the unique } x_i, \tag{3}$$

where

$$K = \min\left(\frac{1}{4} \times \text{number of unique } x_i, 35\right). \tag{4}$$

This knot placement method is simple and straightforward to implement. However, as illustrated in the next subsection, the "equally spaced" nature of this method does not guarantee that knots are placed at *all* critical locations; that is, locations at which the underlying regression function possesses sharp changes. Typically such critical locations correspond to peaks (local maxima) or troughs (local minima) of the regression function. These considerations present the needs for a new knot placement scheme that combines rule (3) and (4) and the detection of critical locations.

3.2. An illustrative example

This subsection presents an example that illustrates the need for alternative knot placement schemes other than those "equally spaced methods". In this example the true but unknown testing regression function is

$$m(x) = \sin(2x) + 2\mathrm{e}^{-16x^2}, \quad -2 \le x \le 2. \tag{5}$$

This test function has been used previously by various researchers (e.g., see Fan and Gijbels (1996)) and is displayed in the left panel of Fig. 1. From this test function, 400 noisy data sets of size $n = 50$ were generated in the following way: the design points $x_i = \frac{4(i-1)}{49} - 2$, $i = 1, \ldots, 50$, are regularly spaced in $[-2, 2]$, while the errors ϵ_i are iid $N(0, 0.1^2)$. For each of the 400 individual data sets we obtained different estimates for $m(x)$ using different numbers of equally spaced knots in $[-2, 2]$. The numbers of knots used were $K = 5, \ldots, 19$, and hence for each data set a total of 15 estimates for $m(x)$ were obtained. The smoothing parameter λ was chosen by generalized cross-validation and the degree of the spline is $p = 3$. For each regression estimate $\hat{m}_\lambda$ the following averaged squared error (ASE)

was computed as a numerical measure for the estimation quality:

$$\text{ASE} = \frac{1}{n}\sum_{i=1}^{n}\{m(x_i) - \hat{m}_\lambda(x_i)\}^2. \tag{6}$$

In the middle and right panels of Fig. 1 the averages of these computed ASEs, taken over the 400 samples, are plotted against the number of knots K.

Examining the plots of ASEs, one interesting observation is that the shape of the ASE paths resembles a saw blade. This can be explained as follows. From the left panel of Fig. 1, one can see that one of the most obvious structures of the testing function is the peak located at $x = 0$, and it is important to place a knot at $x = 0$ to allow additional flexibility for the regression estimate to capture this sharp structure. Due to the equally spaced nature of the knot placement method, the knots sequence excludes 0 for even K, while the sequence includes 0 if K is odd. Thus for even K the regression estimate is not flexible enough at $x = 0$ to capture this sharp structure and results in a higher ASE value. On the other hand, as for odd K a knot is placed at $x = 0$, the regression estimate is capable of modeling this sharp structure which in turn produces a lower ASE value. Such a switching of high and low ASE values explains the "saw blade" behavior of the ASE plots. Even though the ASE values somewhat stabilize when $K \geq 12$, this "saw blade" behavior is still apparent; see the right panel of Fig. 1. Therefore, due to the "unlucky" combination of the regression function structures and the value of K, the "equally spaced methods" can be improved by placing additional knots at which the regression functions are at local extrema.

3.3. The proposed procedure

Driven by the above discussion, the following knot placement procedure is proposed for improving those "equally spaced methods".

(1) Compute an initial estimate $\hat{\boldsymbol{\beta}}_0$ for $\boldsymbol{\beta}$ using (2), where the set of knots $\mathcal{K} = \{\kappa_1, \ldots, \kappa_n\}$ is chosen by (3) and (4). The smoothing parameter λ can be chosen using any standard selection procedure, such as generalized cross-validation.
(2) Estimate $m(x)$ by substituting $\hat{\boldsymbol{\beta}}_0$ into (1). Denote the resulting estimate as $\hat{m}_0(x)$.
(3) Locate those x-values at which $\hat{m}_0(x)$ are either local maxima or local minima. Denote the set of such x-values as $\mathcal{X}$. In our implementation these local maxima or minima are located by conducting the following grid search. From left to right, we scan and compute the values of $\hat{m}_0(x)$ for a set of fine grid points equally spaced in $[x_1, x_n]$. If $\hat{m}_0(x')$ is greater (smaller) than both $\hat{m}_0(x' - \delta)$ and $\hat{m}_0(x' + \delta)$ for some $\delta > 0$, then $\hat{m}_0(x')$ is declared as a local maximum (minimum).
(4) Lastly obtain the new set of knots as the union of $\mathcal{K}$ and $\mathcal{X}$. New estimates for $\boldsymbol{\beta}$ and hence $m(x)$ can be computed with this new set of knots.

We close this section with the following two remarks. First, other smoothing methods, such as local polynomial smoothing (Fan & Gijbels, 1996) and smoothing splines (Green & Silverman, 1994), can also be used to obtain the initial estimate $\hat{m}_0(x)$ for $m(x)$. Second, one could place additional knots at the local extrema of the $\hat{m}_\lambda(x)$ that are obtained at the end of Step 4. In fact, one could iterative this idea. That is, continue adding knots until no new local extremum is found in the resulting $\hat{m}_\lambda(x)$. However, our numerical experience suggests that this iterative scheme does not significantly improve the final regression estimates. Therefore, for speed considerations and simplicity, we do not recommend using this iterative scheme.

4. Simulation study

A simulation study was conducted to evaluate the practical performance of the proposed knot placement scheme. In particular it is compared to the "equally spaced method" previously described in (3) and (4). The experimental setup was originally due to Professor Steve Marron and has been used for example by Wand (2000) and Lee (2003). This setup was designed to study the effects of independently changing one of the following four experimental factors: (i) noise level, (ii) design density, (iii) degree of spatial variation and (iv) noise variance function. Each factor is changed 6 times, so altogether there are 24 different experimental configurations. These configurations are listed in Table 1, and typical noisy data sets simulated from these configurations are displayed in Figs. 2–5. Throughout the whole

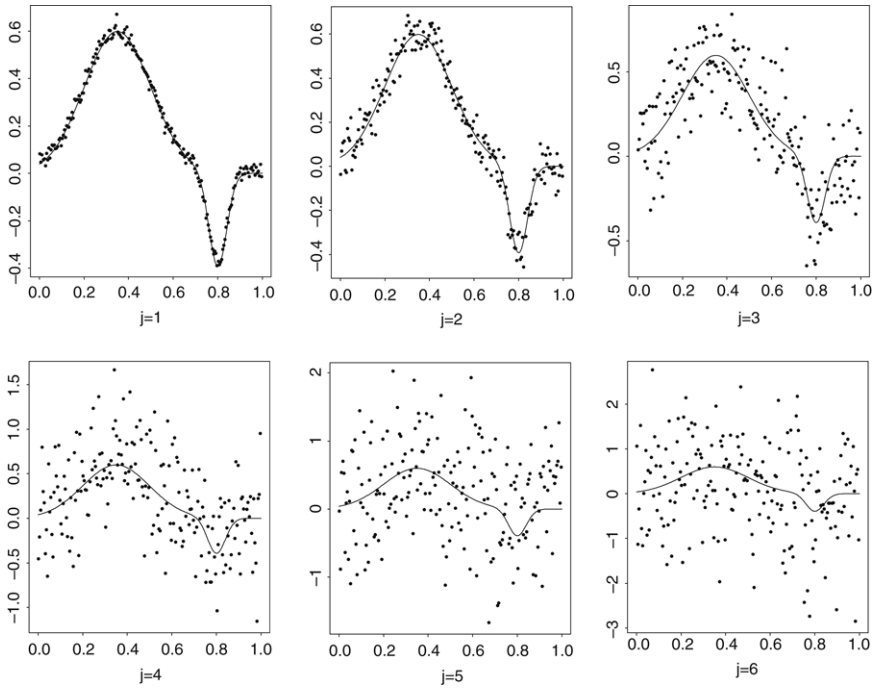

Fig. 2. Simulated noisy data (dots) and the corresponding true regression functions (solid lines) for the experimental factor *noise level*.

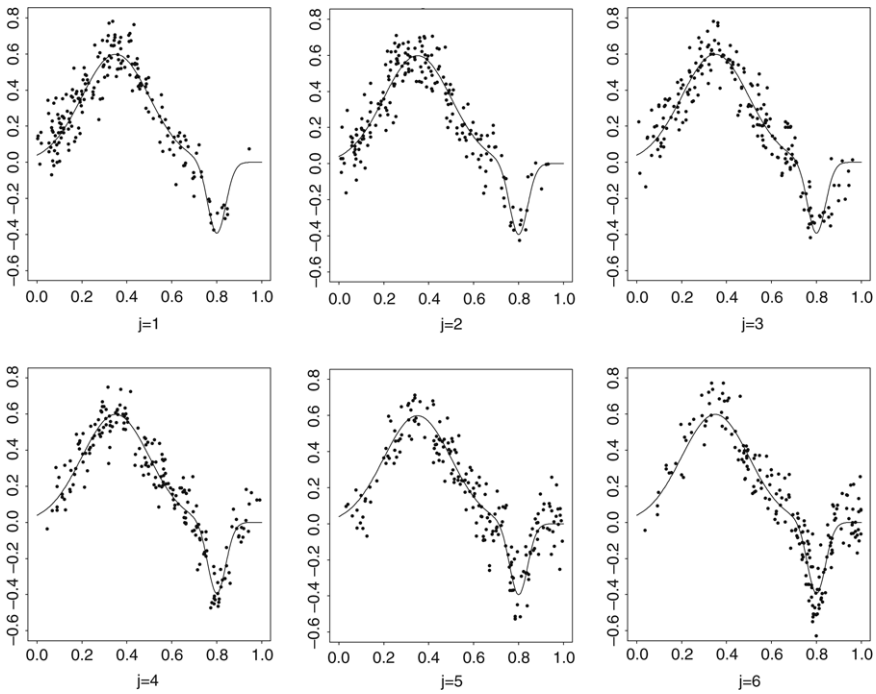

Fig. 3. Similar to Fig. 2 but for the experimental factor *design density*.

simulation study generalized cross-validation was used to select λ, and the order p of the spline used was $p = 3$. Given the simple structures of the testing functions, the number of knots K chosen by (4) should be large enough.

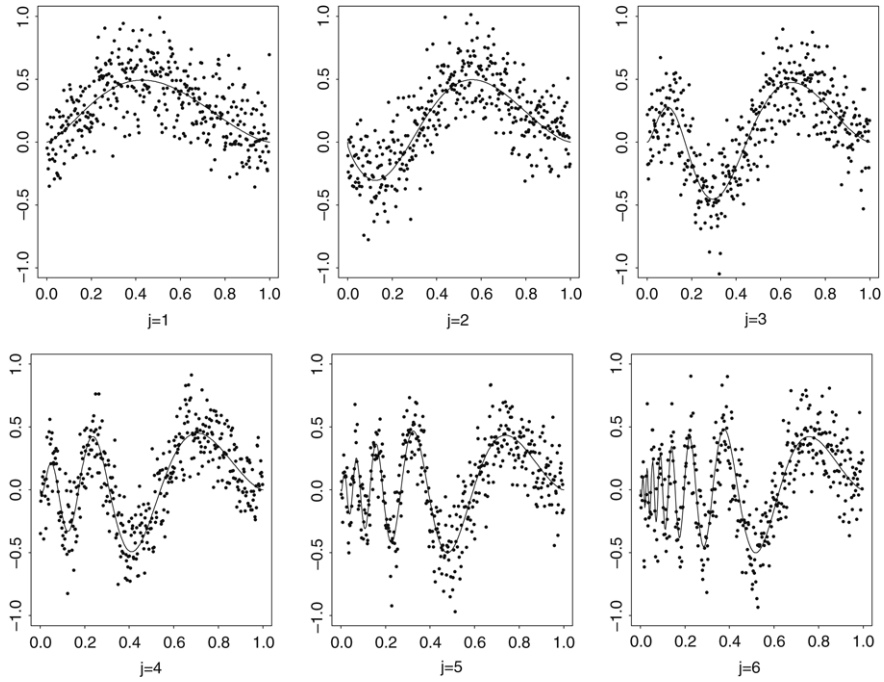

Fig. 4. Similar to Fig. 2 but for the experimental factor *spatial variation*.

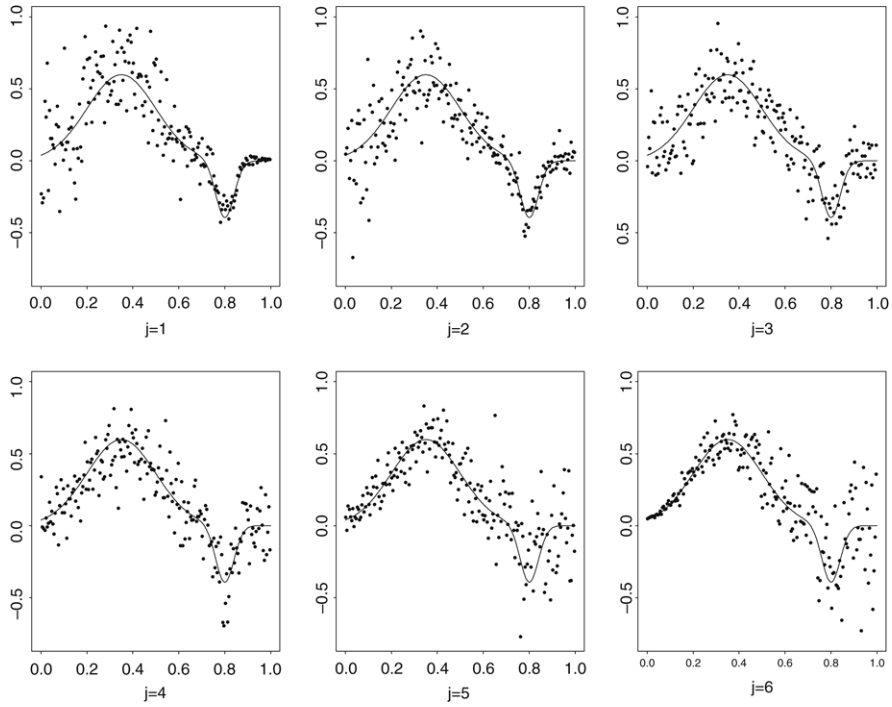

Fig. 5. Similar to Fig. 2 but for the experimental factor *variance function*.

For each of the 24 configurations, 200 data sets were simulated. Then the method of (3) and (4) and the proposed method described in Section 3.3 were applied to select the knots, and the corresponding estimates for m were also obtained. We use the same numerical measure ASE as defined in (6) to evaluate the quality of any regression estimate $\hat{m}_\lambda$.

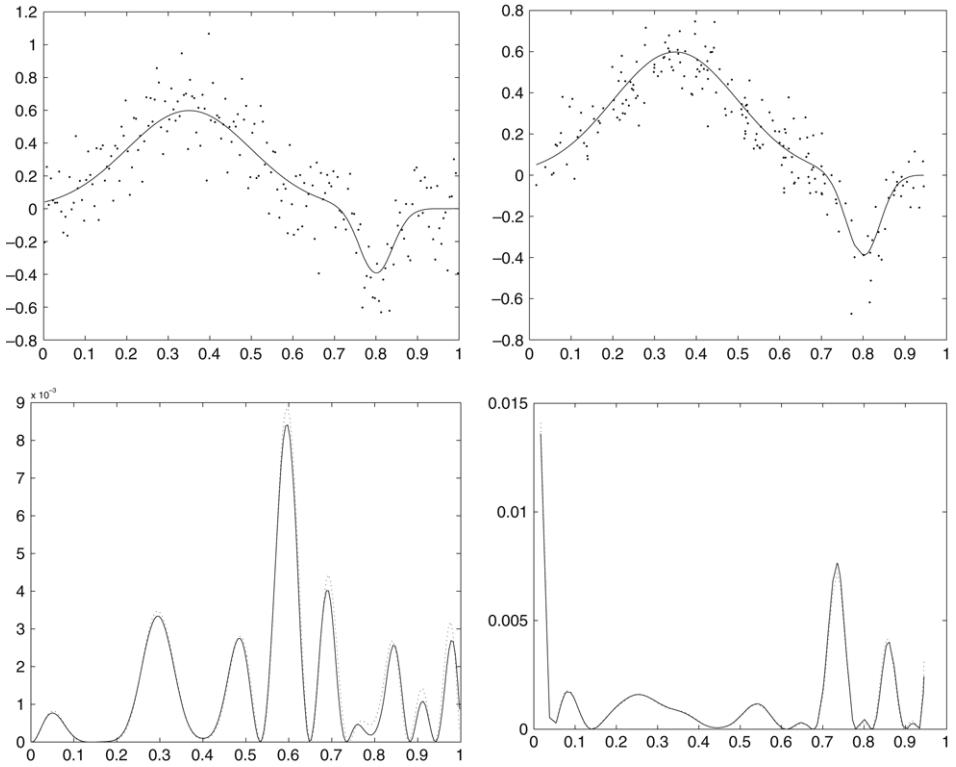

Fig. 6. Plots for visual evaluation. Left column: noise level experimental factor; right column: design density experimental factor. Top row: true regression function and noisy observations. Bottom row: local squared errors of the knot selection method (3) and (4) (dotted lines) and the proposed method (solid lines).

Table 1
Specification of the simulation setup

Factor	Generic form	Particular choices
Noise level	$y_{ij} = f(x_i) + \sigma_j \epsilon_i$	$\sigma_j = 0.02 + 0.04(j-1)^2$
Design density	$y_{ij} = f(X_{ji}) + \sigma \epsilon_i$	$\sigma = 0.1$, $X_{ji} = F_j^{-1}(X_i)$
Spatial variation	$y_{ij} = f_j(x_i) + \sigma \epsilon_i$	$\sigma = 0.2$, $f_j(x) = \sqrt{x(1-x)} \sin\left[\frac{2\pi\{1+2^{(9-4j)/5}\}}{x+2^{(9-4j)/5}}\right]$
Variance function	$y_{ij} = f(x_i) + \sqrt{v_j(x_i)}\epsilon_i$	$v_j(x) = [0.15\{1 + 0.4(2j-7)(x-0.5)\}]^2$
	$j = 1, \ldots, 6$; $n = 200$; $x_i = \frac{i-0.5}{n}$; $\epsilon_i \sim$ iid $N(0,1)$	
	$f(x) = 1.5\phi\left(\frac{x-0.35}{0.15}\right) - \phi\left(\frac{x-0.8}{0.04}\right)$; $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2})$	
	$X_i \sim$ iid Uniform[0, 1]; F_j is the Beta $\left(\frac{j+4}{5}, \frac{11-j}{5}\right)$ c.d.f.	

The means of the paired differences between the ASEs of the method of (3) and (4) and the proposed method are listed in Table 2. From this table one can see that, out of the total 24 experimental configurations, the proposed method gave better performances for 19 configurations, while for the remaining 5 configurations the performances of the two methods are statistically indistinguishable.

To visually evaluate the regression estimates, for each of the above four experimental factors with $j = 3$, a typical noisy data set was generated and regression estimates from the method of (3) and (4) and the proposed method were obtained. These noisy data sets were displayed in the top rows of Figs. 6 and 7. To enhance visibility, plotted in the bottom rows of these figures are the local squared errors $\{x_i, \mathrm{LSE}(x_i)\}$ between the regression estimates and the true function:

$$\mathrm{LSE}(x_i) = \{m(x_i) - \hat{m}_\lambda(x_i)\}^2, \quad i = 1, \ldots, n.$$

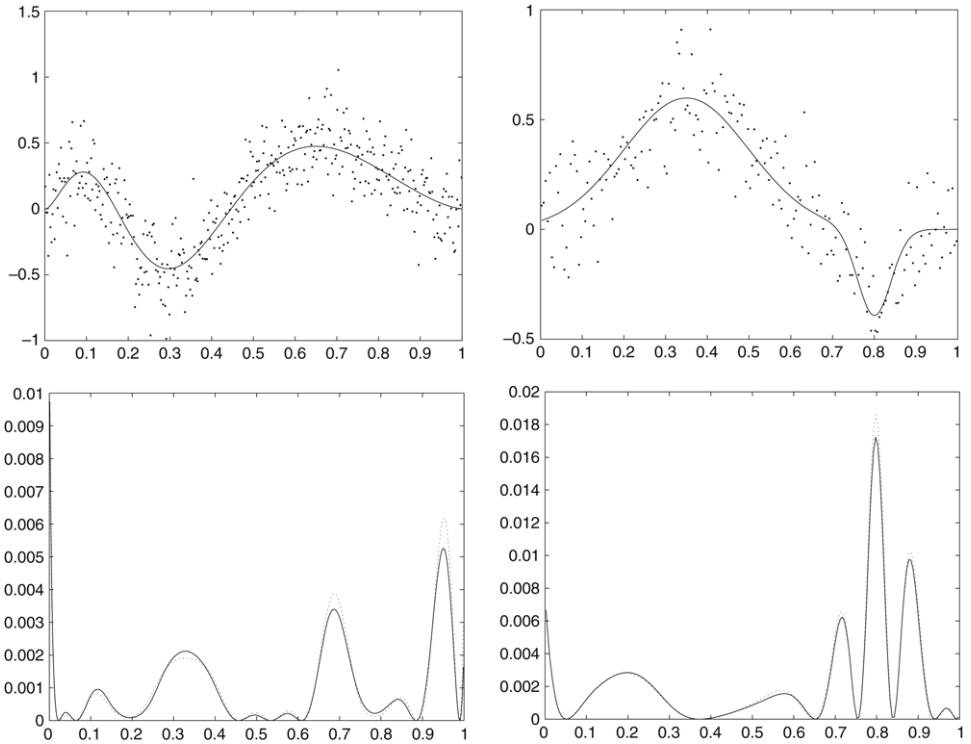

Fig. 7. Plots for visual evaluation. Left column: spatial variation experimental factor; right column: variance function experimental factor. Top row: true regression function and noisy observations. Bottom row: local squared errors of the knot selection method (3) and (4) (dotted lines) and the proposed method (solid lines).

Table 2
Means of the paired differences between the ASEs ($\times 10^5$) of the knot selection method (3)–(4) and the proposed method

j	Noise level	Design density	Spatial variation	Variance function
1	0.123* (0.0189)	2.25* (0.457)	0.213 (0.253)	2.22* (0.915)
2	0.576* (0.100)	2.40* (0.308)	−0.553 (0.379)	1.71* (0.707)
3	3.40* (0.870)	1.04* (0.237)	2.29* (0.640)	2.07* (0.434)
4	3.67 (3.06)	1.32* (0.232)	4.56* (0.546)	2.18* (0.497)
5	17.9* (8.15)	0.856* (0.238)	19.2* (1.20)	1.75 (0.997)
6	23.8 (26.1)	0.569* (0.208)	67.8* (5.16)	1.64* (0.797)

Numbers in parentheses are estimated standard errors. Differences that are significant at the 5% level are indicated by an asterisk.

5. Conclusions

In this paper a simple and yet effective method is proposed for identifying important knot locations for penalized regression spline fitting. The idea is to first obtain an initial estimate of the unknown function and then place additional knots at locations at which this initial function estimate are local extrema. Results from numerical experiments suggest that the proposed knot placement method improves upon methods that allocate knots in an "equally spaced" fashion.

A major reason behind the success of the proposed method is that quite often the critical locations for knot placement are at the local extrema of the target function, and the proposed method aims to identify such extrema. It is also possible that some critical knot locations are at places where the curvature of the target function is high; i.e., for large second derivative. We have also investigated a procedure for placing knots at locations with high curvature. However, numerical results suggest that this curvature based procedure is inferior to the proposed procedure described above. A likely explanation for this is that the locations of high curvature are harder to estimate than the locations of local extrema.

We close this paper by pointing out the following two possible limitations of the proposed methods: (i) if the initial estimate $\hat{m}_0(x)$ obtained from Steps 1 and 2 is poor in quality, then the proposed method may not be able to correctly identify all local extrema, and (ii) the proposed method may not be easily generalized to higher dimensional settings.

References

Banerjee, T., Maiti, T., & Mukhopadhyay, P. (2006). Classification of pathological stage of prostate cancer patients using penalized splines. *Computational Statistics and Data Analysis*, *51*, 1147–1155.

Breidt, F. J., Claeskens, G., & Opsomer, J. D. (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika*, *92*, 831.

Clarke, E. D., Speirs, D. C., Heath, M. R., Wood, S. N., Gurney, W. S. C., & Holmes, S. J. (2006). Calibrating remotely sensed chlorophyll-a data by using penalized regression splines. *Journal of the Royal Statistical Society Series C*, *55*, 331–353.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible parsimonious smoothing and additive modeling (with discussion). *Statistical Science*, *89*, 89–121.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*. London: Chapman and Hall.

Hall, P., & Opsomer, J. D. (2005). Theory for penalised spline regression. *Biometrika*, *92*, 105.

Harezlak, J., Ryan, L. M., Giedd, J. N., & Lange, N. (2005). Individual and population penalized regression splines for accelerated longitudinal designs. *Biometrics*, *61*, 1037–1048.

Lee, T. C. M. (2003). Smoothing parameter selection for smoothing splines: A simulation study. *Computational Statistics and Data Analysis*, *42*, 139–148.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, *11*, 735–757.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge University Press.

Wand, M. P. (2000). A comparison of regression spline smoothing procedures. *Computational Statistics*, *15*, 443–462.

Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian New Zealand Journal of Statistics*, *48*, 445–464.

Wood, S. N., & Augustin, N. H. (2002). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, *157*, 157–177.

Yao, F., & Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society Series B*, *68*, 3–25.