Spatially Adaptive Smoothing Splines

BY ALEXANDRE PINTORE

Department of Statistics, University of Oxford,

1 South Parks Rd, Oxford OX1 3TG, U.K.

pintore@stats.ox.ac.uk

PAUL SPECKMAN

Department of Statistics, University of Missouri,

Columbia, Missouri 65211-6100, U.S.A.

speckman@stats.missouri.edu

AND CHRIS C. HOLMES

Department of Statistics, University of Oxford,

1 South Parks Rd, Oxford OX1 3TG, U.K.

cholmes@stats.ox.ac.uk

SUMMARY

Smoothing splines are models for curve fitting that are widely used in scatter-plot smoothing and additive function estimation. Their formulation as the solution to a variational problem balances fidelity to the data against smoothness through a single parameter

λ. In this paper, we use a reproducing kernel Hilbert space representation to derive the smoothing spline solution when the smoothness penalty is a function $\lambda(t)$ of the design space t. We show how this allows the model to adapt to various degrees of smoothness in the structure of the data. We propose a convenient form for the smoothness penalty function and discuss computational algorithms for automatic curve fitting using a generalised cross-validation measure.

Some key words: Curve fitting, Generalised Cross-Validation, Smoothing spline, Spatial adaption, Splines.

1. Introduction

Curve fitting, also known as scatter-plot smoothing, is a well studied problem in statistics and almost all current data analysis software provide curve fitting routines. The methods are also applicable to high dimensional nonlinear function estimation by assuming additivity of the solution (Hastie and Tibshirani 1990, chapter 4). The smoothing spline is one of the most popular curve fitting methods partly due to empirical evidence supporting its effectiveness and partly due to it's elegant mathematical formulation.

The smoothing spline is derived as the solution to a function minimisation problem where the cost function trades off fidelity to the data in terms of sum squared error, against smoothness of the reconstructed curve in terms of the integrated squared derivative of order m, where m is typically taken to be two. The smoothing spline solution uses a global smoothing parameter λ which implies that the true underlying mean process has a constant degree of smoothness. In this paper we suggest a more general framework that accommodates varying degrees of roughness by seeking solutions where the smoothness penalty $\lambda(t)$ is now a function of the design t. The approach is deemed "spatially adaptive" as for judicious choice of $\lambda(t)$ the model is able to capture highly varying or "wiggly" regions of the underlying mean process without over-fitting the smoother areas, and visa versa. We derive the solution within a reproducing kernel Hilbert space, see Gu (2002, chapter

2) and Wahba (1990, chapter 1).

When applying the method to data we propose a piecewise constant model for the smoothing function $\lambda(t)$. This provides a convenient computational framework with closed form solutions to the corresponding reproducing kernels of the Hilbert space. It is interesting that these kernels show similarities to regression splines in that the use of piecewise constant $\lambda(t)$ induces breaks in the $m^{th}, \ldots, (2m-1)^{th}$ derivatives at the change points in $\lambda(t)$.

There is an extensive literature on spatial adaptation in curve fitting models other than smoothing splines. Local polynomial models with adaptive window widths are discussed in Fan and Gijbels (1996) and regression splines are over-viewed in Hansen and Kooperberg (2002), see also Ruppert and Carroll (2000). Bayesian methods include adaptive regression splines, Smith and Kohn (1996), and mixtures of smoothing splines, Wood et al. (2002). Denison et al. (2002, chapter 3) overview a number of Bayesian basis function approaches. In work related to ours, Cummins et al. (2001) used local Generalised Cross-Validation to fit spatially adaptive smoothing splines, while Wahba (1995) and Abramovich and Steinberg (1996) proposed L-splines for adaptive modelling. Finally, Luo and Wahba (1997) used smoothing spline results and proposed an interesting and promising approach to build the so called hybrid adaptive splines which use the reproducing kernels as their basis. However, none of the approaches so far have developed the adaptive smoothness as a function minimisation problem with an explicit penalty functional.

In Section 2 we describe the theory supporting our approach and derive the general form of the reproducing kernel of the Hilbert space associated with the spatially adaptive smoothing spline. Then as a natural first generalisation of the smoothing spline, we take the smoothness penalty $\lambda(t)$ to be piecewise constant and describe some characteristics of the spatially adaptive smoothing spline under this assumption. Section 3 discusses practical issues surrounding the estimation $\lambda(t)$ and presents some examples. Finally, Section 4 offers a discussion and points to areas of future research.

2. The modelling framework

To begin, we assume the data model

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)' \sim N(\mathbf{0}, \sigma^2 I)$, and f is an unknown function defined on an interval $T \subset \mathbb{R}$. The function f is only known to be "smooth". From here on we will also assume without loss of generality that $T = [0, 1]$. Following Wahba (1990, chapter 1), we let $\mathcal{B}_m$ be the class of functions satisfying the boundary conditions $f^{(\nu)}(0) = 0, \ \nu = 0, 1, \ldots, m-1$ and let $\mathcal{W}_m$ be the collection of functions on T such that

$$\left\{ f : \ f^{\nu}, \ \nu = 0, \ldots, m-1, \ \text{absolutely continuous and } f^{(m)} \in L^2(T) \right\},$$

where $L^2(T)$ is the space of functions square integrable over T. Finally, define $\mathcal{W}_m^0 = \mathcal{W}_m \cap \mathcal{B}_m$.

In this paper we propose an estimate of f in (1) as the solution to the minimisation problem

$$\min_{f \in \mathcal{W}_m} \frac{1}{n} \sum_{i=1}^{n} \{y_i - f(t_i)\}^2 + \int_0^1 \lambda(t) \{f^{(m)}(t)\}^2 dt, \tag{2}$$

for some bounded integrable function $\lambda(t) > 0$ on $t \in T$. We also assume $\lambda(t)^{-1}$ is integrable on T. This is a generalisation of the smoothing spline in which $\lambda(t) = \lambda$, a constant. There is an extensive literature for solving problems of the latter type. For a full account, we refer the reader to Wahba (1990, chapter 1) or Gu (2002, chapter 2).

The optimisation problem (2) is an example of L-spline smoothing with weighted penalty function as described, for example, in Gu (2002, chapter 4). The solution to (2) for given $\lambda(t)$ can be derived using of the reproducing kernel Hilbert space representation of the space $\mathcal{W}_m^0$ with inner product

$$\langle f, g\rangle_\lambda = \int \{f^{(m)}(u)\}\{g^{(m)}(u)\}\lambda(u)\, du$$

associated with the penalty term in (2). This was noted independently in the comment of Wahba (1995) and additional results were reported in Abramovich and Steinberg (1996). The present paper contains a more complete treatment of these problems, generalising the smoothing spline, and suggesting a methodology which leads to explicit computational procedures based on analytical expressions for the resulting basis functions.

We now briefly review the requisite reproducing kernel theory. Letting $(t)_+ = t$ for $t \geq 0$ and $(t)_+ = 0$ otherwise, define

$$G_m(t, u) = \frac{(t-u)_+^{(m-1)}}{(m-1)!},$$

and define the kernel

$$K_\lambda(s, t) = \int_0^1 \lambda(u)^{-1} G_m(s, u) G_m(t, u)\, du = \int_0^1 \lambda(u)^{-1} \frac{(s-u)_+^{m-1}}{(m-1)!} \frac{(t-u)_+^{m-1}}{(m-1)!}\, du \qquad (3)$$

for $s, t \in [0, 1]$. The function $G_m(t, u)$ is the Green's function for the boundary value differential equation $f^{(m)}(t) = g(t)$, $f \in \mathcal{B}_m$, $t \in [0, 1]$, i.e., the unique solution to the boundary value problem is given by

$$f(t) = \int_0^1 G_m(t, u) g(u)\, du, t \in [0, 1].$$

In particular,

$$\frac{\partial^m}{\partial u^m} K_\lambda(t,u) = \lambda(u)^{-1} G_m(t,u), \tag{4}$$

and consequently

$$\int_0^1 \{\frac{\partial^m}{\partial u^m} K_\lambda(t,u)\}\{f^{(m)}(u)\}\lambda(u)\, du = \int_0^1 G(t,u) g(u)\, du = f(t), s \in [0,1].$$

This implies that K_λ is the reproducing kernel for the space $\mathcal{W}_m^0$ with inner product $\langle f, g\rangle_\lambda$. As we show in the next paragraph, for given $\lambda(\cdot)$ and data $\boldsymbol{y} = (y_1, ..., y_n)'$, the resulting estimate $f(t)$ is completely determined. Put differently, the model for $\lambda(\cdot)$ determines the basis function space in which the solution to (2) sits. Thus, the major challenge when considering solutions to (2) lies in the modelling of $\lambda(\cdot)$.

We describe our model in Subsection 2.1. First however we give the form of the general solution to (2) obtained from reproducing kernel Hilbert space theory, assuming for the moment that $K_\lambda(\cdot,\cdot)$ is available, either in closed form or via a numerical technique.

From the general theory of smoothing splines as given in Wahba (1990, chapter 1) or Gu (2002, chapter 2), the solution to (2) lies in the finite dimensional space of functions

$$f(t) = \sum_{j=1}^{n} c_j K_\lambda(t_j, t) + \sum_{j=0}^{m-1} d_j \phi_j(t), \tag{5}$$

where $\phi_j(t) = t^j/j!$, $j = 0, \ldots, m-1$. Now define the $n \times n$ matrix

$$\Sigma_\lambda = \{K_\lambda(t_i, t_j)\}_{i,j=1,\ldots,n} \tag{6}$$

and the $n \times m$ matrix

$$T = \{\phi_j(t_i)\}_{i=1,\ldots,n,\ j=0,\ldots,m-1} . \tag{7}$$

If f satisfies (5), the penalty can be expressed as $\int \{f^{(m)}(u)\}^2 \lambda(u)\, du = c'\Sigma_\lambda c$, and the problem of solving (2) reduces to that of finding vectors $c = (c_1, \ldots, c_n)'$ and $d = (d_0, \ldots, d_{m-1})'$ to minimise

$$\frac{1}{n} \|\{y - (\Sigma_\lambda c + Td)\}\|^2 + \lambda c'\Sigma_\lambda c. \tag{8}$$

Taking partial derivatives in (8) and letting $M_\lambda = \Sigma_\lambda + nI$, necessary and sufficient conditions for optimality are

$$\begin{aligned} \Sigma_\lambda(M_\lambda c + Td) &= \Sigma_\lambda y, \\ T'(\Sigma_\lambda c + Td) &= T'y. \end{aligned}$$

Following Wahba (1990, chapter 1) and Gu (2002, chapter 2), for example, any vectors c and d satisfying

$$M_\lambda c + Td = y, \tag{9}$$

$$T'c = 0 \tag{10}$$

also satisfy the sufficient conditions, and this particular solution is also known to have good numeric properties.

To solve the latter equations, let

$$T = (Q_1 : Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix}$$

be the QR decomposition of T, where Q_1 is $n \times m$, Q_2 an $n \times (n-m)$, $Q = (Q_1, Q_2)$ is an orthogonal matrix, and R is upper triangular. From (10), $c = Q_2\gamma$ for an $(n-m)$ vector γ, and substitution into (9) yields the solution for γ and hence for c and d,

$$\begin{aligned} c &= Q_2(Q_2^{'} M_\lambda Q_2)^{-1} Q_2^{'} y, \\ Rd &= Q_1^{'}(y - M_\lambda c). \end{aligned}$$

Note that the estimate is linear in the data vector y. In particular, the smoothing matrix A_λ is specified by $\hat{y} = A_\lambda y = \Sigma_\lambda c + Td$. In view of (9), $y - A_\lambda y = nc$. Thus

$$I - A_\lambda = nQ_2(Q_2' M_\lambda Q_2)^{-1} Q_2'. \tag{11}$$

These computations can easily be extended to penalised weighted least squares and general L-spline penalties.

Finally, as with ordinary smoothing splines, one can also express the solution to the minimisation problem as the Bayes estimate under a partially improper prior. Let $X(t), t \in [0,1]$ be the mean zero Gaussian process defined by the stochastic integral

$$X(t) = \sqrt{b} \int_0^1 G_m(t,s) \lambda(s)^{-\frac{1}{2}} dW(s), \ t \in [0,1], \tag{12}$$

where $W(t)$ is standard Brownian motion and $b > 0$. Then $X(t_1), \ldots, X(t_n)$ has covariance matrix $b\Sigma_\lambda$ given by (6). Take the prior of f to be

$$f(t) = \sum_{j=0}^{m-1} d_j \phi_j(t) + X(t)$$

with a uniform improper prior on $(d_0, \ldots, d_{m-1})$. Letting $x = (X(t_1), \ldots, X(t_n))'$, the joint posterior distribution of (x, d) is then proportional to

$$\exp\left\{-\frac{1}{2\sigma^2}(y - x - Td)'(y - x - Td) - \frac{1}{2b}x'\Sigma_\lambda^{-1}x\right\}. \tag{13}$$

where T is given by (7). Since the posterior is Gaussian, the Bayes estimate of (x, d) is obtained by maximising (13). But this is equivalent to minimising (2) with $x = \Sigma_\lambda c$ and $b = n/\sigma^2$, proving the well-known equivalence between the Bayesian estimator under the partially improper reproducing kernel prior and the solution to the penalised equation (2). Hence, the spatially adaptive smoothing spline is a Bayes estimate under this partially improper prior. Moreover, it can be shown that the posterior variance of $z = x + Td$ is $\sigma^2 A_\lambda$, where A_λ is given by (11). Refer to Gu (2002, chapter 2) or Speckman and Sun (2003) for details. This formula will be used in Section 3 for the building of Bayesian "confidence" intervals. Note that these results can again be easily extended to penalised weighted least squares and general L-spline penalties.

2.1. Reproducing kernel for piecewise constant $\lambda(t)$

As a natural generalisation to the smoothing spline, we first suggest to model $\lambda(t)$, $t \in [0, 1]$, as a piecewise constant function with jumps at the adaptive smoothing knots τ_i, $i =$

$1, \ldots, S$. Defining $\tau_0 = 0$ and $\tau_{S+1} = 1$, suppose

$$\lambda(t) = \lambda_i \quad for\ t \ \in [\tau_i, \tau_{i+1}), \ i = 0, \ldots, S.$$

There are many reasons for which it appears sensible to model $\lambda(t), \ t \in [0,1]$, in this way:

- it provides a direct and simple generalisation of the smoothing spline derivation for which $\lambda(t)$ is constant over $[0,1]$.
- the resulting estimate (5) is a polynomial spline.
- the solution is analytically tractable.

Routine if tedious calculations lead to the following explicit representation for the new Green's function $K_\lambda(t_i, v)$. First fix t_i for $1 \le i \le n$, then for $v \in (t_i, 1]$,

$$K_\lambda(t_i, v) = \sum_{k=0}^{S} \sum_{j=1}^{m} \lambda_k^{-1} (-1)^j \left\{ \frac{(t_i - \tau_{k+1})_+^{m-1+j} (v - \tau_{k+1})^{m-j}}{(m-1+j)!(m-j)!} - \frac{(t_i - \tau_k)_+^{m-1+j} (v - \tau_k)^{m-j}}{(m-1+j)!(m-j)!} \right\} \tag{14}$$

and for $v < t_i$ with $v \in [\tau_l, \tau_{l+1})$,

$$\begin{aligned} K_\lambda(t_i, v) = & \\ & \sum_{k=0}^{S} \sum_{j=1}^{m} \lambda_k^{-1} (-1)^j \left\{ \frac{(t_i - \tau_{k+1})^{m-1+j} (v - \tau_{k+1})_+^{m-j}}{(m-1+j)!(m-j)!} - \frac{(t_i - \tau_k)^{m-1+j} (v - \tau_k)_+^{m-j}}{(m-1+j)!(m-j)!} \right\} \\ & + \lambda_l^{-1} (-1)^m \frac{(t_i - v)^{2m-1}}{(2m-1)!} \quad \text{for any } v \in (\tau_l, \tau_{l+1}) \end{aligned} \tag{15}$$

The detail for these calculations are given in Appendix A. In the next section we derive some characteristics of the estimate f_λ obtained with a piecewise constant assumption for

$\lambda(\cdot)$ and consider some issues linked with it.

2.2. *Some Properties*

Recall that the ordinary smoothing spline is well known to be a natural spline of order $2m$, that is, $f_\lambda(t) \in \pi^{2m-1}$ for $t \in [t_i, t_{i+1}]$, $i = 1, \ldots, n-1$, $f(t) \in \pi^{m-1}$ for $t \in [0, t_1]$ and $t \in (t_n, 1]$, and $f \in C^{2m-2}([0,1])$, where $\pi^k, k \in \mathbb{N}$ denotes the $(k+1)$-dimensional space of polynomials of degree k or less.

The properties of the spatially adaptive spline depend on smoothness properties of the kernel $K_\lambda(s,t)$ defined in (3). It can be seen from (4) that since $\lambda^{-1}(v)$ is assumed integrable $K_\lambda(t,\cdot) \in C^{m-1}[0,1]$ for any $t \in [0,1]$. Of course, K_λ may be smoother depending on additional assumptions on $\lambda(v)$. In this report we consider the case when $\lambda(v)$ is piecewise constant with possible jumps at $\tau_1, \ldots, \tau_S \subseteq \{t_1, \ldots, t_n\}$ for some $S \leq n$. We refer to $\{\tau_1, \ldots, \tau_S\}$ as adaptive smoothing knots. It follows that $K_\lambda(t_i, v) \in \pi^{2m-1}$ for $v \in (t_{j-1}, t_j)$ for $j \leq i$ and π^{m-1} for $v > t_i$. Thus $f_\lambda(t)$ is a polynomial spline of order $2m$ satisfying

$$f_\lambda(t) \in \pi^{2m-1} \text{ for } t \in [t_i, t_{i+1}], i = 1, \ldots, n-1, \text{ and } f_\lambda \in C^{m-1}[0,1]. \tag{16}$$

In addition, $f_\lambda \in \pi^{m-1}$ for $t \in (t_n, 1]$. It remains to be shown that $f_\lambda \in \pi^{m-1}$ for $t \in [0, t_1)$. The proof of this result is the same as that given in Wahba (1990, chapter 1) for the smoothing spline and we give it here: for $t < t_1$, we can write:

$$\begin{aligned} \sum_{i=1}^{n} c_i K_\lambda(t_i, t) &= \int_0^t G_m(t,u) \sum_{i=1}^{n} c_i \frac{(t_i - u)^{m-1}}{(m-1)!} du \\ &= 0 \ \text{ since } T'\boldsymbol{c} = \mathbf{0}, \end{aligned}$$

and the result is proved. Note that $T'\boldsymbol{c} = 0$ highlights the fact that $\boldsymbol{c}$ is a generalised divided difference with respect to the sequence of knots $t_1, \ldots, t_n$. Finally, if $t_i \neq \tau_j$ for any j, continuity of λ at t_i implies $f_\lambda \in C^{2m-2}(t_{i-1}, t_{i+1})$.

Equation (16) is interesting since it shows that the spatially adaptive smoothing spline has its m^{th}, ..., $(2m-1)^{th}$ derivatives (possibly) discontinuous whereas the smoothing spline only has its $(2m-1)^{th}$ derivative (possibly) discontinuous. Thus, the effect of moving from a constant $\lambda(\cdot)$ to a piecewise constant $\lambda(\cdot)$ is clear. Moreover, allowing lower derivatives to be possibly discontinuous renders the estimate more flexible since the model is less constrained.

An interesting relationship concerning the change in the derivatives of $f_\lambda(t)$ at the knots can be derived, which gives further insight into the behaviour of the spatially adaptive smoothing spline. Here we derive the relationship for a piecewise constant smoothing function $\lambda(t)$ with jumps at $\tau_1, .., \tau_S$ since it is the case we discuss in this paper. First, recall that $f_\lambda(t)$ is given by (5). Then, by differentiating $f_\lambda(t)$ in (5) $(m-1+l)$, $1 \le l \le m$, times, and using (4), we obtain an expression for $f_\lambda^{(m-1+l)}(t)$, $t \in (\tau_j, \tau_{j+1})$, $j = 1, ..., S$, where we take $\tau_0 = 0$ and $\tau_{S+1} = 1$,

$$f_\lambda^{(m-1+l)}(t) = \frac{1}{\lambda(\tau_j^+)} \left(\sum_{k=1}^{n} c_k \frac{(t_k - t)_+^{m-l}}{(m-l)!} \right), \; t \in (\tau_j, \tau_{j+1}), \; 1 \le l \le m, \tag{17}$$

where we define $\tau_j^+ = \lim_{h \to 0,\, h>0} (\tau_j + h)$. Similarly we define $\tau_j^- = \lim_{h \to 0,\, h<0} (\tau_j + h)$. Note also that $\lambda(\tau_j^+)$ could be replaced by $\lambda(\tau_{j+1}^-)$ in (17). Now, evaluating (17) at τ_j^+ and τ_j^- for $j = 1, \ldots, K$ leads to an interesting property of the spatially adaptive spline estimate concerning the change in the derivatives of $f_\lambda(t)$ at the knots, that is,

$$\left| f_\lambda^{(m-1+l)}(\tau_j^+) - f_\lambda^{(m-1+l)}(\tau_j^-) \right| = \left| \sum_{k=1}^{n} c_k \frac{(t_k - \tau_j)_+^{m-l}}{(m-l)!} \right| \left| \frac{\lambda(\tau_j^-) - \lambda(\tau_j^+)}{\lambda(\tau_j^-)\lambda(\tau_j^+)} \right| \quad \text{for } 1 \le l \le m \tag{18}$$

Equation (18) again expresses the fact that the model induces a change point in the

$\{m, m+1, \ldots, 2m-1\}$ derivatives of the process at each knot point. The size of the breaks are constrained partly through the quantity $\left|\{\lambda(\tau_j^-) - \lambda(\tau_j^+)\}/\{\lambda(\tau_j^-)\lambda(\tau_j^+)\}\right|$, so that the resulting solution solves the function minimisation in (2). This has clear connections with regression splines using the truncated polynomial basis and moreover by allowing the knot points of $\lambda(\cdot)$ to be random would lead to a free knot smoothing spline solution of an explicit function optimisation problem.

3. Results

We now consider a set of different examples in order to show how the spatially adaptive spline performs. Its performances will be highlighted by comparing it with the conventional "best" fit smoothing spline obtained on the same examples. The comparison will be visual and will also involve comparing the equivalent degrees of freedom for each model, that is the trace of the smoothing matrix, as well as the kernels induced and the value of the estimated variance for the noise process.

Three examples will be considered. The first two are simulated data sets, which have been chosen because of their inhomogeneous smoothing structure. The third example is designed to show that the spatially adaptive spline performs well when the smoothing structure is homogeneous.

In all examples, we consider four situations, $S = 1,\ 5,\ 10,\ 20$, where the solution with $S = 1$ corresponds to the usual smoothing spline. For $S = 5,\ 10,\ 20$ the smoothing knots are taken on the corresponding 20^{th}, 10^{th} and 5^{th} percentiles of the data. That is, we use 5, 10 and 20 knots positioned on the quantiles of the empirical cumulative distribution function of the $t_i, i = 1, ..., n$. The value of $\lambda(t)$ between two successive knots $[t_i, t_{i+1}]$ is taken to be equal to $\lambda(t_i)$. The best fits obtained for each model together with some of their characteristics for the first three examples are given in Table 1, Subsection 3.5.

For each example we also plot the 95% confidence interval for the "best" spatially adaptive spline fit. To do so, we use methods for building confidence intervals on sampling points in the smoothing spline literature. First we derive interval estimates under the equivalent Bayes model (12). The reader may refer to the end of Subsection 2.1 for more detail. Then it can be shown that these estimates have a certain across the function coverage property when the regression function is fixed and smooth. That is, if one defines the average coverage proportion as

$$ACP(\alpha) = \frac{1}{n}\# \left[i : \{\eta_\lambda(t_i) - \eta(t_i)\} < z_{\alpha/2}\hat{\sigma} a_{i,i}^{\frac{1}{2}} \right],$$

where $a_{i,i}$ is the $(i,i)^{th}$ element of the smoothing matrix. Then, simulation results in Wahba (1983) suggest that for n large,

$$E\{ACP(\alpha)\} \approx 1 - \alpha,$$

where $\eta(t)$ is considered fixed. One can thus compare these intervals estimates with standard parametric confidence intervals. We refer the reader to Wahba (1983), Gu (2002, chapter 3) or Nychka (1988) for more detail.

In order to run our model, we need to estimate the smoothing piecewise constant function $\lambda(t)$. The general issue of how to best estimate $\lambda(t)$ is beyond the scope of this paper, whose main aim is to present the theoretical framework of spatially adaptive splines. We suggest however the following procedure as what we believe to be a good first attempt to model $\lambda(t)$.

3.1. *Estimation of* $\lambda(t)$

We follow the smoothing spline literature (Wahba 1990, chapter 4) and only consider possible changes in the smoothness at the given data points, that is at $\tau_i \in \{t_1, ..., t_n\}$, $i = 1, ..., S$, $S \leq n$. Then, we estimate $\lambda(t)$ by minimising a possibly weighted and penalised version of Generalised Cross Validation function corresponding to (2) and defined by

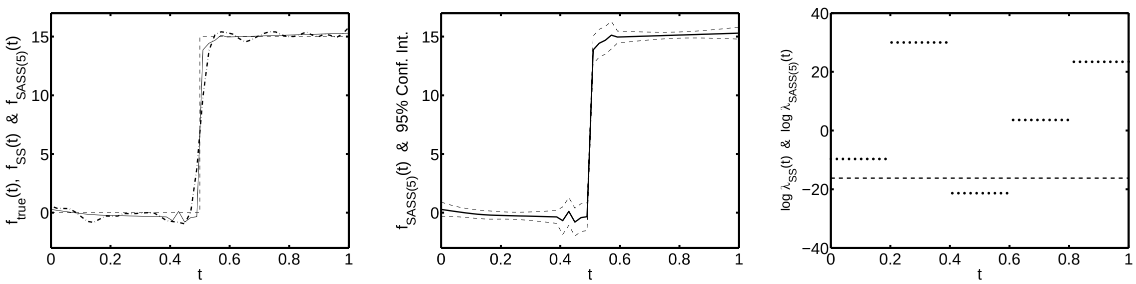
$$V(\lambda) = \frac{1}{n}\|(I - A_\lambda)\boldsymbol{y}\|^2 / \left[\frac{1}{n}\mathrm{tr}\{I - q(S)A_\lambda\}\right]^2, \tag{19}$$

where A_λ is the smoothing matrix previously defined in Section 2 and $q(S) \geq 1$ is an optional cost factor, depending on the number of jumps K allowed, which is sometimes used when fitting models with a large number of parameters. This factor was termed "inflation degrees" of freedom by Luo and Wahba (1997), in the context of hybrid adaptive splines, and was also used by Friedman and Silverman (1989), in the context of adaptive regression splines. It allows one to increase the penalty on the number of degrees of freedom and thus, when the dimension of the parameter space is big, to prevent the model from over-fitting. Its modelling is difficult as it is depends on the nature of the data set as well as S. Setting $q(S) = 1$ yields the original Generalised Cross-Validation of Craven and Wahba (1979) and we note, using (11), that $\mathrm{tr}(I - A_\lambda) = n\mathrm{tr}(Q_2' M_\lambda Q_2)^{-1}$. Unfortunately, unlike in the case of ordinary spline smoothing, there is no simpler form obtained by diagonalising A_λ, and $\mathrm{tr}(A_\lambda)$ must be recomputed for every choice of values $\lambda_1, \ldots, \lambda_S$, where S is the number of jumps allowed.

We minimise (19) using an optimisation procedure to find $\{\hat{\lambda}_i\}_{i=1}^{S}$. In our experiments we have found the Nelder-Mead algorithm on $\log \lambda_j$ to be efficient. We refer the reader to Fletcher (1987, chapter 2) for details on the Nelder-Mead algorithm. In the examples to follow, we do not consider the modelling of $q(S)$ as a function of S but only show how the simulations reinforce the need for penalising the number of degrees of freedom.

3.2. A simulated Heavyside function

We first describe a simulated one-dimensional Heavyside example where we take $n = 50$ data points sampled regularly on [0,1] and consider a "noisy" sample of it with $\sigma = 0.7$. The true function is plotted in dashed line in Fig. 1(a). Clearly, the smoothness of the process varies significantly at the point $t = 0.5$. The results obtained for the "best" fit smoothing spline and the "best" fit spatially adaptive spline with 5 jumps are plotted in Fig. 1(a) and listed in Table 1. We have chosen the model with 5 jumps here since it can be seen from Table 1 that the models with 10 and 20 jumps have a tendency to over-fit. This effect will be discussed further in Subsection 3.5 in relation to the smoothness penalty $q(S)$.



(a) Data and best fits (b) 95% Confidence Interval (c) Estimated $\lambda(t)$'s

Fig. 1. Heavyside function, Subsection 3.2. (a) represents the true function (- -), the "best" fit Smoothing Spline (-.), SS, and Spatially Adaptive Smoothing Spline with 5 jumps (-), SASS(5). (b) presents the "best" Spatially Adaptive Smoothing Spline fit with 5 jumps (-), SASS(5), and the 95% Confidence Interval (- -). (c) shows the value of $\log \lambda(t)$ for the Spatially Adaptive Smoothing Spline with 5 jumps (*), SASS(5), and of $\log \lambda$ for the Smoothing Spline (- -), SS.

It is clear from Fig. 1(a) and Fig. 1(b) that the spatially adaptive spline is able to adapt to the changes of smoothness in the structure of the data in a more flexible way than the smoothing spline. Indeed, in the areas away from $t = 0.5$, the resulting spline is very smooth and close to the true piecewise constant function. The smoothing spline on the other hand doesn't perform as well since it under-smooths in these regions because it aims at catching the high frequency variation in the structure around $t = 0.5$. A single smoothing parameter is thus clearly not enough here to catch the variation in the smoothness of the structure of the data.

This is also highlighted in Fig. 1(c), where we have plotted on the log scale the values $\log\{\lambda(t)\}$ at each knot for the "best" fit spatially adaptive splines with 5 jumps together with the value of the corresponding parameter for the smoothing spline plotted in dashed line. Indeed, Fig. 1(c) shows that the values of the smoothing parameters $\lambda_i, i = 1, \ldots, n$ are greater than the smoothing spline parameter λ in regions far from $t = 0.5$ and smaller around $t = 0.5$. From this behaviour, it appears that the smoothing spline parameter is simply a compromise between the different amounts of smoothness in the data. Note that the variability of the spatially adaptive smoothing spline in Fig. 1(a) around $t = 0.45$ is the result of the use of a constant value for λ around $t = 0.5$ which is smaller than the smoothing spline parameter value for reasons just mentioned.

Finally, it is important to emphasise that Fig. 1(c) reveals another characteristic of the spatially adaptive spline model which is its interpretability. Indeed the values of the smoothing parameters $\lambda_i, i = 1, \ldots, n$ give information on the smoothness structure of the data, not through their values directly, since these are difficult to interpret, but rather via their behaviour relative one to another over the space. This enables one to gain insight into the smoothness structure of the data.

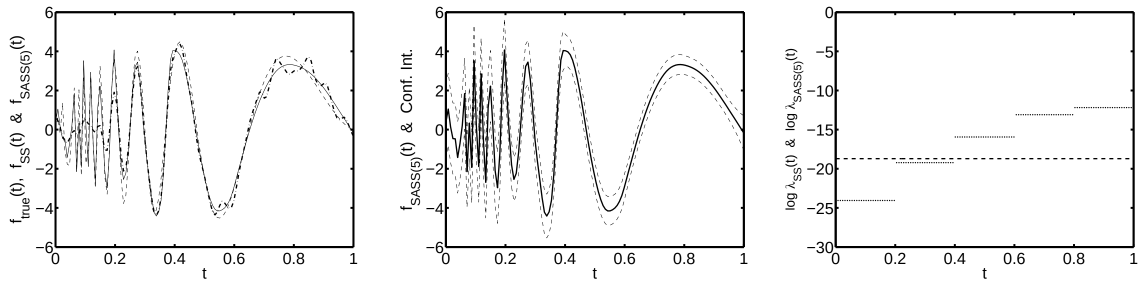
3.3. The Doppler function

The second simulated data example is the Doppler function from Donoho and Johnstone (1995),

$$f(t) = \{t(1-t)\}^{1/2} \sin\{2\pi(1+a)/(t+a)\}, t \in [0,1],$$

where a is a constant. We generated $n = 128$ data points regularly on $[0, 1]$ with $a = 0.05$. Then, following Donoho and Johnstone (1995), the points were re-scaled to show a variance of 7 to which $N(0, 1)$ noise was added. The "true" data are plotted in Fig. 2(a). Clearly, the smoothness in the structure of the data increases with t, thus revealing non-constant smoothness over the space.

The results for the "best" fits obtained for the smoothing spline and the spatially adaptive splines obtained with 5 jumps are shown in Fig. 2(a) and the 95% Confidence Interval for the spatially adaptive spline fit with $S = 5$ are given in Fig. 2(b). As in the previous example, it appears both visually and from the behaviour of the function



(a) Data and best fits (b) 95% Confidence Interval (c) Estimated $\lambda(t)$'s

Fig. 2. Doppler function, Subsection 3.3. (a) represents the true function (- -), the "best" fit Smoothing Spline (-.), SS, and Spatially Adaptive Smoothing Spline with 5 jumps (-), SASS(5). (b) presents the "best" Spatially Adaptive Smoothing Spline fit with 5 jumps (-), SASS(5), and the 95% Confidence Interval (- -). (c) shows the value of $\log \lambda(t)$ for the Spatially Adaptive Smoothing Spline with 5 jumps (*), SASS(5), and of $\log \lambda$ for the Smoothing Spline (- -), SS.

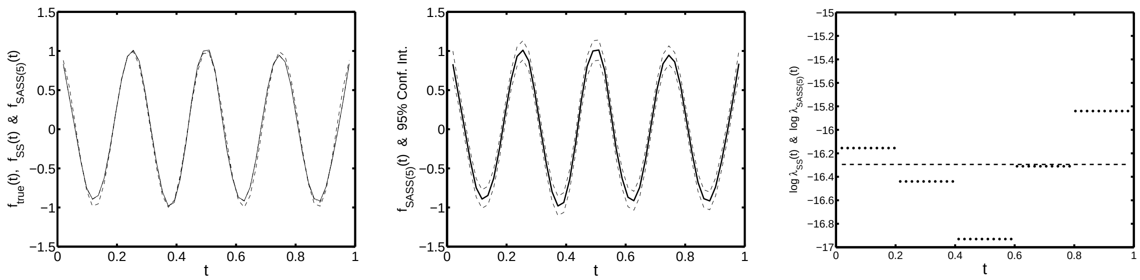
$\log\{\lambda(t)\}$ given in Fig. 2(c) that the spatially adaptive spline with 5 jumps is able to adapt to the local smoothness structure of the data and thus to capture the features of the latter more precisely than the smoothing spline, whose parameter appears again as a compromise between the various degrees of smoothness present.Indeed, as can be seen in Fig. 2(a), the smoothing spline clearly under-smooths for $t \geq 0.5$ and over-smooths for $t \leq 0.2$. The spatially adaptive smoothing spline allows one to greatly attenuate these features.

These results are also confirmed by the estimated variances of the noise process for different models given in Table 1. Note that although we have not plotted its fit here, the spatially adaptive spline with 10 jumps has a behaviour very similar to that with 5 jumps.

3.4. The Cosine function

We consider here a simple cosine wave data set of $n = 50$ points in order to tackle the case of an homogeneous smoothness structure. Indeed, we feel it is important to illustrate that the spatially adaptive spline performs as well as the smoothing spline when the overall smoothness can be taken to be constant. The true data is given in Fig. 3(a) and we consider a "noisy" sample of it with $\sigma = 0.1$.

The resulting "best" fit smoothing spline and spatially adaptive spline with 5 jumps are given in Fig. 3(a), from which it appears visually that the spatially adaptive spline is able to detect the relatively constant smoothness in the structure of the data. This



(a) Data and best fits (b) 95% Confidence Interval (c) Estimated $\lambda(t)$'s

Fig. 3. Cosine function, Subsection 3.3. (a) represents the true function (- -), the "best" fit Smoothing Spline (-.), SS, and Spatially Adaptive Smoothing Spline with 5 jumps (-), SASS(5). (b) presents the "best" Spatially Adaptive Smoothing Spline fit with 5 jumps (-), SASS(5), and the 95% Confidence Interval (- -). (c) shows the value of $\log \lambda(t)$ for the Spatially Adaptive Smoothing Spline with 5 jumps (*), SASS(5), and of $\log \lambda$ for the Smoothing Spline (- -), SS.

behaviour is also confirmed by Fig. 3(b) and Fig. 3(c). Indeed, it is seen in Fig. 3(c) that the different values of $\log\{\lambda(t)\}$ are very close one to another and to the value of the smoothing spline parameter since the difference in the log-scale on the vertical axis is very small. This shows that the spatially adaptive smoothing spline is able to recover a structure exhibiting constant smoothness, which is what we were aiming at checking through this example.

3.5. Simulation results

Table 1 gives the complete results for the three previous examples of Subsections 3.2, 3.3 and 3.4 for the "best" fit smoothing spline and the "best" fit spatially adaptive spline obtained when 5, 10 and 20 jumps are allowed. We have given the results for $q(S) = 1.0$ for each model. The noise variance estimates were obtained by dividing the residual sum of squares by the number of degrees of freedom for the noise component obtained from the fit. We refer the reader to Green and Silverman (1994, chapter 2) for more detail. The simulations were run on a Unix machine and the computational times for the spatially adaptive models were calculated when taking the value of the smoothing spline parameter as an initial state.

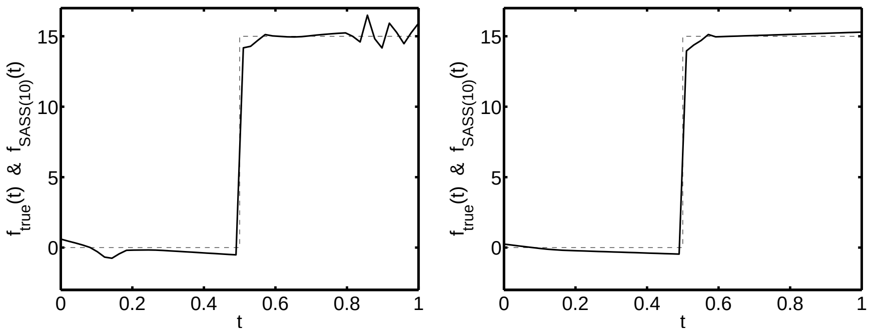
Examples	Results	SS	SAS(5)	SAS(10)	SAS(20)
Heavyside function (Section 3.2)	comp. time (s)	2.46	10.76	16.83	58.84
	deg. of freedom	21.47	14.39	20.54	24.75
	$\hat{\sigma}^2$ ($\sigma^2_{true} = 0.49$)	1.68	0.38	0.24	0.17
Doppler function (Section 3.3)	comp. time (s)	27.72	85.27	421.37	859.63
	deg. of freedom	39.17	39.32	46.03	70.77
	$\hat{\sigma}^2$ ($\sigma^2_{true} = 1.00$)	1.63	0.96	0.87	0.34
Cosine function (Section 3.4)	comp. time (s)	1.04	5.67	32.96	72.44
	deg. of freedom	21.26	21.54	22.82	32.97
	$\hat{\sigma}^2$ ($\sigma^2_{true} = 0.01$)	0.0089	0.0092	0.0081	0.0025

Table 1. Subsection 3.5. Comparisons of the computational time, the number of degrees of freedom and the estimated variance for the Smoothing Spline and the Spatially Adaptive Smoothing Spline with 5, 10 and 20 jumps, denoted by SS, SASS(5), SASS(10) and SASS(20) respectively, for the three examples of Subsections 3.2, 3.3.and 3.4, with $q(S) = 1.0$

From this table and from the analysis given in Subsections 3.2, 3.3 and 3.4 we clearly see that when $q(S) = 1.0$ the model with 5 jumps performs well compared to the smoothing spline. This is highlighted by the results obtained for the estimated noise variance. It is also interesting to note that the number of degrees of freedom for the smoothing spline and the "best" fit spatially adaptive spline, which was obtained for 5 jumps in the given examples in the previous subsections, are very close, something that was confirmed in other simulations. Thus it seems that the spatially adaptive spline is able to make "better use" of the degrees of freedom than the smoothing spline and that this is how flexibility is introduced rather than through adding more degrees of freedom.

As previously mentioned in Subsection 3.2 and as can also be seen from Table 1, models allowing 10 or 20 jumps tend to over-fit the data. This is due to the fact that introducing more parameters into our model allows for more degrees of freedom, and if not penalised in our estimation procedure, this leads to over-fitting of the data. Recalling the role of the parameter $q(S)$ in Subsection 3.1, (19), it is now clear that when the number of jumps S is allowed to increase, the parameter $q(S)$ should also increase as a function of S and the data set.

As an illustration of this phenomenon, we return to the Heavyside simulation example of Subsection 3.2. We give the "best" spatially adaptive spline fit when $q(10) = 1.0$, $q(10) = 1.2$ and $q(10) = 1.4$ in Fig. 4.



(a) 10 jumps and $q(10) = 1.0$ (b) 10 jumps and $q(10) = 1.4$

Fig. 4. Subsection 3.5. True data (- -) and "best" Spatially Adaptive Smoothing Spline fit with 10 jumps (-), SASS(10) when different values for $q(10)$ are taken. (a) corresponds to $q(10) = 1.0$ and (b) to $q(10) = 1.4$.

Clearly, when allowing for 10 jumps, optimising with $q(10) = 1.4$ leads to a better fit (i.e. less over-fit) than with $q(10) = 1.0$. The values of $\hat{\sigma}_{true}$ obtained were 0.24 for $q(S) = 1.0$ and 0.38 for $q(S) = 1.4$, thus confirming the need to penalise $q(S)$ with S. This example shows that the degrees of freedom must be penalised. However, one must be careful as the penalty chosen in any case must also take into account the data set considered. This was confirmed on other unreported simulations.

4. Discussion

In this paper we have derived a spatially adaptive smoothing spline obtained as the solution of a mathematical optimisation problem (2) when the smoothing function $\lambda(t)$ is piecewise constant. We have shown through some examples that our model is able to adapt to changes in the smoothness structure. Comparisons with the smoothing spline have allowed us to gain further insight into the features of both our model and the smoothing spline. Although our model has some resemblance with "free-knot" models in that a jump in the smoothing function corresponds to a break in the m^{th} derivative of the polynomial spline fit of degree $2m - 1$ corresponding to (2), it is different in the sense that our model is an explicit solution to a mathematical approximation problem. Furthermore, it is somewhat more interpretable since $\lambda(t)$ quantifies the smoothness in a precise manner.

Many issues remain open. First, we have modelled the smoothing function as piecewise constant in order to obtain a polynomial spline fit and also as a direct generalisation to the smoothing spline. However, one could model the smoothing function in a different manner, thus not obtaining a polynomial spline anymore but an L-spline in the general meaning of the term. Note that it might be difficult or computationally too expensive to obtain the reproducing kernel (3) analytically. In that case, one could use numerical approximations to (3). Another solution would be for one to consider perhaps performing the minimisation over some other, possibly finite-dimensional, subspace which has good approximation properties with respect to the problem at hand. For instance, the space of natural splines (Eubank 1999, chapter 5) of degree m could be considered for instance. The authors hope to give more detail on this issue in a forthcoming paper.

The estimation of $\lambda(t)$ is another open problem. Indeed, as we have discussed in Subsection 3.1 and partly shown in Subsection 3.5, this problem is far from being straightforward. It is clear for instance that one would need to penalise the number of degrees of freedom in the Generalised Cross-Validation when allowing for more jumps. It seems that techniques taken from other spatially adaptive curve fitting methods referred to in Subsection 3.1 could be used and adapted here. The difficulty however comes from the fact that unlike with the smoothing spline, the fit must be recomputed at each step since the reproducing kernels change with $\lambda(t)$. This can be computationally expensive, especially when n is big. It is therefore necessary that the estimation procedure for $\lambda(t)$ take this latter problem into account.

Finally it is clear that our method could be generalised to higher dimensions provided one is able to do the calculations and integrals needed in order to obtain analytical forms for the reproducing kernels or at least some good, possibly numerical, approximation to it. Such a generalisation raises the issue of the extension of the method we have proposed for modelling $\lambda(t)$. Tessellation methods (Okabe et al. 2000, chapter 2) might provide one with a possible generalisation of the piecewise constant assumption to higher dimensions although this would have to be looked at in greater detail.

APPENDIX A. DERIVATION OF $K_\lambda(\cdot,\cdot)$ FOR PIECEWISE CONSTANT $\lambda(\cdot)$.

We derive in this section the expressions (14) and (15) obtained for $K_\lambda(t_i, v),\ v \in [0,1],\ i = 1, \ldots, n$. The derivation is simple and relies essentially on successive applications of the *Integration-by-parts* method. Let us first consider the case $v \in [t_i, 1],\ i = 1, \ldots, n$ corresponding to (14).

$$
\begin{aligned}
K_\lambda(t_i, v) &= \int_0^1 \lambda(u)^{-1} G_m(t_1, u) G_m(v, u) du,\ v \in [t_i, 1],\ i = 1, \ldots, n \\
&= \int_0^1 \lambda(u)^{-1} \frac{(t_i - u)_+^{m-1}}{(m-1)!} \frac{(v-u)_+^{m-1}}{(m-1)!} du \\
&= \sum_{k=0}^{S} \lambda_k^{-1} \int_{\tau_k}^{\tau_{k+1}} \frac{(t_i - u)_+^{m-1}}{(m-1)!} \frac{(v-u)^{m-1}}{(m-1)!} du
\end{aligned} \tag{20}
$$

Now, we use an Integration-by-parts to obtain

$$
\begin{aligned}
\int_{\tau_k}^{\tau_{k+1}} \frac{(t_i - u)_+^{m-1}}{(m-1)!} \frac{(v-u)^{m-1}}{(m-1)!} du &= -\left\{ \frac{(t_i - u)_+^{m}}{m!} \frac{(v-u)^{m-1}}{(m-1)!} \right\}_{\tau_k}^{\tau_{k+1}} \\
&- \int_{\tau_k}^{\tau_{k+1}} \frac{(t_i - u)_+^{m}}{m!} \frac{(v-u)^{m-2}}{(m-2)!} du
\end{aligned}
$$

And hence, a straightforward recursion gives

$$
\begin{aligned}
\int_{\tau_k}^{\tau_{k+1}} \frac{(t_i - u)_+^{m-1}}{(m-1)!} \frac{(v-u)^{m-1}}{(m-1)!} du &= \sum_{j=1}^{m} (-1)^j \left\{ \frac{(t_i - u)_+^{m-1+j}}{(m-1+j)!} \frac{(v-u)^{m-j}}{(m-j)!} \right\}_{\tau_k}^{\tau_{k+1}} \\
&= \sum_{j=1}^{m} (-1)^j \left\{ \frac{(t_i - \tau_{k+1})_+^{m-1+j}}{(m-1+j)!} \frac{(v - \tau_{k+1})_+^{m-j}}{(m-j)!} \right\} \\
&- \sum_{j=1}^{m} (-1)^j \left\{ \frac{(t_i - \tau_k)_+^{m-1+j}}{m-1+j} \frac{(v - \tau_k)_+^{m-j}}{(m-j)!} \right\}
\end{aligned} \tag{21}
$$

And, by introducing (21) in (20) we obtain (14).

We now move on to the proof of (15). Consider now that $v \in [0, t_i],\ i = 1, \ldots, n$. Then by a similar argument as for the derivation of (14) we have

$$
\begin{aligned}
K_\lambda(t_i,v) &= \int_0^1 \lambda(u)^{-1} G_m(t_1,u) G_m(v,u) du, \;\; v \in [0,t_i], \;\; i=1,\ldots,n \\
&= \int_0^1 \lambda(u)^{-1} \frac{(t_i-u)_+^{m-1}}{(m-1)!} \frac{(v-u)_+^{m-1}}{(m-1)!} du \qquad (22)
\end{aligned}
$$

Now, let the index l denote the knot such that:

$$
\tau_l = \max_{i \in [1,\ldots,S]} \{(\tau_i \leq v) \text{ and } (\tau_{i+1} > v)\}
$$

Then, we can rewrite (22) as

$$
K_\lambda(t_i,v) = \sum_{k=0}^{l-1} \lambda_k^{-1} \int_{\tau_k}^{\tau_{k+1}} \frac{(t_i-u)^{m-1}}{(m-1)!} \frac{(v-u)^{m-1}}{(m-1)!} + \lambda_l^{-1} \int_{\tau_l}^{v} \frac{(t_i-u)^{m-1}}{(m-1)!} \frac{(v-u)^{m-1}}{(m-1)!}
$$

And we have, using successive Integration-by-parts as for the derivation of (21), we obtain

$$
\begin{aligned}
\int_{\tau_l}^{v} \frac{(t_i-u)^{m-1}}{(m-1)!} \frac{(v-u)^{m-1}}{(m-1)!} &= \sum_{j=1}^{m} (-1)^j \left\{ \frac{(t_i-u)^{m-1+j}(v-u)^{m-j}}{(m-1+j)!(m-j)!} \right\}_{\tau_l}^{v} \\
&= \sum_{j=1}^{m-1} (-1)^{j+1} \frac{(t_i-\tau_l)^{m-1+j}(v-\tau_l)^{m-j}}{(m-1+j)!(m-j)!} \\
&\quad + (-1)^m \left\{ \frac{(t_i-v)^{2m-1}}{(2m-1)!} - \frac{(t_i-\tau_l)^{2m-1}}{(2m-1)!} \right\} \qquad (23)
\end{aligned}
$$

And using the results given by (23) and (21), we have

$$K_\lambda(t_i, v) =$$
$$\sum_{k=0}^{l-1}\sum_{j=1}^{m} \lambda_k^{-1}(-1)^j \left\{ \frac{(t_i-\tau_{k+1})^{m-1+j}(v-\tau_{k+1})^{m-j}}{(m-1+j)!(m-j)!} - \frac{(t_i-\tau_k)^{m-1+j}(v-\tau_k)^{m-j}}{(m-1+j)!(m-j)!} \right\} +$$
$$\sum_{j=1}^{m-1} \lambda_l^{-1}(-1)^{j+1} \left\{ \frac{(t_i-\tau_l)^{m-1+j}(v-\tau_l)^{m-j}}{(m-1+j)!(m-j)!} \right\} + \frac{(-1)^m}{\lambda_l} \left\{ \frac{(t_i-v)^{2m-1}}{(2m-1)!} - \frac{(t_i-\tau_l)^{2m-1}}{(2m-1)!} \right\}$$

which can be rewritten as:

$$K_\lambda(t_i, v) = \sum_{k=0}^{S}\sum_{j=1}^{m} \lambda_k^{-1}(-1)^j \left\{ \frac{(t_i-\tau_{k+1})^{m-1+j}(v-\tau_{k+1})_+^{m-j}}{(m-1+j)!(m-j)!} - \frac{(t_i-\tau_k)^{m-1+j}(v-\tau_k)_+^{m-j}}{(m-1+j)!(m-j)!} \right\} \quad (24)$$

$$+\ \lambda_l^{-1}(-1)^m \frac{(t_i-v)^{2m-1}}{(2m-1)!} \quad \text{for any } v \in [0, t_i],\ i = 1, \ldots, n \text{ and } v \in (\tau_l, \tau_{l+1})$$

thus obtaining (15).

References

Abramovich, F. and Steinberg, D. M. (1996). Improved inference in nonparametric regression using l_k-smoothing splines, *J. of Stat. Plan. and Inf.* **49**: 327–341.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation, *Numer. Math.* **31**: 377–403.

Cummins, D., Filloon, T. and Nychka, D. (2001). Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage, *J. Amer. Stat. Assoc.*

96: 233–246.

Denison, D., Holmes, C., Mallick, B. and Smith, A. (2002). *Bayesian methods for nonlinear classification and regression.*, Chichester: Wiley.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Stat. Assoc.* **90**: 1200–1224.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*, Marcel Dekker.

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

Fletcher, R. (1987). *Practical Methods of Optimization. 2nd edition*, John Wiley & Sons.

Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling, *Technometrics* **31**: 3–39.

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

Gu, C. (2002). *Smoothing Spline ANOVA Models*, Springer.

Hansen, M. and Kooperberg, C. (2002). Spline adaptation in extended linear models, with discussion, *Stat. Sci.* **17**: 2–51.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall, London, chapter 4.

Luo, Z. and Wahba, G. (1997). Hybrid adaptive splines, *J. Amer. Stat. Assoc.* **92**: 107–116.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines, *J. Amer. Stat. Assoc.* **81**: 96–103.

Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams*, Chichester: Wiley. 2nd edn.

Ruppert, D. and Carroll, R. J. (2000). Spatially-adaptive penalties for spline fitting, *Aust. and New Zeal. J. Stat.* **42**: 205–223.

Smith, M. and Kohn, R. (1996). Nonparametric regression using bayesian variable selection, *J. Econom.* **75**: 317–344.

Speckman, P. and Sun, D. (2003). Fully bayesian spline smoothing and intrinsic autoregressive priors, *Biometrika* **90**: 289–302.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline, *J. Roy. Stat. Soc. B* **45**: 133–150.

Wahba, G. (1990). *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM.

Wahba, G. (1995). Discussion of "wavelet shrinkage: Asymptopia?" by d. l. donoho et al., *J. Roy. Stat. Soc. B* **57**: 360–361.

Wood, S., Jiang, W. and Tanner, M. (2002). Bayesian mixture of splines for spatially adaptive nonparametric regression, *Biometrika* **89**: 513–521.