

Least squares splines with variable knots using a smoothing spline basis

T. Prvan*

(Received 7 August 2000)

Abstract

A method for constructing a least squares spline with variable knots using a smoothing spline basis is presented. The inherent stability problems of the usual formulation of the smoothing spline are avoided by using the Kalman Filter, the Fixed-interval, discrete-time smoother and the interpolation smoother as computational tools.

*School of Mathematics and Statistics, University of Canberra, ACT 2601,
AUSTRALIA

⁰See <http://anziamj.austms.org.au/V42/CTAC99/Prva> for this article and ancillary services, © Austral. Mathematical Soc. 2000. Published 27 Nov 2000.

Contents

1	Introduction	C1200
2	Stochastic formulation of a smoothing spline	C1203
3	Method	C1206
3.1	Constructing the smoothing spline basis	C1207
3.2	Least squares spline with fixed knots using a smoothing spline basis	C1208
3.3	Least squares spline with variable knots using a smoothing spline basis	C1208
3.4	Extensions	C1209
3.5	Remarks	C1210
4	A simulation study	C1212
5	Conclusion	C1215
	References	C1215

1 Introduction

Least squares is a familiar approach to fitting a smooth curve to the n data points (t_i, y_i) , $i = 1, \dots, n$. The least squares approximation is the best

approximation with respect to a norm whose square product is given by

$$\sum_{i=1}^n (y_i - f(t_i))^2$$

where $f(t)$ is our curve fitted to the data. Often

$$f(t) = \sum_{i=1}^n a_i B_{i,k}(t, \zeta)$$

where $B_{i,k}(t, \zeta)$ is a B-spline of order k with knot points ζ_1, \dots, ζ_n denoted by ζ . Typically the B-spline basis is cubic ($k = 4$). The knots do not need to coincide with data points.

In Jupp [4] and references contained therein, the objective function associated with least squares splines is minimised over the knot points as well as the coefficients of the B-spline basis. To obtain the least squares spline with variable knots they minimise

$$S(\zeta, \mathbf{a}) = \sum_{i=1}^N (y_i - f(t_i))^2 = \sum_{i=1}^N \left(y_i - \sum_{j=1}^n a_j B_{k,j}(t_i; \zeta) \right)^2 \quad (1)$$

over \mathbf{a}, ζ . Here $\mathbf{a}^T = [a_1, \dots, a_n]$. The IMSL subroutine DBSVLS [3] may be used to compute this variable knot B-spline least squares approximation.

Instead of using a B-spline basis we now want to consider a basis of smoothing splines. Recall that B-splines have compact support whereas

smoothing splines do not. A smoothing spline also has the additional parameter λ which controls the level of smoothness. A stochastic formulation of the smoothing spline will be used. This involves using the Kalman Filter, the Fixed-interval, discrete-time smoother and the Interpolation smoother as computational tools and avoids the inherent stability problems of the usual formulation of the smoothing spline (see, for example, Osborne and Prvan [5]). Besides being a more stable implementation, this formulation can be easily extended to vary the continuity properties and degree of the piecewise polynomial; for more details refer to Prvan [7].

Schwetlick and Schütze [8] minimise the function associated with the smoothing spline over the coefficients of the B-spline basis, knots and smoothing parameter to produce a regression spline with variable knots. The same constraint as in the approach presented in this paper is placed on the knots; that is, the knots are ordered. Our approach is conceptually simpler with the smoothing parameter being introduced into the basis.

Details about the stochastic formulation of the smoothing spline and related curves are given in Section 2. The construction of a smoothing spline basis is outlined in Section 3, along with the least squares spline with variable knots using a smoothing spline basis. Section 4 contains a simulation study.

2 Stochastic formulation of a smoothing spline

The data $(t_1, y_1), \dots, (t_n, y_n)$ are given. A smoothing spline f is the minimiser of

$$\sum_{i=1}^n (y_i - f(t_i))^2 + \mu \int_{t_1}^{t_n} (f^{(p)}(t))^2 dt. \quad (2)$$

The resultant curve fitted to the data is a piecewise polynomial of degree $2p - 1$. The data are assumed to be decomposed as a signal plus noise

$$y_i = f(t_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n.$$

The resultant curve is a piecewise polynomial of degree $2p - 1$ with $2p - 2$ continuous derivatives. Wecker and Ansley [10] presented a stochastic formulation of the smoothing spline utilising a result by Wahba [9]. She showed that a polynomial smoothing spline is the solution to the stochastic differential equation

$$\frac{d^p x}{dt^p} = \sigma \sqrt{\lambda} \frac{d\omega}{dt}. \quad (3)$$

Here $\omega(t)$ is a Wiener process (see for example Billingsley [2] with unit dispersion parameter, $\lambda = \frac{1}{\mu}$ and $\mathbf{x}(t_1) = [x(t_1), \dots, x^{(p-1)}(t_1)]^T$ has a diffuse prior (i.e. $\mathbf{x}(t_1) \sim N(\mathbf{0}, \gamma^2 I_p)$ and $\gamma^2 \rightarrow \infty$). The solution is

$$x(t) = \lim_{\gamma^2 \rightarrow \infty} x(t | n),$$

where $x(t)$ is the first entry of $\mathbf{x}(t)$. The quantity $x(t | n)$ is the expected value of $x(t)$ conditioned on the data y_1, \dots, y_n . The stochastic differential

equation can be written in the matrix companion form

$$\frac{d\mathbf{x}(t)}{dt} = \begin{pmatrix} \mathbf{0}_{p-1} & I_{p-1} \\ 0 & \mathbf{0}_{p-1}^T \end{pmatrix} \mathbf{x}(t) + \sigma\sqrt{\lambda} \begin{pmatrix} \mathbf{0}_{p-1} \\ \frac{d\omega}{dt} \end{pmatrix}. \quad (4)$$

The fundamental matrix solution of the associated homogeneous differential equation is denoted by $T(t_i, t_1)$. The solution of (4) can be written recursively as

$$\mathbf{x}(t_i) = T(t_i, t_{i-1})\mathbf{x}(t_{i-1}) + \mathbf{u}(t_i, t_{i-1}) \text{ for } i = 2, \dots, n. \quad (5)$$

The random vector $\mathbf{u}(t_i, t_{i-1})$ is normally distributed with zero mean and covariance $\sigma^2\lambda\Omega(t_i, t_{i-1})$, where

$$\Omega(t_i, t_{i-1}) = \int_{t_{i-1}}^{t_i} T(t_i, s)\mathbf{e}_p\mathbf{e}_p^T T(t_i, s)^T ds. \quad (6)$$

The observation equation now becomes

$$y_i = \mathbf{e}_1^T \mathbf{x}(t_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1 \dots n. \quad (7)$$

The notation \mathbf{e}_j is used to denote a p -vector having all zeros except for a 1 in the j -th position. The diffuse prior can be dealt with explicitly by setting γ sufficiently large. For details on how to implicitly deal with the diffuse prior refer to Wecker and Ansley [10].

The Kalman filter, the Fixed-interval, discrete-time smoother and the Interpolation smoother are implemented on the state space formulation (7) and (5) to obtain $\mathbf{x}(t | n)$ and hence the smoothing spline and its first $p - 1$

derivatives evaluated at t ; that is, $\mathbf{x}(t | n) = [f(t), f'(t), \dots, f^{(p-1)}(t)]^T$. The assumption made is that λ is given. The smoothing parameter is usually chosen by generalised cross validation or maximum likelihood, for more details refer to Osborne and Prvan [5], [6] and references contained therein. In Osborne and Prvan [5] it was shown that the effect of the diffuse prior disappears after p steps of the Kalman Filter.

For reference the Kalman filter for the state space formulation (5) and (7) is

$$\begin{aligned} \mathbf{x}_{i|i-1} &= T(t_i, t_{i-1})\mathbf{x}_{i-1|i-1}, \\ S_{i|i-1} &= T(t_i, t_{i-1})S_{i-1|i-1}T(t_i, t_{i-1})^T + \lambda\Omega(t_i, t_{i-1}), \\ d_i &= \mathbf{e}_1 S_{k|k-1} \mathbf{e}_1^T + \sigma^2, \\ \mathbf{x}_{i|i} &= \mathbf{x}_{i|i-1} + S_{i|i-1} \mathbf{e}_1 d_i^{-1} (\mathbf{y}_i - \mathbf{e}_1^T \mathbf{x}_{i|i-1}), \\ S_{i|i} &= S_{i|i-1} - S_{i|i-1} \mathbf{e}_1 d_i^{-1} \mathbf{e}_1^T S_{i|i-1} \\ & \quad i = 2, \dots, n. \end{aligned}$$

The fixed-interval, discrete-time smoother is

$$\mathbf{x}_{j|n} = \mathbf{x}_{j|j} + A_j(\mathbf{x}_{j+1|n} - \mathbf{x}_{j+1|j}),$$

with

$$A_j = S_{j|j} T(t_{j+1}, t_j)^T$$

and

$$S_{j|n} = S_{j|j} + A_j(S_{j+1|n} - S_{j+1|j})A_j^T$$

for $j = n, \dots, 1$. The quantities $\mathbf{x}_{n|n}$ and $S_{n|n}$ obtained from the forward pass of the Kalman filter initiate this backward recursion.

The interpolation smoother for $t_{i-1} \leq t < t_i$ is

$$\mathbf{x}(t|n) = T(t, t_{i-1})\mathbf{x}_{i-1|i-1} + A(t_i, t)(\mathbf{x}_{i|n} - \mathbf{x}_{i|i-1})$$

where

$$A(t_i, t) = \{T(t, t_{i-1})S_{i-1|i-1}T(t_i, t)^T + \Gamma(t_i, t)\}S_{i|i-1}^{-1}$$

and

$$\Gamma(t_i, t) = \Omega(t, t_i)T(t_i, t)^T$$

with

$$\begin{aligned} S(t|n) &= \Omega(t, t_{i-1}) + T(t, t_{i-1})S_{i-1|i-1}T(t, t_{i-1})^T \\ &\quad - A(t_i, t)(S_{i|i-1} - S_{i|n})A(t_i, t)^T. \end{aligned}$$

It is worth noting that the smoothed covariance is an end product in itself and does not enter the recursion for the smoothed state vectors.

In practice, square root formulations would be used in implementing these recursions.

3 Method

3.1 Constructing the smoothing spline basis

Let the knot points ζ_1, \dots, ζ_n be fixed. A basis of smoothing splines is constructed by generating n pseudo data sets each of n elements $(\zeta_1, z_1^i), \dots, (\zeta_n, z_n^i)$. The i -th set is denoted by (ζ, \mathbf{z}^i) and \mathbf{z}^i , $i = 1, \dots, n$, are linearly independent. For fixed λ an obvious choice is obtained by setting $\mathbf{z}^i = \mathbf{e}_i$ for $i = 1, \dots, n$ where $\mathbf{e}_i \in \mathfrak{R}^n$ here. The i -th smoothing spline basis curve is obtained by using (ζ, \mathbf{z}^i) as the data and then following the approach in Osborne and Prvan [5], [6] to fit the smoothing spline to this pseudo data set. The same smoothing parameter λ , constrained to be positive, will be used for each pseudo data set.

Significant savings in computation may be made by realizing that the covariance matrices obtained from the Kalman filter, fixed-interval, discrete-time smoother and interpolation smoother are independent of \mathbf{z}^i and depend only on ζ , so the covariance matrices need only be found for the first set (ζ, \mathbf{z}^1) . Subsequent applications of the filters or smoothers only require that the steps for the state vectors be executed.

Let $\phi_j(t; \lambda, \zeta)$ denote the j -th smoothing spline basis function, which depends on the smoothing parameter λ and knots ζ . The approximating curve is

$$f(t) = \sum_{j=1}^n a_j \phi_j(t; \lambda, \zeta),$$

where $\phi_j(t; \lambda, \zeta) = \mathbf{e}_1^T \mathbf{x}^{(j)}(t | n)$. The superscript j denotes that the j -th

pseudo data set (ζ, z^j) has been used.

3.2 Least squares spline with fixed knots using a smoothing spline basis

For fixed knot points, we minimise

$$\tilde{S}(\zeta, \lambda, \mathbf{a}) = \sum_{i=1}^N \left(y_i - \sum_{j=1}^n a_j \phi_j(t_i; \lambda, \zeta) \right)^2 \quad (8)$$

over \mathbf{a} and λ . For fixed λ the problem is a linear least squares problem.

3.3 Least squares spline with variable knots using a smoothing spline basis

The least squares spline with variable knots using a smoothing spline basis is obtained by minimising $\tilde{S}(\zeta, \lambda, \mathbf{a})$ over \mathbf{a} , λ and the choice of knots ζ_1, \dots, ζ_n . It is reasonable that λ depends on the data. We have n linear parameters, a_1, \dots, a_n , and $n + 1$ non linear parameters, $\zeta_1, \dots, \zeta_n, \lambda$. Set $\zeta_1 = t_1$ and $\zeta_n = t_N$. To minimise the effect of the diffuse prior set $\zeta_{i+1} = t_1 + i\delta$ for $i = 1, \dots, p - 1$ where δ is chosen small enough so that the first p knots lie in $[t_1, t_2]$. We end up having $n - p - 1$ free knots. The stochastic formulation

requires $\zeta_1 < \zeta_2 < \dots < \zeta_n$ so we need to build this constraint into any constrained optimisation procedure used. Although it is possible to obtain the derivatives of the objective function with respect to the knots, this is usually very complicated.

Since λ needs to be fixed for smoothing spline calculations it makes sense to split the optimisation into two steps.

1. For fixed λ minimise $\tilde{S}(\boldsymbol{\zeta}; \lambda, \mathbf{a})$ over $a_1, \dots, a_n, \zeta_{p+1}, \dots, \zeta_{n-1}$ which is accomplished in two stages because the problem is a linear least squares problem when the knots are fixed.
2. For fixed $a_1, \dots, a_n, \zeta_{p+1}, \dots, \zeta_{n-1}$ we minimise over λ .

We continue the two step process until a minimum has been reached. For the simulation study considered in Section 4 convergence was not slow.

3.4 Extensions

The approach can be extended to fitting a curve which has variable continuity properties and degree. This is achieved by replacing $\mathbf{e}_p \mathbf{e}_p^T$ in (6) by a positive semi definite matrix V_i . The choice of the V_i determines the degree of the polynomial in the $(i-1)$ -th interval. If V_i is different from V_{i-1} the number of continuous derivatives at the $(i-1)$ -th knot point is decreased. More details

about varying the continuity and degree of the curve fitted can be found in Prvan [7].

If two knots appear to be coalescing we could replace them with one knot point and reduce the continuity across that knot point. This approach is currently being investigated as a means of determining the number of knots to use.

As a by-product of the stochastic formulation we estimate definite integrals by replacing \mathbf{e}_1^T in the observation equation by \mathbf{e}_2^T . The estimate of $\int_{t_1}^t f(t)dt$ is $\sum_{i=1}^n a_i \mathbf{e}_1^T \mathbf{x}^{(j)}(t|n)$ and the estimate of $f(t)$ is $\sum_{i=1}^n a_i \mathbf{e}_2^T \mathbf{x}^{(j)}(t|n)$.

3.5 Remarks

The stochastic setting enables approximate pointwise confidence intervals to be attached to the curve. The approximate $(1 - \alpha)100\%$ confidence interval (C.I.) is

$$f(t) \pm \hat{\sigma} U_{\frac{\alpha}{2}} \sqrt{\sum_{j=1}^n a_j^2 \text{var}(\phi_j(t; \lambda, \zeta))}$$

where $\hat{\sigma}^2 = \tilde{S}(\zeta, \lambda, \mathbf{a})/(n - p)$, $\text{var}(\phi_j(t; \lambda, \zeta)) = \mathbf{e}_1^T S^{(j)}(t|n) \mathbf{e}_1$ and $U_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ percentage point of the standard normal distribution. The superscript j denotes the j -th pseudo data set has been used to obtain the smoothed state covariance matrix $S^{(j)}(t|n)$. As noted earlier, $S^{(j)}(t|n)$, $j = 1, \dots, n$, are identical because they depend on ζ not \mathbf{z}^i . This smoothed state

covariance is obtained from the interpolation smoother after performing a forward pass of the Kalman filter and a backward pass of the fixed-interval, discrete-time smoother. For discussions on estimating the error for smoothing splines more accurately refer to Van der Linde [11] and references contained therein.

Like Schwetlick and Schütze [8] we can find a good location of variable knots in a subinterval of $[t_1, t_N]$ with the knots fixed outside this subinterval. As mentioned in [8] this option is useful when fitting several data sets where one data set only differs from the others over a small subsection of $[t_1, t_N]$.

The approach above can be implemented without recourse to the stochastic setting. The $\phi_j(t; \lambda, \zeta)$ could be obtained by running any smoothing spline software using the pseudo data (ζ, z^j) for $j = 1, \dots, n$. This would remove the necessity of fixing the first p knots when fitting a smoothing spline of degree $2p - 1$. The continuity properties or degree of the piecewise polynomial being fitted cannot be varied using existing software. Developing the code is not straightforward. As mentioned earlier, the stochastic formulation has superior stability properties.

4 A simulation study

One hundred data sets consisting of 74 points are generated for each of three levels of noise ($\sigma = 0.1$, $\sigma = 0.5$ and $\sigma = 1$) where the underlying signal is

$$f(t) = \frac{x^3}{x^2 + 1}.$$

The 74 data points abscissae are uniformly spaced in the interval $[-3.8, 3.5]$. Each data set (300 in total) has a least squares spline with two variable knots using a cubic B-spline basis fitted to it and a least squares spline with variable knots using a cubic smoothing spline basis fitted to it.

Since each data set is used to fit a least squares spline with 2 variable knots using a cubic smoothing spline basis and then using a cubic B-spline basis we can perform one sample t-tests on the differences of the goodness of fit measures to see if there is significant differences between the two approaches. We will use $\frac{1}{n} \sum_{i=1}^n (f(t_i) - \hat{f}(t_i))^2$, which for simplicity we will call the residual sum of squares (note this is different from the statistical definition of residual sum of squares which is $\frac{1}{n} \sum_{i=1}^N (y_i - \hat{f}(t_i))^2$), and integrated squared error which is $\frac{1}{n} \int_{t_1}^{t_N} (f(t) - \hat{f}(t))^2 dt$, where \hat{f} is our fitted curve. A 95% confidence interval (C.I.) for the difference between these two measures of goodness of fit are obtained. Here the difference is smoothing spline approach value minus B-spline approach value for the measure of goodness of fit being looked at. Results are summarised in the table below. A p-value less than or equal to α (typically 0.05 or 0.01) indicates there is sufficient evidence to suggest a

significant difference with respect to the goodness of fit measure looked at. The confidence interval or mean of the differences tells you in which direction the difference is.

TABLE 1: Results for True Sum of Squared Errors.

σ	p-value	mean of differences	95% C.I. for differences
0.1	0.00	0.02691	(0.01991, 0.03391)
0.5	1.6×10^{-2}	-0.06796	(-0.10948, -0.02643)
1	4.0×10^{-4}	-0.1738	(-0.2689, -0.07884)

TABLE 2: Results for Integrated Squared Errors.

σ	p-value	mean of differences	95% C.I. for differences
0.1	0.00	0.001907	(0.001864, 0.001950)
0.5	1.2×10^{-3}	-0.002627	(-0.004194, -0.001059)
1	0.00	-0.01600	(-0.0.02241, -0.009590)

Boxplots (Figure 1) give the pattern of variability in the variables being looked at in an easy to compare graphical summary. The white horizontal line lying inside the box gives the median value (point at which 50% of the observations are smaller and 50% of the observations are larger), the height of the box gives the interquartile range (IQR) which is the range of the middle 50% of the values, and the whiskers extend to whichever is closer: the most extreme observation or $1.5 \times \text{IQR}$ away from the box.

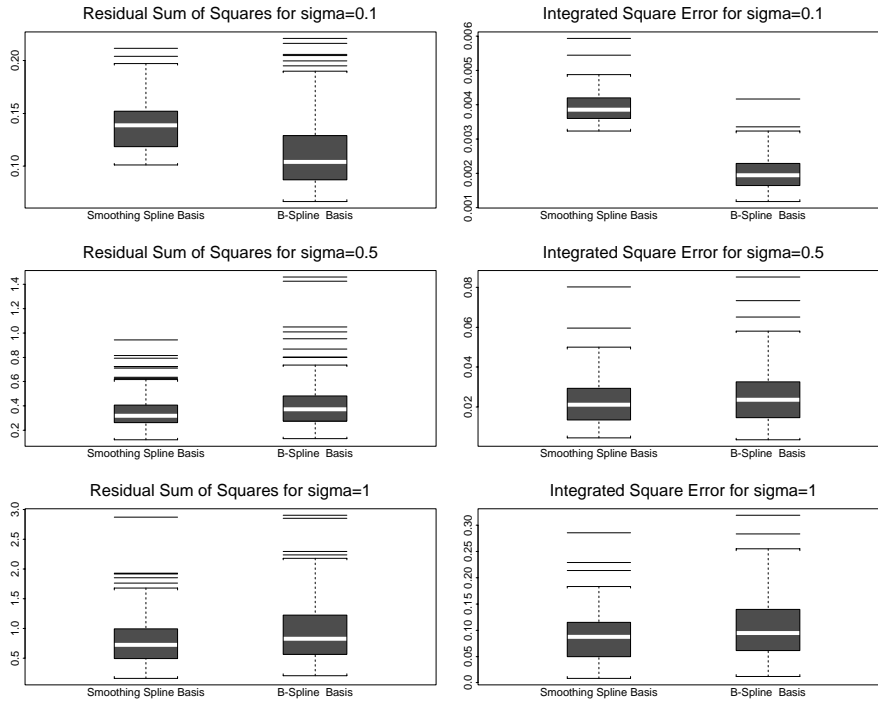


FIGURE 1: Simulation Study Results

As can be seen from the boxplots in Figure 1 and t-tests for differences summarised in Table 1 and Table 2 for negligible noise ($\sigma = 0.1$) the least squares spline using a B-spline method on average is superior in terms of the two goodness of fit measures used whereas when the noise is noticeable ($\sigma = 0.5$ and $\sigma = 1$) the converse is true.

5 Conclusion

Least squares splines with variable knots using a smoothing spline basis appear to perform better when the data is noisy. Least squares splines with variable knots using a B-spline basis appear to perform better when the noise is negligible.

References

- [1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, 1979.
- [2] P. Billingsley. *Probability and Measure*. Wiley, 1979. C1203

- [3] IMSL. *User's manual : MATH/LIBRARY : FORTRAN subroutines for mathematical applications, Version 2.0*. IMSL, Houston, 1991. [C1201](#)
- [4] D.L. Jupp. Approximation to data by splines with free knots. *SIAM J. Numer. Anal.*, 15:328–343, 1978. [C1201](#)
- [5] M.R. Osborne and T. Prvan. On algorithms for generalised smoothing splines. *J. Austral. Math. Soc. Ser. B*, 29:322–241, 1988. [C1202](#), [C1205](#), [C1205](#), [C1207](#)
- [6] M.R. Osborne and T. Prvan. Smoothness and conditioning in generalised smoothing spline calculations. *J. Austral. Math. Soc. Ser. B*, 30:43–56, 1988. [C1205](#), [C1207](#)
- [7] T. Prvan. Smoothing splines with variable continuity properties and degree. *Appl. Math. Lett.*, 10:75–80, 1997. [C1202](#), [C1210](#)
- [8] H. Schwetlick and T. Schütze. Least squares approximation by splines with free knotd. *B.I.T.*, 35:361–384, 1995. [C1202](#), [C1211](#), [C1211](#)
- [9] G. Wahba. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Statist. Assoc.*, 40:364–372, 1978. [C1203](#)
- [10] W. Wecker and C.F. Ansley. Signal extraction approach to non linear regression and spline smoothing. *J. Amer. Statist. Assoc.*, 78:81–89, 1983. [C1203](#), [C1204](#)

- [11] A. Van der Linde. Computing the error for smoothing splines. *Comp. Stat.*, 10:143–154, 1995. **C1211**