SIAM J. SCI. COMPUT.
Vol. 15, No. 5, pp. 1126–1133, September 1994
© 1994 Society for Industrial and Applied Mathematics
009

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

MONOTONIC SMOOTHING SPLINES FITTED BY CROSS VALIDATION*

S. N. WOOD†

Abstract. A practical method for calculating monotonic cubic smoothing splines is given. Linear sufficient conditions for monotonicity are employed, and the spline coefficients are obtained using quadratic programming. The method enables efficient cross-validation estimates of the smoothing parameter to be made and confidence intervals to be calculated for the resulting spline. The results are easy to extend to histogram data.

Key words. smoothing spline, spline, monotonic, cross validation

AMS subject classifications. 65D07, 65D10

1. Introduction. There are a number of practical spline based methods for monotonic interpolation [1], [3], [5], [6], [9], [10], [12]–[14], [17], [18], [22], [30], but monotonic smoothing splines have received less attention. Unconstrained smoothing splines are useful tools for data analysis principally because of the cross-validation methods pioneered in [4] for choosing the amount of smoothing appropriate for any dataset and the further work in [27] on estimating confidence intervals associated with the resulting spline even when the variance of the original data is unknown. Similar results are not available for monotonic smoothing splines [24]. A cross-validation method for linearly constrained thin-plate splines has been developed in [25], but monotonicity or confidence intervals were not considered. Subsequent work [20], [15] on monotonically constrained cubic smoothing splines requires the level of smoothing to be chosen by eye, whilst [7] is concerned with efficient algorithms for more restrictive constraints involving convexity and does not address the choice of smoothing parameter. Confidence intervals are not considered in [15] or [7]. Ramsay [20] plots confidence intervals on one figure, but does not discuss their calculation or properties.

The practical methods for calculating monotonic smoothing splines proposed in [15] and [20] involve constraints on the parameters of B-splines, which yield sufficient (but not necessary) conditions for monotonicity. In this note I instead use the piecewise polynomial representation of a spline and some sufficient conditions for monotonicity first used in [14] for monotonic interpolation with piecewise cubics. This yields a quadratic programming problem with linear constraints, and by considering some properties of the active set method for positive-definite quadratic programming given in [11] it is possible to extend the cross-validation and confidence interval methods of [4] and [27] to monotonic cubic smoothing splines. A demonstration of the technique applied to artificial data is given at the end of this article.

2. The method. The set of n "knots" $\{(x_i, a_i)\}$ such that $x_i < x_{i+1}$ for $1 \le i \le n-1$ can be interpolated by a curve $s(x)$ defined by

$$s(x) = s_i(x) = a_i\phi_{0i}(x) + a_{i+1}\phi_{1i} + c_i\gamma_{0i}(x) + c_{i+1}\gamma_{1i}(x) \quad \text{for } x_i \le x \le x_{i+1},$$

where the basis functions ϕ and γ are defined in Table 1; $c_i = s''(x_i)$, $a_i = s(x_i)$, and $s_i(x)$ is a section of cubic polynomial. The c_i's can be obtained from the a_i's by requiring continuity of $s(x)$, $s'(x)$, and $s''(x)$ at each x_i, and setting $c_1 = c_n = 0$, this yields the matrix system

$$\mathbf{Dc} = \mathbf{Ha},$$

where $\mathbf{D}$ and $\mathbf{H}$ are defined in Table 1, $\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$ and $\mathbf{c} = (c_2, c_3, \ldots, c_{n-1})^T$. The interpolant calculated in this manner is a cubic spline.

*Received by the editors August 14, 1991; accepted for publication (in revised form) August 2, 1993.

†Natural Environment Research Council (NERC) Centre for Population Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, United Kingdom (snw@ic.ac.uk).

TABLE 1
Definitions for a cubic spline $h_j = x_{j+1} - x_j$.

Definitions of **H** and **D**. All elements are 0 except the following.		
$H_{j,j} = 1/h_j$ $D_{j,j} = (h_j + h_{j+1})/3$	$H_{j,j+1} = -(1/h_j + 1/h_{j+1})$	$H_{j,j+2} = 1/h_{j+1}$ $1 \le j \le n-2$
$D_{j,j+1} = h_{j+1}/6$	$D_{j+1,j} = h_{j+1}/6$	$1 \le j \le n-3$
Definitions of **P** and **U**. All elements are 0 except the following.		
$P_{j,j} = 2$		$1 \le j \le n$
$P_{j,j-1} = h_j/(h_j + h_{j+1})$ $U_{j,j+1} = 3P_{j,j+1}/h_j$	$P_{j,j+1} = 1 - P_{j,j-1}$ $U_{j,j} = -(U_{j,j+1} + U_{j,j-1})$	$U_{j,j-1} = -3P_{j,j-1}/h_{j-1}$ $2 \le j \le n-1$
$P_{1,2} = 1$ $P_{n,n-1} = 1$	$U_{1,1} = -3/h_1$ $U_{n,n-1} = -3/h_{n-1}$	$U_{1,2} = -U_{1,1}$ $U_{n,n} = -U_{n,n-1}$
Basis functions for a cubic spline.		
$\phi_{0j}(x) = (x_{j+1} - x)/h_j$ $\phi_{1j} = (x - x_j)/h_j$	$\gamma_{0j}(x) = \{\phi_{0j}(x)^3 - \phi_{0j}(x)\}h_j^2/6$ $\gamma_{1j}(x) = \{\phi_{1j}(x)^3 - \phi_{1j}(x)\}h_j^2/6$	

If $\mathbf{y}$ is a vector of observations, which we wish to approximate with the spline $s(x)$, then we seek $\mathbf{a}$ the vector of the values of the spline at the knots such that

$$\int_{x_1}^{x_n} [s''(x)]^2 dx + \frac{\lambda}{n} \|\mathbf{W}(\mathbf{a} - \mathbf{y})\|^2 \tag{1}$$

is minimised. λ is the "smoothing parameter," $\mathbf{W}$ is a diagonal matrix of observation weights, and $\|\cdot\|$ is the standard Euclidian norm. The resulting curve is a cubic smoothing spline. Extensive discussion of spline theory is given in [2], [8], [23], [28], and at a more introductory level in [16]. It is easy to show that (1) is equivalent to minimising

$$\frac{1}{2}\mathbf{a}^{\mathrm{T}}\mathbf{G}_\lambda \mathbf{a} + \mathbf{c}_\lambda^{\mathrm{T}}\mathbf{a}, \tag{2}$$

where $\mathbf{c}_\lambda^{\mathrm{T}} = -2\frac{\lambda}{n}\mathbf{y}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}$ and $\mathbf{G}_\lambda = 2[\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{H} + \frac{\lambda}{n}\mathbf{W}^{\mathrm{T}}\mathbf{W}]$.$\mathbf{G}_\lambda$ is clearly nonnegative definite.

The minimisation of (2) will not, in general, give rise to an $s(x)$, which is monotonic, but conditions for monotonicity of a cubic over an interval $[x_i, x_{i+1}]$ are well known and can be expressed in terms of β and δ where $\beta = s'(x_{i+1})/\Delta_i$, $\delta = s'(x_i)/\Delta_i$, and $\Delta_i = (a_{i+1} - a_i)/(x_{i+1} - x_i)$. The necessary and sufficient conditions for monotonicity are that δ and β should lie within the region of Fig. 1, defined by the union of A, B, and C. These nonlinear constraints are used in several shape-preserving interpolation schemes, notably [30], [9], and [1]. Minimisation of (2) subject to these constraints requires quadratic programming with nonlinear constraints. This is not an attractive proposition (see [11, Chap. 6]) and instead I propose to follow [14] in using the linear sufficient conditions for monotonicity $0 \le \beta \le 3$ and $0 \le \delta \le 3$; that is β and δ lie within region A of Fig. 1. If $\mathbf{b} = (s'(x_1), s'(x_2), \ldots, s'(x_n))^{\mathrm{T}}$ then, for a cubic spline, $\mathbf{b}$ can be written as a linear transformation of $\mathbf{a}$: $\mathbf{b} = \mathbf{B}\mathbf{a}$, where $\mathbf{B} = \mathbf{P}^{-1}\mathbf{U}$ and $\mathbf{P}$ and $\mathbf{U}$ are defined in Table 1. Writing $\mathbf{B}_i$ for the ith row of $\mathbf{B}$, it is now possible to write the sufficient conditions for monotonicity over $[x_i, x_{i+1}]$ in terms of $\mathbf{a}$:

$$3(a_{i+1} - a_i)/h_i - \mathbf{B}_i\mathbf{a} \ge 0, \quad 3(a_{i+1} - a_i)/h_i - \mathbf{B}_{i+1}\mathbf{a} \ge 0,$$

$$\mathbf{B}_i\mathbf{a} \ge 0, \quad \mathbf{B}_{i+1}\mathbf{a} \ge 0 \quad \text{and} \quad a_{i+1} - a_i \ge 0$$

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

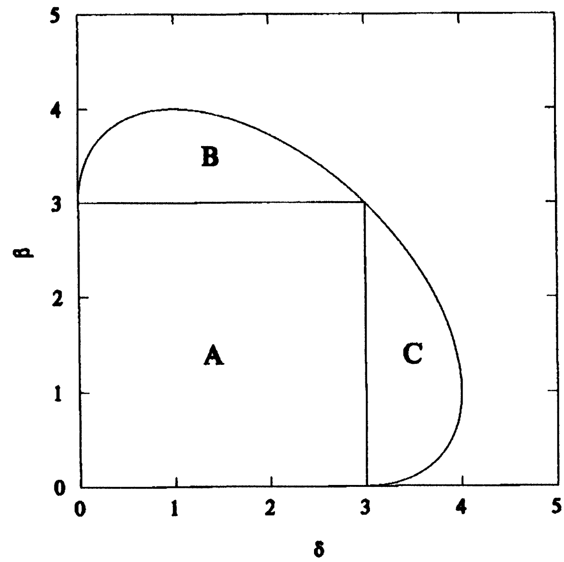

FIG. 1. *The monotonicity region for a section of a cubic spline.* A *is the subregion used for this work.* β *and* δ *are defined in* §2.

for $s(x)$ nondecreasing and

$$\mathbf{B}_i\mathbf{a} - 3(a_{i+1} - a_i)/h_i \geq 0, \quad \mathbf{B}_{i+1}\mathbf{a} - 3(a_{i+1} - a_i)/h_i \geq 0,$$

$$\mathbf{B}_i\mathbf{a} \leq 0, \quad \mathbf{B}_{i+1}\mathbf{a} \leq 0, \quad \text{and} \quad a_i - a_{i+1} \geq 0$$

for $s(x)$ nonincreasing, where $h_i = x_{i+1} - x_i$. Clearly, the sufficient conditions for monotonicity of $s(x)$ over $[x_1, x_n]$ can be written as

$$\mathbf{Ca} \geq [\mathbf{0}], \tag{3}$$

where $[\mathbf{0}]$ is a vector of zeros. Additional absolute constraints on any a_i are trivial to add to this constraint set. It is fairly easy to show that these constraints are equivalent to those used in [15] (which are apparently similar to those used in [20]). Note also that any straight line sloping up or down, as appropriate, will satisfy the monotonicity constraints.

Minimisation of (2) subject to (3) is a positive definite quadratic programming problem, which will be solved using the algorithm given in [11]. To extend cross-validation methods from unconstrainted to constrained cubic splines some features of this algorithm must be considered.

The method starts with a feasible vector $\mathbf{a}$, which violates none of the inequality constraints in (3) but which may be constrained to satisfy t of the constraints as equality constraints. These t exact constraints are termed an "Active set" and can be expressed $\mathbf{C}_A\mathbf{a} \geq [\mathbf{0}]$. The algorithm is straightforward.

The $t \times n$ matrix $\mathbf{C}_A$ is factorized $\mathbf{C}_A\mathbf{Q} = [\mathbf{0}, \mathbf{T}]$, where $\mathbf{T}$ is a $t \times t$ matrix such that $T_{i,j} = 0$ if $i + j < n$, $\mathbf{0}$ is an appropriate matrix of zeros, and $\mathbf{Q}$ is an $n \times n$ orthogonal matrix constructed from the product of the identity matrix and a series of Householder matrices applied to the right. The first $n - t$ columns of $\mathbf{Q}$ define a matrix $\mathbf{Z}$, the columns of which form a basis for the null space of $\mathbf{C}_A$, the space in which any movement is possible whilst still satisfying the active set as equality constraints. The remaining columns of $\mathbf{Q}$ form the basis

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$\mathbf{\Omega}$ of the space orthogonal to $\mathbf{Z}$: this is the "range space" of the active set. Given $\mathbf{Z}$ a search direction $\mathbf{p}$ can be found that is guaranteed not to violate the active constraints. The search direction $\mathbf{p}$ comes from the solution of $\mathbf{Z}^T\mathbf{G}_\lambda\mathbf{Z}\mathbf{p}_z = -\mathbf{Z}^T\mathbf{g}$, where $\mathbf{p} = \mathbf{Z}\mathbf{p}_z$ and $\mathbf{g} = \mathbf{G}_\lambda\mathbf{a}+\mathbf{c}_\lambda$. The vector $\mathbf{a}^* = \mathbf{a}+\mathbf{p}$ is the position of the minimum of the quadratic problem within the null space of $\mathbf{C}_A$, but if a step from $\mathbf{a}$ to $\mathbf{a}^*$ would violate one or more of the inactive constraints, then a shorter step must be taken to the constraint nearest $\mathbf{a}$ and that constraint must be included in an updated active set matrix $\mathbf{C}_A$. The algorithm then proceeds from the beginning with the new $\mathbf{C}_A$ and $\mathbf{a}$. If the solution has reached a constrained minimum $\mathbf{a}^*$, then it is necessary to find out whether all the active constraints are currently required. This is achieved by evaluating the Lagrange multipliers associated with each active constraint. If no constraints need to be deleted from $\mathbf{C}_A$, then the problem has been solved. Extensive computational details are given in [11]. What matters here are the spaces defined by $\mathbf{\Omega}$ and $\mathbf{Z}$, which will be used in the method of choosing λ.

Assume a data model $y_i = f(x_i) \pm \varepsilon_i$, where $f(x)$ is the underlying but unknown smooth function and the ε_i's are error terms such that $\mathbf{E}(\varepsilon_i) = 0$, $\mathbf{E}(\varepsilon_i\varepsilon_i) = w_i^{-2}\sigma^2$, and $\mathbf{E}(\varepsilon_i\varepsilon_i) = 0$ if $i \neq j$. Let $s_\lambda(x)$ be the monotonic smoothing spline fitted to $\mathbf{y}$ by solving (2) subject to (3). The spline $s_\lambda(x)$ is intended to be an approximation of $f(x)$. A reasonable measure of the success of this approximation is provided by

$$R(\lambda) = \sum_{i=1}^{n} w_i^2[f(x_i) - s_\lambda(x_i)]^2/n = \|\mathbf{W}(\mathbf{f} - \mathbf{a}^*)\|^2/n, \tag{4}$$

where $W_{i,i} = w_i$ and $W_{i,j} = 0$ if $i \neq j$. The term λ should be chosen to minimize the expected value of $R(\lambda)$. In the unconstrained case it has been suggested that λ should be chosen so that $\|\mathbf{a}^* - \mathbf{y}\|^2/n = \sigma^2$, [21], but in [26] it is shown that this leads to oversmoothing and a practical estimator of $R(\lambda)$ is derived. Reference [26] uses the fact that unconstrained smoothing with spline functions can be summarised by a linear equation of the form $\mathbf{a} = \mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is called the "influence matrix." In general, such an equation cannot be constructed for the constrained spline function, since the constraints also enter the solution. However, it is possible to estimate λ by minimisation of an estimator of $R(\lambda)$ for any particular active set $\mathbf{C}_A$. As mentioned above the null space of $\mathbf{C}_A$, $\mathbf{Z}$ has a complementary "range" space $\mathbf{\Omega}$ and an initial estimate $\mathbf{a}^0$ of the solution $\mathbf{a}^*$ of the spline fitting problem can therefore be written $\mathbf{a}^0 = \mathbf{\Omega}\mathbf{a}_\Omega^0 + \mathbf{Z}\mathbf{a}_Z^0$. The solution $\mathbf{a}^*$ is therefore $\mathbf{a}^* = \mathbf{K}\mathbf{a}_\Omega^0 + \tilde{\mathbf{A}}\mathbf{y}$, where $\mathbf{K} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}^T\mathbf{G}_\lambda\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{G}_\lambda)\mathbf{\Omega}$ and $\tilde{\mathbf{A}} = 2\frac{\lambda}{n}\mathbf{Z}(\mathbf{Z}^T\mathbf{G}_\lambda\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{W}^T\mathbf{W}$. The attractive feature of this representation is that $\mathbf{a}_\Omega^0$ is uniquely determined by the constraints in the active set, unlike $\mathbf{a}^0$. It is now easy to show that for a given $\mathbf{C}_A$, minimisation of the $\mathbf{E}\{R(\lambda)\}$ can be achieved by minimisation of

$$\tilde{R}(\lambda) = \|\mathbf{W}(\mathbf{a}^* - \mathbf{y})\|^2/n + 2\sigma^2\mathrm{Tr}(\tilde{\mathbf{A}})/n \tag{5}$$

(cf. [4, Eqn. (1.8)]). So one way of finding the value of λ is to search for the turning point of (5), which is consistent with the active set implied by such a turning point. This method is adopted in the work reported below.

In the unconstrained case, [4] shows that the value of λ minimising the expected value of (4) can be estimated without prior knowledge of σ using the process of generalized cross validation. If $s_\lambda^{[k]}(x)$ is the spline function calculated using every point in the data vector $\mathbf{y}$ except y_k, then

$$V(\lambda) = \sum_{k=1}^{n} u_k(s_\lambda^{[k]}(x) - y_k)^2/n$$

is taken as a measure of the badness of λ (u_k is a weight compensating for uneven mesh, etc.). So λ is chosen by minimisation of $V(\lambda)$. Once again the presence of the constraints in the solution of the monotonic spline problem precludes direct use of the results in [4], but by using intermediate results from the paper it is easy to show that for a given active set minimisation of:

$$\tilde{V}(\lambda) = n\|\mathbf{W}(\mathbf{a}^* - \mathbf{y})\|^2/[\mathrm{Tr}(\mathbf{I} - \tilde{\mathbf{A}})]^2 \tag{6}$$

is equivalent to minimisation of $V(\lambda)$ with respect to λ. Equation (6) can be used in the same way as (5) to obtain an estimate of λ. $\tilde{\mathbf{A}}$ has the same role in choosing the smoothing parameter for the monotonic spline that $\mathbf{A}$ has for unconstrained spline smoothing. Minimisation of (6) should be compared with the method for constrained multivariate thin-plate splines given in [25].

$\tilde{\mathbf{A}}$ can be used to extend the results on confidence intervals for unconstrained spline smoothing given in [27] to the constrained case. For the spline fitted by cross validation, an estimate of σ^2, $\sigma_v^2 = \|\mathbf{W}(\mathbf{a}^* - \mathbf{y})\|^2/\mathrm{Tr}(\mathbf{I} - \tilde{\mathbf{A}})$, can be obtained. Ninety-five percent "confidence intervals," $p_v = 2\sigma_v\sqrt{\mathrm{Tr}(\tilde{\mathbf{A}}[\mathbf{W}^{\mathrm{T}}\mathbf{W}]^{-1})/n}$ or $p_R = 2\sigma\sqrt{\mathrm{Tr}(\tilde{\mathbf{A}}[\mathbf{W}^{\mathrm{T}}\mathbf{W}]^{-1})/n}$, can also be calculated depending on whether or not σ is known a priori. Finally note that all the results given above can easily be applied to area approximating splines as used in [29], for example.

3. Some examples.

The three monotonic relationships, $f_i(x)$, shown in Fig. 2 were "sampled" at 45 equally spaced points with three levels of additive Guassian noise: standard deviation 0.05, 0.1, and 0.2. The noise was generated using methods given in [19]. Fifty replicate sets of data were generated for each relationship at each noise level. Three splines $s_\lambda^V(x)$, $s_\lambda^R(x)$, and $s_\lambda^t(x)$, with smoothing parameters λ_V, λ_R, and λ_t, were fitted to each set of data by minimisation of $\tilde{V}(\lambda)$, $\tilde{R}(\lambda)$, and $\|\mathbf{a}^* - \mathbf{f}\|^2/n$, respectively. The spline $s_\lambda^t(x)$ is closest to the function from which the data was sampled and enables the efficacy of $s_\lambda^V(x)$ and $s_\lambda^R(x)$ to be examined. For the spline fitted by cross validation, σ_V was calculated and 95% "confidence intervals," were produced for both $s_\lambda^V(x)$ and $s_\lambda^R(x)$. In all cases $\mathbf{W}$ was set to the identity matrix.

Several test statistics were calculated for each replicate as follows.

(i) XE is the proportion extra mean square deviation from $f_i(x)$ of $s_\lambda^V(x)$ and $s_\lambda^R(x)$ in comparison to the mean square deviation of $s_\lambda^t(x)$ from $f_i(x)$: this is intended to see how well the spline recaptures the true underlying function, in comparison with the best achievable.

(ii) RM measures the closeness of the estimated smoothing parameter to the one that would have led to the most accurate reconstruction of the underlying function. RM is the ratio of $\log(\lambda_V)$ or $\log(\lambda_R)$ to $\log(\lambda_t)$.

(iii) CIC, the percentage of $f_i(x)$ within the region the region defined by $s_\lambda^{V/R}(x) \pm p_{V/R}$, tests the effectiveness of the confidence intervals.

Means and standard deviations for these quantities are given in Table 2 for each treatment. Table 2 also records the mean number of constraints active at the solution K and the mean estimated value of σ, σ_V. Figure 2 shows three randomly selected examples from the set of simulations.

The estimates of σ, σ_V, are excellent, although the confidence intervals obtained from them are slightly too narrow when calculated from the minimisation of (6) and are slightly too wide when calculated by minimisation of (5). The deterioration in the ability of the splines $s_\lambda^V(x)$ and $s_\lambda^R(x)$ to recapture the underlying relationship $f(x)$ as the level of noise increases is unsurprising. If σ is large enough, then minimisation of (5) or (6) is bound to produce a straight line whatever the form of $f(x)$. This tendency for the method to produce a smoother relationship as the noise level becomes very large also explains the systematic tendency to underestimate λ and, therefore, k relative to $s_\lambda^t(x)$.

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

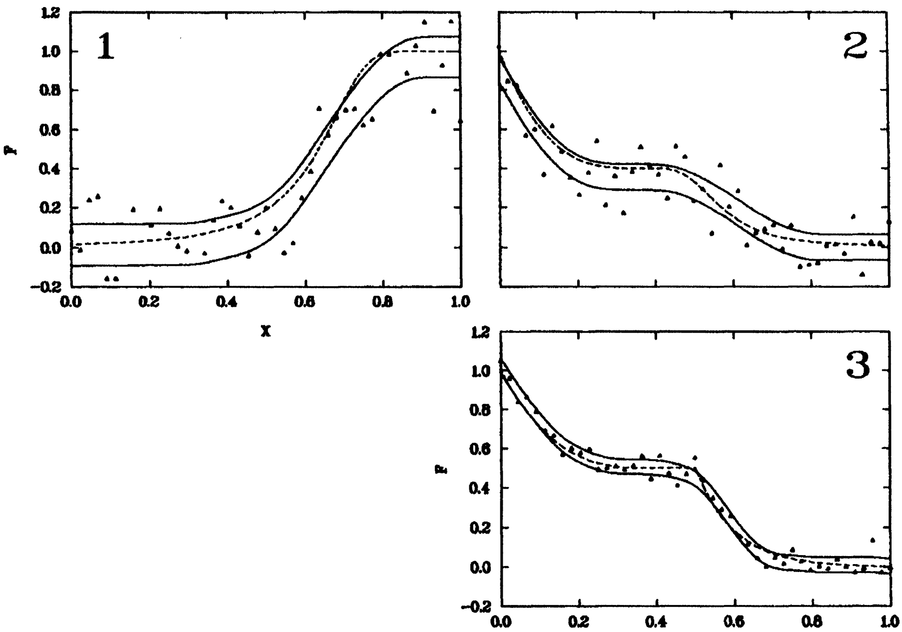

FIG. 2. *The dotted lines show the functions $f_i(x)$ used to test the spline fitting algorithms, as detailed in* §3. *The numbers in the corner of each graph are the function numbers of Table* 2. *The continuous lines show* 95% *confidence regions for a spline fitted to the data shown by cross validation.* σ =0.2, 0.1, *and* 0.05 *for graphs* 1, 2, *and* 3, *respectively.*

TABLE 2

A summary of the results of §3. *The figures are means over* 50 *replicates of the statistics. Figures in brackets,* (), *are standard deviations. The figures in square brackets,* [], *are the mean numbers of constraints active for* s_λ^t. *The notation* $\pm$ *is used to prefix the standard error of the mean. The function,* f, *numbers refer to Fig.* 2. *Columns* $\tilde{R}$ *are for splines fitted by minimisation of* (5) *and* $\tilde{V}$ *for splines fitted by minimisation of* (6).

		Noise level σ					
f	Stat	0.05		0.1		0.2	
		$\tilde{V}$	$\tilde{R}$	$\tilde{V}$	$\tilde{R}$	$\tilde{V}$	$\tilde{R}$
1	XE	0.09(.10)	0.10(.11)	0.11(.14)	0.12(.16)	0.15(.15)	0.15(.14)
	RM	0.97(.09)	0.97(.09)	0.94(.09)	0.94(.09)	0.86(.11)	0.86(.12)
	CIC	92(7)	98(2)	94(8)	98(3)	90(11)	94(6)
	K	8	8 [9]	8	8 [11]	8	8 [12]
	σ_V	0.047 ± 0.001		0.097 ± 0.002		0.0197 ± 0.003	
2	XE	0.08(0.09)	0.09(.09)	0.15(.15)	0.12(.13)	0.31(.42)	0.29(.35)
	RM	0.95(0.08)	0.94(.07)	0.92(.11)	0.92(.10)	0.76(.61)	0.76(.61)
	CIC	92(7)	98(2)	92(8)	97(2)	82(18)	93(7)
	K	6	5 [8]	6	6 [8]	4	4 [8]
	σ_V	0.047 ± 0.001		0.101 ± 0.001		0.201 ± 0.003	
3	XE	0.09(.12)	0.10(.12)	0.12(.15)	0.12(.13)	0.18(.25)	0.15(.21)
	RM	0.93(.10)	0.92(.10)	0.92(.12)	0.92(.12)	0.74(.48)	0.83(.33)
	CIC	91(6)	99(1)	91(8)	98(3)	77(22)	89(13)
	K	8	8 [11]	7	7 [10]	5	5 [9]
	σ_V	0.049 ± 0.004		0.100 ± 0.001		0.200 ± 0.004	

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

4. Discussion. Unconstrained smoothing splines are useful because reliable methods exist for choosing the smoothing parameter λ. The approach described in the previous sections of this note extends this power to monotonic smoothing splines.

One possible criticism of the method presented here (or of the other published methods) is that monotonicity is imposed by unnecessarily restrictive sufficient conditions. There are two replies to this. First, Fig. 1 shows that the linear conditions cover a large proportion of the necessary parameter space for monotonicity, but this is a weak argument on its own. More persuasive is a point first made in [2, p. 248]: the smoothing spline as presented here has far more knots than are actually required to represent the noisy data to which it is fitted. Put another way, the curve $s_\lambda(x)$ is potentially far more flexible than it needs to be. This implies that it is unlikely that there is any significant advantage gained by using the full nonlinear monotonicity constraints, which will simply allow $s_\lambda(x)$ slightly more flexibility. This argument becomes increasingly dubious as errors in the data become very small (not a problem with which the author has much experience). Similarly, it is less convincing as the number of knots in a least squares spline decreases relative to the number of datapoints.

Two further practical problems are associated with monotonic smoothing splines. First, the method described here will be slow if applied to large datasets. For noisy data this could be overcome using splines with fewer knots than there are datapoints: least-squares splines. The remaining drawback is the tedious computer programming required to implement the methods, especially given that standard quadratic programming packages do not give easy access to the matrices $\mathbf{Z}$ and $\mathbf{\Omega}$. This difficulty can be overcome by writing to the author for a "C" computer program implementing the method described in this note.

Acknowledgement. Thanks to Hans Metz for several helpful discussions.

REFERENCES

[1] R. K. Beatson and H. Wolkowicz, *Post-processing piecewise cubics for monotonicity*, SIAM J. Numer. Anal., 26 (1989), pp. 480–502.

[2] C. DeBoor, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.

[3] P. Costantini, *Co-monotone interpolating splines of arbitrary degree—a local approach*, SIAM J. Sci. Statist. Comput., 8 (1987), pp. 1026–1034.

[4] P. Craven and G. Wahba, *Smoothing noisy data with spline functions*, Numer. Math., 31 (1979), pp. 377–403.

[5] R. Delbourgo and J. A. Gregory, *C^2 quadratic spline interpolation to monotonic data*, IMA J. Numer. Anal., 3 (1983), pp. 141–152.

[6] S. Dietze and J. W. Schmidt, *Determination of shape preserving spline interpolants with minimal curvature via dual programs*, J. Approx. Theory, 52 (1988), pp. 43–57.

[7] T. Elfving and L.-E. Andersson, *An algorithm for computing constrained smoothing spline functions*, Numer. Math., 52 (1988), pp. 583–595.

[8] R. L. Eubank, *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, 1988.

[9] F. N. Fritsch and R. E. Carlson, *A method for constructing local monotone piecewise cubic interpolants*, SIAM J. Numer. Anal., 17 (1980), pp. 238–246.

[10] F. N. Fritsch and J. Butland, *A method for constructing local monotone piecewise cubic interpolants*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 300–304.

[11] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, 1981.

[12] J. A. Gregory, *Shape preserving rational spline interpolation*, in Rational Approximation and Interpolation, Lecture Notes in Mathematics 1105, Springer-Verlag, New York, 1984, pp. 431–441.

[13] J. A. Gregory and R. Delbourgo, *Piecewise rational quadratic interpolation to monotonic data*, IMA J. Numer. Anal., 2 (1982), pp. 123–130.

[14] J. M. Hyman, *Accurate monotonicity preserving cubic interpolation*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 645–654.

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

[16] P. Lancaster and K. Salkauskas, *Curve and Surface Fitting*, Academic Press, London, UK, 1986.

[17] D. F. McAllister and J. A. Roulier, *An algorithm for computing a shape-preserving osculatory quadratic spline*, ACM Trans. Math. Software, 7 (1981), pp. 331–347.

[18] R. MORANDI AND P. COSTANTINI, *Piecewise monotone quadratic histosplines*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 397–406.

[19] W. H. PRESS, B. P. FLANNERY, S. A. TEUKOLSKY, AND W. T. VETTERLING, *Numerical Recipes in C*, Cambridge University Press, Cambridge, UK, 1988.

[20] J. O. RAMSAY, *Monotone regression splines in action*, Statist. Sci., 3 (1988), pp. 425–461.

[21] C. H. REINSCH, *Smoothing by spline functions*, Numer. Math., 10 (1967), pp. 177–183.

[22] M. SAKAI AND M. C. LOPEZ DE SILANES, *A simple rational spline and its application to monotonic interpolation to monotonic data*, Numer. Math., 50 (1986), pp. 171–182.

[23] B. W. SILVERMAN, *Some aspects of the spline smoothing approach to non-parametric regression curve fitting*, J. Roy. Statist. Soc., B, 47 (1985), pp. 1–52.

[24] F. UTRERAS, *Smoothing noisy data under monotonicity constraints: existence, characterization and convergence rates*, Numer. Math., 47 (1985), pp. 611–625.

[25] M. VILLALOBOS AND G. WAHBA, *Inequality-constrained multivariate smoothing splines with application to the estimation of posterior probabilities*, J. Amer. Statist. Assoc., 82 (1987), pp. 239–248.

[26] G. WAHBA, *Smoothing noisy data with spline functions*, Numer. Math., 24 (1975), pp. 383–393.

[27] ———, *Bayesian "confidence intervals" for the cross-validated smoothing spline*, J. Roy. Statist. Soc. Sci., B, 45 (1983), pp. 133–150.

[28] ———, *Spline Methods for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.

[29] S. N. WOOD AND R. M. NISBET, *Estimation of Mortality Rates in Stage Structured Populations*, Lecture Notes in Biomathematics 90, Springer-Verlag, Berlin, 1991.

[30] Z. YAN, *Piecewise cubic curve fitting algorithm*, Math. Comp., 49 (1987), pp. 203–213.

Downloaded 05/06/20 to 131.130.169.5. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php