Alternatives to the Exponential Spline in Tension

By Steven Pruess

Abstract. A general setting is given for smooth interpolating splines depending on a parameter such that as this parameter approaches infinity the spline converges to the piecewise linear interpolant. The theory includes the standard exponential spline in tension, a rational spline, and several cubic splines. An algorithm is given for one of the cubics; the parameter for this example controls the spacing of new knots which are introduced.

1. Introduction. Cline [1] has written algorithms for the spline in tension which are useful in constructing convex and monotone approximations. The derivation and properties of this spline are given by Schweikert [13], Späth [15], [16], Hill [7] and Pruess [12]. Dube [3] has described local bases for such splines with varying continuity restrictions. With the exception of [16] these papers are concerned with the exponential (hyperbolic) spline in tension as given in Example 1 at the end of this section. In applications where the spline is to be evaluated many times, it is desirable to have an algorithm which avoids the three hyperbolic function evaluations each time the value of the spline is needed. Also, there are some numerical difficulties with the hyperbolic representation for small tension parameters; an alternate form is needed for this case [12]. Rational functions as suggested by Späth [16] provide one alternative, but recently de Boor [2] and Nielson [10] have exhibited cubic splines having similar properties to the spline in tension. In this paper a general setting is given which encompasses the above examples. This allows the results of [12] on the qualitative behavior of the exponential splines to be extended to this new class; in fact, the results can be sharpened since the assumption of quasi-uniform tension parameters is not needed here. An algorithm for one of the cubic splines is given which includes a discussion as to how the tension parameters can be automatically chosen. No theorem on convergence as the maximum mesh width goes to zero has been given, since this is usually not of interest for such splines. However, if one is desired, its proof would follow the pattern of that for Theorem 1 in [12].

Another approach to constructing monotone or convex approximations is through Bézier curves. Some representative papers are by Forrest [5], Gordon and Riesenfeld [6] and Wielinga [17]. More recent examples are found in Passow and Roulier [11] and McAllister, Passow and Roulier [8]. McAllister and Roulier [18] have also used these ideas to generate a C^1-quadratic spline with knots chosen automatically to match convexity or monotonicity. Finally, Nielson's v-splines [9] can also be used for parametric representations of a curve.

For the data (x_i, f_i), $1 \leqslant i \leqslant N$ with $x_1 < x_2 < \cdots < x_N$ we seek an interpolatory function $s(x)$ in $C^2[x_1, x_N]$. The class of splines considered here can be represented in terms of either first derivative or second derivative values at knots. In the first case, in $[x_i, x_{i+1}]$

$$s(x) = \{f_{i+1} - f_i - h_i s_i'\}G((x_{i+1} - x)/h_i)$$

(1.1)

$$- \{f_{i+1} - f_i - h_i s_{i+1}'\}G((x - x_i)/h_i) + \lambda_i(x),$$

for the second,

$$(1.2) \qquad s(x) = h_i^2 \{s_{i+1}''F((x - x_i)/h_i) + s_i''F((x_{i+1} - x)/h_i)\} + \lambda_i(x).$$

Here, $h_i = x_{i+1} - x_i$, $s_i' = s'(x_i)$, $s_i'' = s''(x_i)$, $\lambda_i(x) = f_{i+1}(x - x_i)/h_i + f_i(x_{i+1} - x)/h_i$. For the notation to be consistent, $F, G \in C^2[0, 1]$, $G(0) = G(1) = G'(0) = 0$, $G'(1) = 1$ and $F(0) = F(1) = F''(0) = 0, F''(1) = 1$. In most cases G and F are interval dependent and should have an i subscript (for simplicity, this subscript is omitted when no confusion is possible). I have always preferred representation (1.2) so the results of this paper are derived from it; however, one could equally well use (1.1) and prove analogous theorems.

Some possible choices for F are:

(1) $F_i(z) = (\sinh p_i h_i z/\sinh p_i h_i - z)/(p_i^2 h_i^2)$, the exponential spline in tension ($p_i > 0$),

(2) $F(z) = \{z^3/[1 + p(1 - z)] - z\}/(2p^2 + 6p + 6)$ a rational spline of Späth [16], ($p \geqslant 0$),

(3) $F(z) = [H(1 - z) + (1 + p^2)H(z)]/[2(1 + p^2)^2 - 2]$ with

$$H(z) = \frac{2}{3}z^3 - z^2 + \frac{1}{3}[pz - (p - 1)]_+^3,$$

a cubic spline of Nielson [10] ($p \geqslant 1$), and

(4) $F(z) = \{[pz - (p - 1)]_+^3 - z\}/(6p^2)$, a cubic spline of the author ($p \geqslant 1$). As usual $g_+ := \max(g, 0)$. The reason for the complexity of Nielson's example is that it is derived for representation (1.1) in which case

$$G(z) = \left\{\frac{2}{3}z^3 - z^2 + \frac{1}{3}[pz - (p - 1)]_+^3\right\}\Big/p.$$

De Boor [2] also has a piecewise cubic described in terms similar to (1.1). His example has an extra parameter to control "roundness" of the graph; also, it has tension parameters associated with data points rather than intervals so it does not fit into the scheme described herein. Späth [16], too, considers more general types of splines in tension but with no proofs of their behavior.

2. **Asymptotic Properties.** In order that the class of splines given by (1.2) have desirable mathematical and computational properties, various restrictions are imposed on the function $F(x)$:

(R-1) $F \in C^2[0, 1]$,

(R-2) $F(0) = F''(0) = F(1) = 0, F''(1) = 1,$

(R-3) $F'' \geqslant 0$ in $[0, 1]$,

(R-4) F depends on a parameter p in $[\alpha, \infty)$, α some fixed constant, in such a way that

(a) for $p = \alpha$ $F(z) = (z^3 - z)/6$,

(b) in any closed subinterval of $(0, 1)$ $F''(z) = o(1/p)$ uniformly as $p \to \infty$,

(c) $-F'(0)/F'(1)$ is a nonincreasing function of p.

(R-5) There exist positive constants C_1 and C_2, independent of p, but possibly dependent on $\{x_i\}$, such that for $p \geqslant \alpha$

(a) $0 < F'(1) \leqslant C_1/p$,

(b) $0 < -F'(0)/[F'(1)]^2 \leqslant C_2$,

(c) for p sufficiently large $C_1/(2p) < F'(1)$.

The necessity of the first two restrictions is discussed in the introduction. The third is desirable to get the proper convexity. (R-4(a)) provides a useful starting point for constructing tension spline approximations, viz., the standard cubic interpolatory spline. The more technical restrictions are not claimed to be necessary for the results of this paper to hold; however, they are sufficiently general to be satisfied by most of the examples listed in the introduction and the literature (cf. Section 3). In practice, it is useful to have a different value of p for each interval; thus F depends on i. It is then assumed that each F_i satisfies the above restrictions with α independent of i.

Note that (R-4(b)) indicates that s approaches the broken line interpolant as the parameters grow. Unfortunately, because of the continuity imposed on $s''(x)$, the values $\{s_i''\}$ also grow, so proofs of asymptotic behavior are not immediate.

Formula (1.2) and the first two restrictions insure that s interpolates the data and $s, s'' \in C[x_1, x_N]$. To get $s' \in C[x_1, x_N]$ set

$$(2.1) \quad d_i = h_i F_i'(1), \quad e_i = -h_i F_i'(0), \quad b_i = (h_i + h_{i-1}) \cdot [x_{i-1}, x_i, x_{i+1}]f,$$

where square brackets indicate the usual divided difference operator. Then, $s'(x_i^-) = s'(x_i^+)$, $1 < i < N$, implies

$$(2.2) \qquad\qquad e_{i-1}s_{i-1}'' + (d_{i-1} + d_i)s_i'' + e_i s_{i+1}'' = b_i.$$

To produce a square linear system in the unknown $\{s_i''\}$ two end conditions are imposed on $s(x)$. Some typical choices at $x = x_1$ (conditions at x_N are analogous) are: (a) $s'(x_1) = f'(x_1)$, (b) $s''(x_1) = 0$, (c) $s''(x_1) = f''(x_1)$, (d) $s''(x_1) = P''(x_1)$, where P is the cubic which interpolates f over x_1, x_2, x_3, x_4. These end conditions result in equations with structure similar to (2.2), e.g., at $x = x_1$ (a) says $d_1 s_1'' + e_1 s_2'' = \hat{b}_1$, (b)–(d) result in a reduced system of order $N - 2$ whose first equation is $(d_1 + d_2)s_2'' + e_2 s_3'' = \hat{b}_2$. The formulas for $\hat{b}$ vary with the type of end condition used.

The resulting system, written as $As'' = b$, is symmetric and tridiagonal. That A is also strictly diagonally dominant follows from a lemma which is also needed in developing the asymptotic behavior of $s(x)$. The proof follows directly from (2.1) and

(R-4), (R-5). The constants in the next two lemmas are nonzero and may depend on $\{x_i\}$.

LEMMA 1. *For each i there exists constants independent of* $\{p_i\}$ *such that for* $p_i \geqslant \alpha$

$$(2.3) \qquad 0 < d_i \leqslant \text{constant}/p_i, \qquad 0 < e_i \leqslant \text{constant}/p_i^2,$$

$$(2.4) \qquad e_i \leqslant \tfrac{1}{2} d_i,$$

$$(2.5) \qquad e_i/d_i^2 \leqslant \text{constant},$$

$$(2.6) \qquad \text{for } p_i \text{ sufficiently large } d_i > \text{constant}/p_i.$$

THEOREM 1. *The symmetric tridiagonal matrix A is strictly diagonally dominant; hence, Gauss elimination without pivoting can be used to stably solve* $As'' = b$.

Proof. From (2.3) the entries of A are all positive; from (2.4) $d_{i-1} + d_i - e_{i-1} - e_i \geqslant e_{i-1} + e_i > 0$ as desired.

To deduce the asymptotic behavior of $s(x)$ for large tension parameters it is first necessary to study the behavior of $\{s_i''\}$. This can be established by simple matrix arguments.

LEMMA 2. *There exists a constant K independent of* $\{p_i\}$ *such that*

$$\max_i |s_i'' - b_i/(d_{i-1} + d_i)| \leqslant K \cdot \max_j |b_j|.$$

Proof. Set $D = \text{diag } A$, $B = I - D^{-1}A$; then $s'' - D^{-1}b = A^{-1}DBD^{-1}b$ is an identity. Thus, with the usual matrix and vector infinity norms $\|s'' - D^{-1}b\| \leqslant \|A^{-1}D\| \cdot \|BD^{-1}\| \cdot \|b\|$. From (2.4) $\|I - D^{-1}A\| \leqslant 1/2$, which implies $\|A^{-1}D\| \leqslant 2$. A typical row of BD^{-1} is $(\dots, 0, e_{i-1}/(d_{i-1} + d_i)^2, 0, e_i/(d_i + d_{i+1})^2, 0, \dots)$ so $\|BD^{-1}\| \leqslant 2 \max e_i/d_i^2 \leqslant \text{constant by (2.5)}$.

This lemma has an immediate consequence of some importance. Since $1/(d_{i-1} + d_i)$ behaves asymptotically (for p_{i-1} and p_i large) like $\min(p_{i-1}, p_i)$ from (2.3) and (2.6), it follows that for a sequence of choices of $\{p_i\}$:

(1) if for some i both p_{i-1} and $p_i \rightarrow \infty$, then $s_i'' \sim b_i/(d_{i-1} + d_i)$ which means s_i'' grows like $\min(p_{i-1}, p_i)$;

(2) if for some i either p_{i-1} or p_i is uniformly bounded over the sequence, then so is s_i''.

The behavior of $s(x)$ in the interior of $[x_i, x_{i+1}]$ is now established.

LEMMA 3. *As* $p \rightarrow \infty$ (a) $F' = O(1/p^2)$ *uniformly in any closed subinterval of* $(0, 1)$, (b) $F = O(1/p^2)$ *uniformly in* $[0, 1]$.

Proof. First note that (R-5(a)), (R-5(b)) imply that $-F'(0) \leqslant C_1^2 C_2/p^2$. Since $F'' \geqslant 0$ in $[0, 1]$, $F(z) \geqslant F(0) + zF'(0) = zF'(0)$ so $|F(z)| = -F(z) \leqslant -zF'(0) \leqslant C_1^2 C_2/p^2$. For (a) consider the interval $[0, \alpha]$ with $0 < \alpha < 1$. For any x such that $F'(x) \leqslant 0$, F' nondecreasing from (R-3)) implies $0 \leqslant -F'(x) \leqslant -F'(0) \leqslant C_1^2 C_2/p^2$.

If α is large enough that $F'(\alpha) > 0$, then $F(0) = F(1) = 0$ implies F' has zero mean on $[0, 1]$, or $-\int_0^\alpha F'(t)\,dt = \int_\alpha^1 F'(t)\,dt$. But then

$$(1 - \alpha)F'(\alpha) \leqslant \int_\alpha^1 F'(t)\,dt = -\int_0^\alpha F'(t)\,dt \leqslant -F'(0)\alpha \leqslant \alpha C_1^2 C_2/p^2.$$

Thus, $F'(\alpha) \leqslant \alpha C_1^2 C_2/(p^2(1 - \alpha))$, so again F' nondecreasing gives the correct bound for $F'(x)$ when $F'(x) > 0$, $x \leqslant \alpha$.

The following theorem is an analog of a result in [12] concerning exponential splines. Here $\lambda(x)$ is the broken line interpolant for the data over $[x_1, x_N]$.

THEOREM 2. *Given a sequence of tension splines for which* $p_i \rightarrow \infty$ *for some i. Then,* (a) *in any closed subinterval of* (x_i, x_{i+1}), $s'' \rightarrow 0$, $s' \rightarrow \lambda'$ *uniformly, and* (b) $s \rightarrow \lambda$ *uniformly in* $[x_i, x_{i+1}]$.

Proof. In $[x_i, x_{i+1}]$ for $0 \leqslant j \leqslant 2$,

$$D^j(s - \lambda) = h_i^{2-j}\{s_{i+1}''(D^jF)((x - x_i)/h_i) + (-1)^j s_i''(D^jF)((x_{i+1} - x)/h_i)\},$$

so from (2.6), Lemmas 2 and 3, $\exists K$ such that

$$|D^j(s - \lambda)| \leqslant h_i^{2-j}\{|b_{i+1}|/(d_i + d_{i+1}) + |b_i|/(d_{i-1} + d_i) + K\} \cdot O(1/p_i^2)$$

uniformly on closed subintervals of (x_i, x_{i+1}) for $j = 1, 2$, and on all of $[x_i, x_{i+1}]$ when $j = 0$. The right-hand side is $O(1/p_i)$ as $p_i \rightarrow \infty$ by (2.6).

It follows that in any closed subinterval of (x_i, x_{i+1}) s' has the same sign as λ' for p_i sufficiently large which is important for monotonicity. In order to match convexity or monotonicity of the data in all of $[x_i, x_{i+1}]$, p_{i-1} and p_{i+1} must also be allowed to change. Since b_i is a positive multiple of the second divided difference of f over x_{i-1}, x_i, x_{i+1}, the data is concave upward (downward) when b_i is positive (negative). The following theorem states that the tension splines (1.2) satisfying (R-1)–(R-5) do have the proper asymptotic behavior to match convexity or monotonicity in the data.

THEOREM 3. *If* b_i, b_{i+1} *are positive, then for* p_{i-1}, p_i, p_{i+1} *sufficiently large* $s''(x)$ *is positive in* $[x_i, x_{i+1}]$. *If* λ' *is positive in* $[x_i, x_{i+1}]$, *then for* p_{i-1}, p_i, p_{i+1} *sufficiently large* $s'(x)$ *is positive in* $[x_i, x_{i+1}]$. *The same statements hold with positive replaced by negative.*

Proof. Lemma 2 implies that for p_{i-1}, p_i, p_{i+1} sufficiently large s_i'' and s_{i+1}'' have the same signs as b_i and b_{i+1}, respectively. The first conclusion then follows from (R-3). For the second part, Theorem 2 implies that for p_{i-1}, p_i, p_{i+1} sufficiently large s' has the same sign as λ' in $[x_{i-1}, x_{i+2}]$ except perhaps in small neighborhoods of the knots. Set $\lambda_j' = [x_j, x_{j+1}]f$, the slope of λ' in the jth interval, then from (1.2) and (2.1), $s'(x_j^+) = -s_{j+1}''e_j - s_j''d_j + \lambda_j'$ and $s'(x_j^-) = s_j''d_{j-1} + s_{j-1}'' + \lambda_{j-1}'$. Thus, from continuity of $s'(x)$

$$(2.7) \quad (d_{j-1} + d_j)s'(x_j) = d_{j-1}\lambda_j' + d_j\lambda_{j-1}' + d_je_{j-1}s_{j-1}'' - d_{j-1}e_j s_{j+1}''.$$

But, for $j = i$ or $j = i + 1$

$$|d_j e_{j-1} s_j'' - d_{j-1} e_j s_{j+1}''|/(d_{j-1} + d_j) \leqslant \max\{|s_{j-1}''|, |s_{j+1}''|\} \cdot \max\{e_{j-1}, e_j\}$$

$$= O(1/\min(p_{i-1}, p_i, p_{i+1}))$$

by (2.3) and comment (1) following Lemma 2. Hence, as $p_{i-1}, p_i, p_{i+1} \longrightarrow \infty$, for $j = i$ or $i + 1$, $s'(x_j) \sim (d_{j-1}\lambda_j' + d_j\lambda_{j-1}')/(d_{j-1} + d_j)$, which is a weighted average of quantities with the same sign as λ'.

3. Examples. In this section the examples listed in the introduction are examined to see if they satisfy (R-1)–(R-5). Most of the work for the exponential spline is found in [12]. Note that this spline reduces to the usual cubic interpolatory spline only in the limit as $p \longrightarrow 0$; the special formulas needed for small p are a nuisance in any implementation. For Spath's example of a rational spline, $-F'(0) = 1/(2p^2 + 6p + 6)$ and $F'(1) = (p + 3)/(2p^2 + 6p + 6)$ so (R-5) can be shown to hold with $C_1 = 1/2$, $C_2 = 2$. For (R-3),

$$F''(z) = \{(p + 1)^3/[1 + p(1 - z)]^3 - 1\}/[p(2p^2 + 6p + 6)]$$

is nonnegative; also, if $0 \leqslant z \leqslant 1 - \delta$, $0 < \delta < 1$, $p(2p^2 + 6p + 6)F''(z) \leqslant (p + 1)^3/(p^3\delta^3)$ so $F''(z) \leqslant 1/(p^3\delta^3)$. The remaining restrictions are easily verified. This spline requires at least eight multiplications and three divisions per spline evaluation which is inferior to the two cubics. Also, the formulas for derivatives are rather complicated.

The cubic with $F(z) = \{[pz - (p - 1)]_+^3 - z\}/(6p^2)$ has $-F'(0) = 1/(6p^2)$, $F'(1) = (3p - 1)/(6p^2)$ so (R-5) holds with $C_1 = 1/2$, $C_2 = 3/2$. Since $F''(z) = [pz - (p - 1)]_+$, (R-2) and (R-3) are valid; moreover, for $z \in [0, 1 - \delta]$, $|F''(z)| \leqslant (1 - \delta p)_+ = 0$ as $p \longrightarrow \infty$. Finally, $-F'(0)/F'(1) = 1/(3p - 1)$ is nonincreasing in p.

Nielson's cubic (Example 3) has $-F'(0) = 1/\{2p(p^3 + 2)\}$, $F'(1) = (1 + p^2)/\{2p(p^3 + 2)\}$, so (R-5) holds with $C_1 = 1/2$, $C_2 = 2$. Restrictions (R-1), (R-2) are easily verified, and $-F'(0)/F'(1) = 1/(1 + p^2)$ is nonincreasing in p. However,

$$F''(z) = \{(2z - 1) + (1 + p^2)[pz - (p - 1)]_+ + [1 - pz]_+\}/(2 + p^2)$$

and as $p \longrightarrow \infty$, $F''(z) \sim (2z - 1)/(2 + p^2)$ so (R-3) does not hold. This restriction is used in the proof of Lemma 3 and for the first result of Theorem 3. The conclusions of Lemma 3 can be shown to hold for this case, e.g., in closed subintervals of $(0, 1)$ as $p \longrightarrow \infty$, $F'(z) \sim -z(1 - z)/(p^2 + 2) = O(1/p^2)$. Thus, Theorem 2 is still valid for this example; however, when b_i and b_{i+1} have the same sign this spline will have extra inflection points in $[x_i, x_{i+1}]$ for large p_{i-1}, p_i, p_{i+1}. Fortunately, since $s(x) - \lambda(x) = O(1/p)$ there, these oscillations can be made small enough not to be visible in graphics applications.

4. Algorithm. In this section an algorithm is described for implementing the fourth example given in the introduction. An implementation of Nielson's cubic would proceed similarly except it should be based on the representation (1.1).

There are three phases to the algorithm: (1) computation of $\{s_i''\}$ for a fixed choice of $\{p_i\}$, (2) choice of $\{p_i\}$ so that $s(x)$ satisfies monotonicity or convexity constraints, (3) evaluation of $s(x)$ or its derivatives at some point x. The first phase requires solving the positive definite tridiagonal system for $\{s_i''\}$. For example (4) the coefficients needed are $d_i = (3p_i - 1)h_i/(6p_i^2)$ and $e_i = h_i/(6p_i^2)$. To choose the parameters $\{p_i\}$ one typically starts with all $p_i = 1$; and then by examining the computed $\{s_i''\}$ or first derivative values the parameters are modified, $\{s_i''\}$ recomputed, etc. Details are given below. Once a fixed set of parameters is determined, $s(x)$ could be evaluated directly based on (1.2). However, to take full advantage of the fact that $s(x)$ is a cubic spline, it is better to explicitly calculate the new set of knots given implicitly by $\{p_i\}$ and $\{x_i\}$, then compute the four coefficients in the standard representation of the cubic in each subinterval between knots. These coefficients can be obtained by expanding out the F terms in (1.2). Finally, for each x, first find the knot subinterval in which x lies, e.g., by a binary search, then evaluate the cubic or its derivatives by Horner's scheme. As an example of "new knots", if on the interval $[x_i, x_{i+1}]$ the final $p_i = 3/2$, then the new set of knots would include $x_i, x_i + h_i(p_i - 1)/p_i = x_i + h_i/3, x_i + h_i/p_i = x_i + 2h_i/3, x_{i+1}$. Note that for this cubic spline (example (4)), whenever $p_i > 2$ $s(x)$ is strictly linear in $[x_i + h_i/p_i, x_{i+1} - h_i/p_i]$.

The theory given in Section 2 suggests ways of choosing the $\{p_i\}$ for monotonicity or convexity. For the latter case, s_i'' is forced to have the same sign as b_i. If some $b_i = 0$, Spåth [15] suggests splitting the problem at x_i and using a linear fit for $x \in [x_{i-1}, x_{i+1}]$. This can be done by zeroing certain off-diagonal entries in the coefficient matrix A and putting in slope end conditions at x_{i-1}^- and x_{i+1}^+. For nonzero b_i, if some $s_i'' b_i < 0$, then (2.2) suggests how to modify p_i so that s_i'' and b_i have the same sign: force $|e_{i-1}s_{i-1}'' + e_i s_{i+1}''| \leqslant |b_i|$. Since $e_i = h_i/(6p_j^2)$, try replacing p_j by $sqrt(h_i \cdot \max(|s_{i+1}''|, |s_{i-1}''|)/(3|b_i|))$ for $j = i - 1$ and i. In order to guarantee convergence, in practice p_j is set equal to the maximum of the above quantity and 1.2 times the previous p_j. Of course, changing $\{p_i\}$ will also result in new values for $\{s_i''\}$ so this process must be iterated. Most examples require three or fewer iterations.

Choosing the parameters to ensure monotonicity is more difficult. Clearly, if $s'(x_i)$ has the wrong sign, then p_{i-1} and p_i should be increased; (2.7) suggests forcing

$$|d_i e_{i-1} s_{i-1}'' - d_{i-1} e_i s_{i+1}''| < |d_{i-1} \lambda_i' + d_i \lambda_{i-1}'|.$$

Substituting in the appropriate expressions for d_j, e_j we get

$$p_{i-1}^2 p_i^2 > (h_{i-1} h_i/36)|(3p_i - 1)s_{i-1}'' - (3p_{i-1} - 1)s_{i+1}''|/|d_{i-1}\lambda_i' + d_i \lambda_{i-1}'|.$$

In practice p_{i-1} and p_i have been replaced by

$$sqrt\{(p_{i-1} + p_i) \cdot \max(|s_{i-1}''|, |s_{i+1}''|)/[4 \cdot |p_i \lambda_i'/h_i + p_{i-1}\lambda_{i-1}'/h_{i-1}|]\}.$$

If $\lambda_{i-1}' = \lambda_i' = 0$ then the data is locally flat so the problem can be decomposed into the intervals $[x_1, x_{i-1}]$ and $[x_{i+1}, x_N]$ with $s(x)$ constant in $[x_{i-1}, x_{i+1}]$.
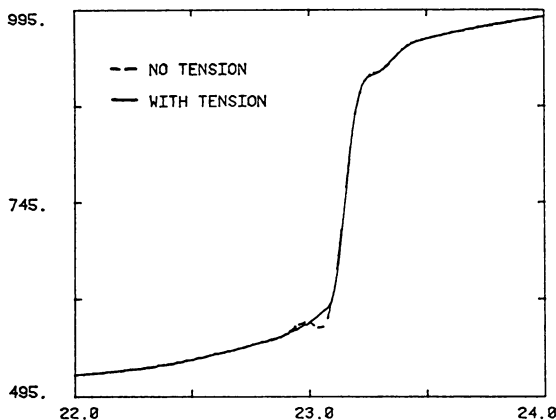
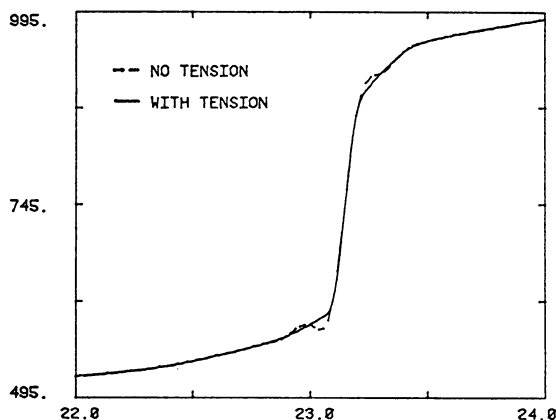FIGURE 1. *Monotonicity forced*

FIGURE 2. *Convexity matched*

Unfortunately, it is possible, though in general unlikely, for s' to change sign twice in (x_i, x_{i+1}). This can occur only if $s_i'' s_{i+1}'' < 0$; when this inequality holds for some i, the extremal value of s' in $[x_i, x_{i+1}]$ should be checked to make sure s' has the desired sign. The critical points of s' depend on the size of p_i: when $p_i \geqslant 2$ they occur at the new knots $x_i + h_i/p_i$ and $x_{i+1} - h_i/p_i$ with extremal value

$$(4.1) \qquad m = \lambda_i' + h_i(s_i'' - s_{i+1}'')/(6p_i^2).$$

When $p_i < 2$ the extremal value is either m or $m - (2 - p_i)^2 h_i - s_i'' s_{i+1}''/[2p_i(s_{i+1}'' - s_i'')]$; the latter comes from the critical point $x_i + h_i[(p_i - 1) + (p_i - 2)s_i''/(s_{i+1}'' - s_i'')]/p_i$. In practice, it has proved sufficient merely to use (4.1) and replace p_i by $sqrt|h_i \cdot (s_i'' - s_{i+1}'')/(6\lambda_i'')|$.

As a test of these ideas consider the data $\{x_i\} = \{1, 5, 10, 15, 20, 21, 22, 22.5, 22.6, 22.7, 22.8, 22.9, 23, 23.1, 23.2, 23.3, 23.4, 23.5, 24, 26, 30\}$, $\{y_i\} = \{373, 415, 438, 459, 491, 503, 523, 543, 550, 557, 565, 575, 590, 620, 860, 915, 944, 958, 986, 1067, 1125\}$ taken from Shampine and Allen [14, p. 40]. Figure 1 shows the results of the first iteration with no tension, i.e., all $p_i = 1$, and the second (final) iteration when monotonicity was forced. Only the region of rapid change in the data is graphed. The only parameters changed were p_{12} and p_{13} which were set to 2.21.

When convexity was matched the graphs in Figure 2 resulted. Here three iterations were needed; the final modified parameters were $p_2 = p_3 = 2.49$, $p_{10} = p_{11} = 2.35$, $p_{12} = p_{13} = 3.13$, $p_{15} = p_{16} = 2.28$, and $p_{18} = p_{19} = 2.92$. Since the physical quality of interest for this problem is the inflection point, the second graph is preferred.

Acknowledgement. The author is grateful to the referees for their careful reading of this paper.

Department of Mathematics and Statistics
The University of New Mexico
Albuquerque, New Mexico 87131

1. A. CLINE, "Curve fitting in one and two dimensions using splines under tension," *Comm. ACM*, v. 17, 1974, pp. 218–223.

2. C. DE BOOR, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.

3. P. DUBE, "Univariate blending functions and alternatives," *Comput. Graphics and Image Processing*, v. 6, 1977, pp. 394–408.

4. P. DUBE, "Automatic generation of parameters for preliminary interactive design of free-form curves." (Submitted for publication.)

5. A. FORREST, "Interactive interpolation and approximation by Bézier polynomials," *Comput. J.*, v. 15, 1972, pp. 71–79.

6. W. GORDON & R. RIESENFELD, "B-spline curves and surfaces," in *Computer Aided Geometric Design* (Barnhill and Riesenfeld, Eds.), Academic Press, New York, 1974, pp. 95–126.

7. D. HILL, *Estimation of Probability Functions Using Splines*, Doctoral Thesis, Univ. of New Mexico, Albuquerque, 1973.

8. D. McALLISTER, E. PASSOW & J. ROULIER, "Algorithms for computing shape preserving spline interpolations to data," *Math. Comp.*, v. 31, 1977, pp. 717–725.

9. G. NIELSON, "Some piecewise polynomial alternatives to splines under tension," in *Computer Aided Geometric Design* (Barnhill and Riesenfeld, Eds.), Academic Press, New York, 1974, pp. 209–236.

10. G. NIELSON, Unpublished notes, Dept. of Mathematics, Arizona State University.

11. E. PASSOW & J. ROULIER, "Monotone and convex spline interpolation," *SIAM J. Numer. Anal.*, v. 14, 1977, pp. 904–909.

12. S. PRUESS, "Properties of splines in tension," *J. Approximation Theory*, v. 17, 1976, pp. 86–96.

13. D. SCHWEIKERT, "An interpolation curve using splines in tension," *J. Math. and Phys.*, v. 45, 1966, pp. 312–317.

14. L. SHAMPINE & R. ALLEN, *Numerical Computing*, Saunders, Philadelphia, Pa., 1973.

15. H. SPÄTH, "Exponential spline interpolation," *Computing*, v. 4, 1969, pp. 225–233.

16. H. SPÄTH, *Spline Algorithms for Curves and Surfaces*, Utilitas Mathematica Publ. Inc., Winnipeg, 1974.

17. R. WIELINGA, "Constrained interpolation using Bézier curves," in *Computer Aided Geometric Design* (Barnhill and Riesenfeld, Eds.), Academic Press, New York, 1974, pp. 153–172.

18. D. McALLISTER & J. ROULIER, "Interpolation by convex quadratic splines," *Math. Comp.*, v. 32, 1978, pp. 1154–1162.