ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 22, Number 2, Spring 1992

WEIGHTED SPLINES AS OPTIMAL INTERPOLANTS

K. ŠALKAUSKAS AND L. BOS

ABSTRACT. We consider interpolation by C^1 cubic splines which minimize a weighted semi-norm. The weight function is piecewise constant but on a subdivision potentially different from that determined by the interpolation knots.

Introduction. Suppose that we are given a set of data (x_i, f_i), $1 \le i \le N$, with $x_1 < x_2 < \cdots < x_N$. As is well known [**1**], among all functions f which interpolate the data (i.e., $f(x_i) = f_i$, $1 \le i \le N$) and which have an absolutely continuous first derivative and square integrable second derivative, the one for which $\int_{x_1}^{x_N} (f^{(2)}(x))^2\,dx$ is a minimum, is the natural cubic interpolating spline. This function is twice continuously differentiable on $[x_1, x_N]$ and is such that its restrictions to each of the subintervals, $[x_i, x_{i+1}]$, is a cubic polynomial. The adjective "natural" indicates that it may be extended by straight lines to a C^2 function, on all of $\mathbf{R}$, whose second derivative is in $L_2(\mathbf{R})$. This condition is easily seen to be equivalent to having zero second derivatives at the end points, x_1 and x_N. The use of the functional $\int_{x_1}^{x_N} (f^{(2)}(x))^2\,dx$ is motivated by the fact that it is a linearization of the bending energy of a thin elastic rod of uniform stiffness. Although cubic splines have found widespread application, there are data sets for which natural splines are not appropriate. Figure 1 below illustrates one such example.

Because of this, the first author introduced in [**4**] the weighted cubic spline, minimizing instead the weighted functional (or semi-norm)

$$|v|^2 := \int_{x_1}^{x_N} w(x)(v^{(2)}(x))^2\,dx$$

in the hope of being able to choose a weighting for which the resulting interpolant is not as unexpectedly oscillatory. To motivate our choice

This work was supported by NSERC Canada grants A8521 and A8389, respectively.
Received by the editors on February 14, 1989 and in revised form on July 4, 1989.

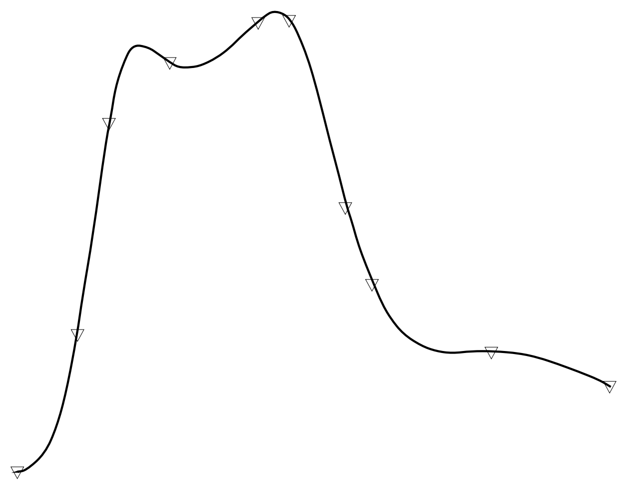

FIGURE 1. A natural cubic interpolating spline.

of weight function, consider a thin elastica, interpolating the data. Its bending energy is given by $\int_{x_1}^{x_N} EI\{1+[s'(x)]^2\}^{-5/2}[s''(x)]^2\,dx$. A description of this may be found in [**3**]. In the construction of natural cubic spline interpolants, it is assumed that $s'(x)$ is close to zero and that EI = constant, which we may take to be unity. Suppose now that the elastica is softened at the interior knots by replacing $EI = 1$ by $EI = 1-\alpha\sum_{i=2}^{N-1} p_\varepsilon(x-x_i)$, where $p_\varepsilon(x)$ is a rectangular pulse of width 2ε centered at the origin, of unit height and $0<\alpha<1$. A typical such EI is plotted in Figure 2.

When α is close to 1, an interpolant of this sort will be essentially

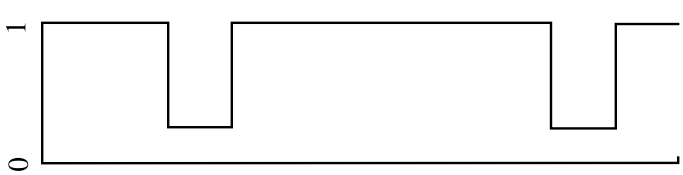

FIGURE 2. A "softened" EI.

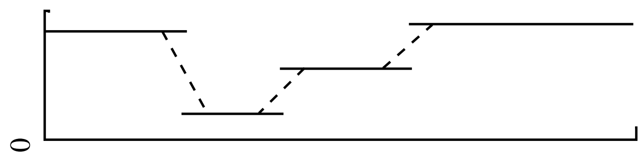

FIGURE 3. A piecewise constant weight function convolved with a unit pulse.

piecewise linear, with its second derivative concentrated on the supports of the translated pulses. Consequently, $\{1 + [s'(x)]^2\}^{-5/2}$ is very nearly piecewise constant. A reasonable first approximation to $\{1 + [s'(x)]^2\}^{-5/2}$ is the piecewise constant weight function $w(x)$ given by

$$w(x)|_{[x_i,x_{i+1})} := (1 + m_i^2)^{-5/2} \tag{1}$$

where m_i is the slope $(f_{i+1} - f_i)/(x_{i+1} - x_i)$. A better approximation is the convolution $(1/2\varepsilon)p_\varepsilon * w$, which incorporates an approximation to the transition between the linear segments and increases the "soft zone" to 4ε. An example of such a weight function is given in Figure 3. The resulting interpolant would be expected to be not as angular.

In general, one may take w to be a positive function in $L_1(x_1, x_N)$. We state some basic existence properties of weighted splines. First, let Σ denote the space of C^1 piecewise cubics on $[x_1, x_N]$ with knots x_i and H^2 the space of functions on $[x_1, x_N]$ with absolutely continuous first derivative and square integrable second derivative.

Theorem 1. (a) *There exists a unique $v \in \Sigma$ of minimal semi-norm interpolating the data.*

(b) *If w is piecewise constant on the partition $\pi_1 : x_1 < \cdots < x_N$, then there is a unique $v \in H^2$ of minimal semi-norm interpolating the data, and, moreover, $v \in \Sigma$.*

The proof can be found in [**4**].

The piecewise constant choice of weight function mentioned above has proven to be quite successful. In [**4**] the exponent -3 was used but

FIGURE 4. Weighted spline with weight function given by (1).

there seems to be very little difference between the two. With these weights, the weighted functional seems to be a closer approximation to the bending energy of a thin elastic rod of nonuniform stiffness. The figures illustrate the behavior. Figure 1 is again the graph of the natural cubic spline which interpolates the indicated data. Figure 4 is the graph of the weighted spline which interpolates the same data using the weights given by (1)—evidently a much better interpolant. Note, however, the sharp bends in the graph and the resulting impression of angularity. This seems to be a characteristic of weighted splines using weights (1). In Figure 5 some of the angularity has been removed by using the slightly "smoother" weight function obtained by convolving the piecewise constant weight function with a unit pulse as mentioned above. Here $\varepsilon = .005$. Note, however, in this example the minimization was done in the smaller space Σ. The defining equations may be found in [**4**]. At present, we have only empirical and heuristic evidence for the effectiveness of this smoothing procedure.

In this work we present a more general procedure for influencing the shape of a weighted spline and for adding some flexibility and "roundness" to the weighted spline if desired. Specifically, we consider the use of weight functions which are still piecewise constant, but are so on a partition which may differ from that determined by the data

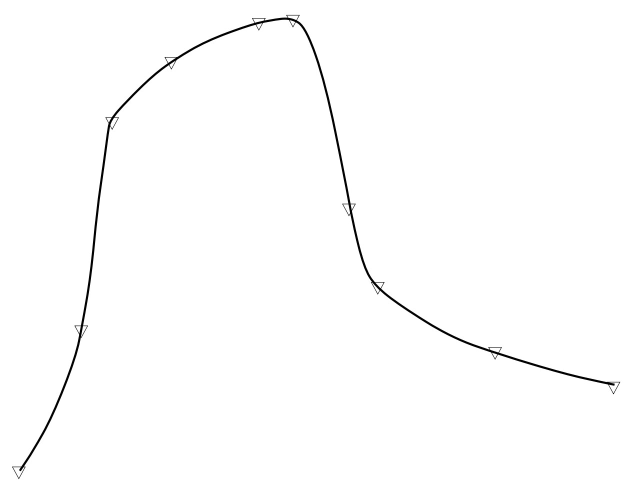

FIGURE 5. Weighted spline with convolved weight function.

points. We show that the interpolants which minimize the resulting weighted semi-norm are still C^1 piecewise cubics, but with knots at the data points and also at the discontinuities of the weight function. We give continuity conditions which characterize this cubic spline, and, finally, we give some examples of its use.

Weighted splines. Let $X := \{v \in C^1(\mathbf{R}) : Dv$ is absolutely continuous and $D^2 v \in L_2(\mathbf{R})\}$. We use operator notation for the derivative. Introduce the weighted semi-inner product

$$(u, v)_w := \int_{\mathbf{R}} w D^2 u D^2 v,$$

where w satisfies

(a) $w(t) > 0$, $t \in \mathbf{R}$,

(b) w is continuous from the right and piecewise constant on a partition $\pi_2 : \tau_1 < \tau_2 < \cdots < \tau_M$, satisfying $x_1 \leq \tau_1$ and $\tau_M \leq x_N$.

Note that τ's are not constrained in any way by the data points except for the first and last. Associated with this semi-inner product is, of course, the semi-norm

$$|v|_w^2 := (v, v)_w.$$

This makes X into a semi-Hilbert space. Our main result may now be stated.

Theorem 2. *There exists a unique $s \in X$ of minimal semi-norm interpolating the data. Furthermore,*

(a) *s is a C^1 piecewise cubic on $\mathbf{R}$ with knots $\pi_1 \cup \pi_2$,*

(b) *wD^2s has zero jump at each knot and $D^2s(t) = 0$ for t outside (x_1, x_N),*

(c) *wD^3s has zero jump at each τ_i not an x_j.*

(Here we say that f has zero jump at a point ξ if $\lim_{x\to\xi^+} f(x) = \lim_{x\to\xi^-} f(x)$.)

Proof. Suppose for the moment that such an $s(x)$, as described by (a), (b) and (c), exists. We first show that it is then indeed the unique minimizer in X. Let $f \in X$ be any other interpolant of the same data and consider

$$\int_{\mathbf{R}} (D^2s(x) - D^2f(x))D^2s(x)w(x)\,dx.$$

As $s(x)$ is piecewise cubic (on the refined partition), $D^2s(x)w(x)$ is piecewise linear. By hypothesis (b), it is also continuous and therefore absolutely continuous and we may integrate by parts to obtain

$$\int_{\mathbf{R}} (Ds(x) - Df(x))D(D^2s(x)w(x))\,dx.$$

Now, as $D^2s(x)$ is zero outside (x_1, x_N), this is actually

$$\int_{x_1}^{x_N} (Ds(x) - Df(x))D(D^2s(x)w(x))\,dx,$$

which we may write as

$$\sum_{i=1}^{N-1} \int_{x_i}^{x_{i+1}} (Ds(x) - Df(x))D(D^2s(x)w(x))\,dx.$$

But, by (c), on each subinterval $[x_i, x_{i+1}]$, $D(D^2s(x)w(x)){=}D^3s(x)w(x)$ and is continuous, i.e., constant, and we obtain

$$\sum_{i=1}^{N-1}\{D^3s(x)w(x)|_{[x_i,x_{i+1})}\}\int_{x_i}^{x_{i+1}}(Ds(x)-Df(x))\,dx.$$

However, as

$$\int_{x_i}^{x_{i+1}}(Ds(x)-Df(x))\,dx=0 \qquad \text{(by interpolation)},$$

we see that

$$\int_{\mathbf{R}}(D^2s(x)-D^2f(x))D^2s(x)w(x)\,dx=0,$$

which we may express as

$$(s-f,s)_w=0, \tag{2}$$

in terms of the semi-inner product of X. Therefore,

$$|f|_w^2=|(f-s)+s|_w^2=|f-s|_w^2+|s|_w^2\geq|s|_w^2,$$

and we see that s is optimal. In fact, $|f|_w^2=|s|_w^2$ only when $|f-s|_w=0$. But then $f-s$ must be a linear. This linear is in fact zero as $f-s$ is zero at at least the two point: x_1 and x_N; $s(x)$ is thus unique.

We now show that such an $s(x)$ exists. Suppose that $t_1<t_2<\cdots<t_n$ with $t_1=x_1$ and $t_n=x_N$ are the knots of $\pi_1\cup\pi_2$. The determination of any piecewise cubic on this partition is equivalent to finding the four coefficients of each of the cubics on $[t_i,t_{i+1}]$. Outside of $[x_1,x_N]$ it is extended uniquely by straight lines. The partition gives $n-1$ subintervals and hence we must determine $4(n-1)$ coefficients. The condition (a) gives us two continuity conditions at each of the $n-2$ interior knots for a total of $2(n-2)$. Condition (b) gives one continuity condition at each of the $n-2$ interior knots and two end conditions for a total of n. Condition (c) gives one continuity condition at each of the $n-N$ noninterpolated knots and then there are also N interpolation

conditions which together yield n conditions. Thus, the total number of conditions on our piecewise cubic is

$$2(n-2)+n+n=4n-4=4(n-1),$$

exactly the number of coefficients to be found. As each condition is linear in the unknowns, we have, in effect, a $4(n-1)$ square system of linear equations to solve. For such systems, existence is equivalent to uniqueness and by our previous work, we are done. $\square$

The orthogonality conditions (2) can be exploited to yield a smaller set of equations for determining the spline s. Since we have shown (Theorem 2) that s is a C^1 piecewise cubic with knots t_i, $i=1,2,\dots,n$, it has a unique representation on $[x_1,x_N]$ in terms of the Hermitian basis $\{\varphi_i\}_1^n \cup \{\psi_i\}_1^n$ defined by

$$\varphi_i(x)=\begin{cases} -2(x-t_{i-1})^2(x-t_i-h_{i-1}/2)/h_{i-1}^3, & x\in[t_{i-1},t_i),\\ 2(x-t_i+h_i/2)(x-t_{i+1})^2/h_i^3, & x\in[t_i,t_{i+1}),\\ 0, & \text{otherwise,}\end{cases}$$

and

$$\psi_i(x)=\begin{cases} (x-t_{i-1})^2(x-t_i)/h_{i-1}^2, & x\in[t_{i-1},t_i),\\ (x-t_i)(x-t_{i+1})^2/h_i^2, & x\in[t_i,t_{i+1}),\\ 0, & \text{otherwise.}\end{cases}$$

Here $h_i := t_{i+1}-t_i$, and $t_0 < t_1$, $t_{n+1} > t_n$ are fixed but arbitrary. This familiar basis satisfies the cardinality conditions

$$\varphi_i(t_j)=\delta_{ij}, \qquad D\varphi_i(t_j)=0, \qquad \psi_i(t_j)=0, \qquad D\psi_i(t_j)=\delta_{ij}.$$

We may now write

$$s(x)=\sum_{i=1}^n s(t_i)\varphi_i(x)+\sum_{i=1}^n Ds(t_i)\psi_i(x). \tag{3}$$

The values of $s(t_i)$ are known when $t_i=x_j$; then $s(t_i)=f_j$, by interpolation. The remaining $n-N$ values of $s(t_i)$ and n values of $Ds(t_i)$ can be found by making use of (2), which says that s is orthogonal to every interpolant of zero (at the x_i's). The functions in

$\{\psi_i\}_1^n$ are such interpolants, hence s satisfies the n equations $(s, \psi_i)_w = 0$, $i = 1, \dots , n$. These semi-inner products, when computed for s in the form (3), yield a set of equations identical in form to those derived in [**4**], which in turn are just like the tridiagonal system arising in the construction of ordinary cubic splines. Thus, if we denote

$$w_i = w(x)|_{[t_i,t_{i+1})}, \qquad Ds(t_i) = m_i,$$
$$\lambda_i = w_{i-1}h_i/(w_{i-1}h_i + w_ih_{i-1}), \qquad \mu_i = 1 - \lambda_i, \quad i = 2, \dots , n-1,$$

the equations are

$$\begin{gathered} 2m_1 + m_2 = -3[f_1 - s(t_2)]/h_1, \\ \lambda_i m_{i-1} + 2m_i + \mu_i m_{i+1} = 3\lambda_i[s(t_i) - s(t_{i-1})]/h_{i-1} \\ +3\mu_i[s(t_{i+1}) - s(t_i)]/h_i, i = 2, \dots , n-1, \\ m_{n-1} + 2m_n = 3[f_N - s(t_{n-1})]/h_{n-1}. \end{gathered}$$

It should be remembered here that $x_1 = t_1$, $x_N = t_n$, and $s(t_i)$ is an unknown if t_i is not an x_j. It will be convenient to define $I := \{i | 1 \le i \le n \text{ and } t_i \ne x_j, 1 \le j \le N\}$, and note that I consists of $n-N$ distinct indices. When the unknowns $s(t_i)$, $i \in I$, are transposed to the left-hand side, there results a sparse, linear, $n \times (2n-N)$ system consisting of tri-diagonal bands. We obtain $n - N$ more equations by noting that if $i \in I$, then φ_i is also an interpolant of zero. There are thus $n - N$ equations $(s, \varphi_i)_w = 0$. The resulting system is also sparse and consist of tri-diagonal bands. The equations in this system are

$$\begin{gathered} -w_{i-1}m_{i-1}/h_{i-1}^2 + m_i(w_i/h_i^2 - w_{i-1}/h_{i-1}^2) + w_im_{i+1}/h_i^2 \\ -2w_{i-1}s(t_{i-1})/h_{i-1}^3 + 2s(t_i)[w_{i-1}/h_{i-1}^3 + w_i/h_i^3] - 2w_is(t_{i+1})/h_i^3 = 0, \\ i \in I. \end{gathered}$$

We now show that the above equations have a solution.

Proposition 3. *The coefficient matrix of the equations* $(s, \psi_i)_w = 0$, $1 \le i \le n$ *and* $(s, \varphi_i)_w = 0$, $i \in I$, *is nonsingular.*

Proof. We will show that the homogeneous system has a unique solution. The "right-hand side" is certainly zero if $f_1, \dots , f_N$ vanish. Let $f_1, \dots , f_N$ be a data set for which the system is homogeneous.

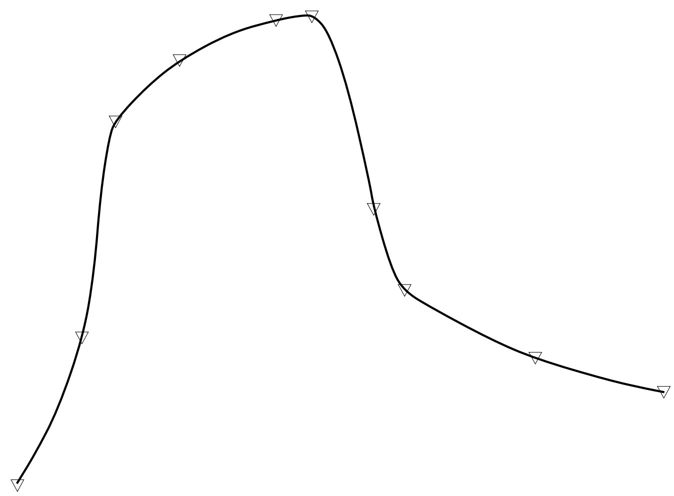

FIGURE 6. One noninterpolated knot mid-way between the two left-most data points.

Suppose that s_1 and s_2 are splines interpolating this data, whose parameters are obtained from distinct solutions of the system. Let $z := s_1 - s_2$. This is an interpolant of zero and can be written in the form $z = \sum_{i=1}^n \alpha_i\psi_i + \sum_{i\in I}\beta_i\varphi_i$. Since s_1 and s_2 satisfy the orthogonality conditions, so does z, i.e., z is orthogonal to the span of $\{\psi_i\}_1^n \cup \{\varphi_i\}_{i\in I}$ and hence to itself. It follows that $z = 0$ and $s_1 = s_2$. □

An alternative way of generating a banded system is to use the B-spline basis described by Foley [**2**], with modifications stemming from the fact that there are potentially more knots than data points, and thus the third derivative condition of Theorem 2 has to be invoked.

To illustrate the effect of allowing refined partitions for the weight function, consider the same data set as in Figures 1, 4 and 5. In Figure 6 we have taken the weights of Figure 4 and added a knot half way between the two left-most interpolation points. Clearly, the resulting spline turns more gently at the second data point. Figure 7 shows the additional improvement obtained by adding a knot between the second and third data points from the right. In these examples we have chosen the weights and additional knots so as to add some roundness to the

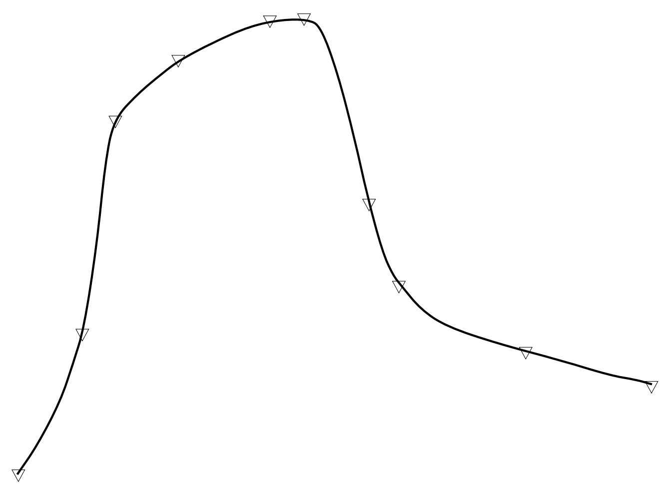

FIGURE 7. Second noninterpolated knot mid-way between the 8th and 9th data points.

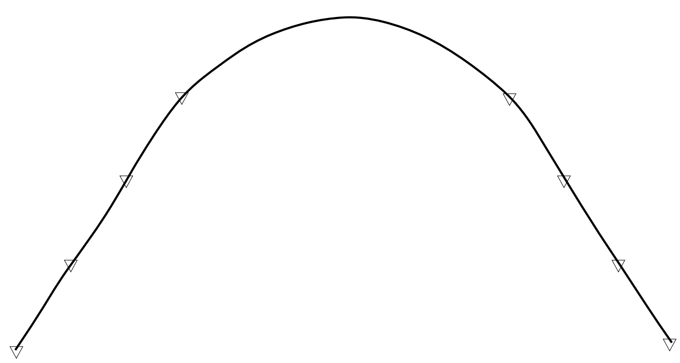

FIGURE 8. Interpolant of absolute value, weights given by (1).

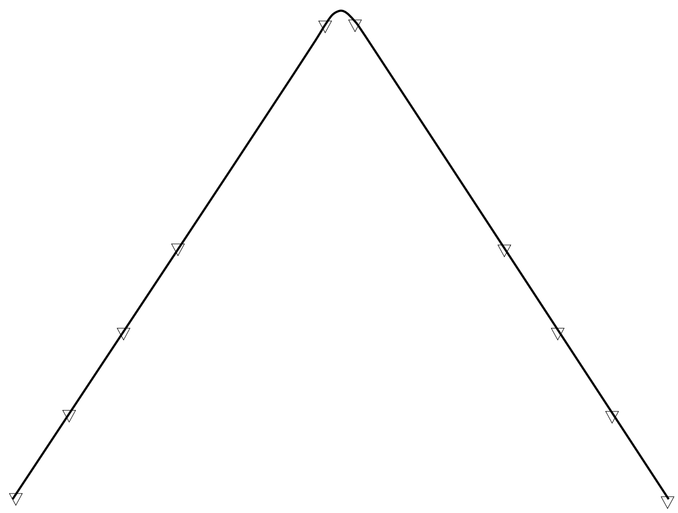

FIGURE 9. Interpolant of absolute value, two noninterpolated knots.

curves. By Theorem 2(b), the jump in the second derivative at a knot is determined by the ratio of the weights on either side of the knot. Thus, the desired effect is accomplished by selecting the weights so as to reduce the ratios of consecutive weights.

In general, the choice of weights and τ's may influence the shape in other ways. Figure 8 shows the weighted spline with weights given by (1) interpolating the absolute-value function at the indicated points. In Figure 9 we have added τ_1 and τ_2 symmetrically about the origin, with a weight of 0.001 assigned on $[\tau_1, \tau_2)$, and the other weights as before. This tends to concentrate the curvature on (τ_1, τ_2).

REFERENCES

1. C. de Boor, *A practical guide to splines*, Springer-Verlag, New York, 1978.

2. T.A. Foley, *Local control of interval tension using weighted splines*, CAGD **3** (1986), 281–294.

3. M.A. Malcolm, *On the computation of nonlinear spline functions*, SIAM J. Numerical Analysis **14** (1977), 254–282.

4. K. Šalkauskas, C^1 *splines for interpolation of rapidly varying data*, Rocky Mountain J. Mathematics **14** (1984), 239–250.

Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4