MINIMAL NORM CONSTRAINED INTERPOLATION

Larry Dean Irvine

OLD DOMINION UNIVERSITY
Norfolk, Virginia

Grant NAG1-363
September 1985

NOV 26 1985

NASA
National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23665
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>THE NATURAL SPLINE INTERPOLANT</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>A MINIMAL NORM INTERPOLATION PROBLEM IN THE $L_p[a,b]$ SPACES</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>THE CONVEX SPLINE INTERPOLANT</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>THE SHAPE-PRESERVING SPLINE INTERPOLANT</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>CONSTRAINED MINIMIZATION IN A DUAL SPACE</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>LIST OF REFERENCES</td>
<td>76</td>
</tr>
<tr>
<td>A</td>
<td>A PROGRAM FOR CONSTRUCTING THE NATURAL CUBIC SPLINE INTERPOLANT TO GIVEN DATA</td>
<td>77</td>
</tr>
<tr>
<td>B</td>
<td>A PROGRAM FOR CONSTRUCTING THE SHAPE-PRESERVING CUBIC SPLINE INTERPOLANT TO GIVEN DATA</td>
<td>82</td>
</tr>
</tbody>
</table>
This Page Intentionally Left Blank
Acknowledgements

I wish to express my appreciation to Dr. Philip Smith, Dr. Robert Smith, and Dr. John Swetits for their advice, assistance, and encouragement during my doctoral work and research. I wish to thank Dr. Surendra N. Tiwarj for the financial support which I received from the Institute for Computational and Applied Mechanics.

I also wish to thank Barbara Jeffrey for typing the dissertation.
ABSTRACT

MINIMAL NORM CONSTRAINED INTERPOLATION

Larry Dean Irvine
Old Dominion University, 1985
Director: Dr. Philip W. Smith

In computational fluid dynamics and in CAD/CAM a physical boundary, usually known only discretely (say, from a set of measurements), must often be approximated. An acceptable approximation must, of course, preserve the salient features of the data (convexity, concavity, etc.) In this dissertation we compute a smooth interpolant which is locally convex where the data are locally convex and is locally concave where the data are locally concave.

Such an interpolant is found by posing and solving a minimization problem. The solution is a piecewise cubic polynomial. We actually solve this problem indirectly by using the Peano kernel theorem to recast this problem into an equivalent minimization problem having the second derivative of the interpolant as the solution.

We are then led to solve a nonlinear system of equations. We show that with Newton's method we have an exceptionally attractive and efficient method for solving this nonlinear system of equations.

We display examples of such interpolants as well as convergence results obtained by using Newton's method. We list a FORTRAN program to compute these shape-preserving interpolants.

Next we consider the problem of computing the interpolant of minimal norm from a convex cone in a normed dual space. This is an extension of de Boor's work on minimal norm unconstrained interpolation.
1. The Natural Spline Interpolant

We consider the problem of computing an interpolant to given data. Throughout our discussion we shall denote the data

\[(t_i, y_i) \quad i = 1, 2, \ldots, n\]

where \(a = t_1 < t_2 < \ldots < t_n = b\) and in this chapter we place no restrictions on the numbers \(y_i\). There are, of course, many such interpolants which we can form. For example, we can calculate the unique polynomial \(p\) of order \(n\) (degree \(n-1\) or less) which interpolates the data. However, as pointed out in [deB(1), chapter 2], for large \(n\) (and especially for equally spaced points \(t_i\)) the polynomial interpolant is notorious for large changes in its first derivative near the endpoints. Figure (1.1) displays the polynomial interpolant to the function

\[f(t) = \frac{1 - \sin(7 \pi t)}{2}\]

at the points \(t_i = (i-1)/10\) for \(i = 1, 2, \ldots, 11\). Since \(0 \leq y_i \leq 1\) for each \(i\), we expect a good interpolant to remain reasonably close to these bounds. However, because of its behavior near the endpoints, the polynomial interpolant fails to model the data well. This behavior is typical of high-order polynomial interpolants.

In order to decrease the unnaturally large changes in the first derivative characteristic of the polynomial interpolant, we wish to calculate the interpolant which "bends" the least over all suitable interpolants. The norm of the second derivative of an interpolant will furnish a measure of the bending of the interpolant so we pose a minimization problem on \(L_2^{(2)}[a,b]\), the Sobolev space of functions with
Figure (1.1): The Polynomial Interpolant.
second derivatives in the normed linear space $L_2[a,b]$. Let $A$ denote the set of all interpolants in the Sobolev space. We consider the minimization problem

Find $f_\star \in A$ such that $\|f_\star (2)\|_2 \leq \|f(2)\|_2$ for all $f \in A$. (1.1)

We shall see that the solution to (1.1) is piecewise cubic with two continuous derivatives; that is

$$f_\star (t) = p_1(t) \quad \text{if} \quad t_1 \leq t \leq t_{1+1}$$

for $i = 1, 2, \ldots, n-1$ where $p_1$ is a cubic polynomial and $f_\star$ is in $C^2[a,b]$. We follow the pattern in [deB(1), chapter 5], taking advantage of the fact that $L_2[a,b]$ is not only a normed linear space, but also a Hilbert space with an inner product defined by

$$(f, g) = \int_a^b f(t)g(t)dt$$

for any two elements $f$ and $g$ in $L_2[a,b]$.

Assume $f$ is an element of $A$. (The set $A$ is nonempty since it contains the polynomial interpolant.) We shall use the Peano kernel theorem to obtain a set of equations for $f(2)$. By the Fundamental Theorem of Calculus we have

$$f(t) = f(a) + \int_a^t f(1)(s)ds \quad (1.2)$$

We integrate $\int_a^t f(1)(s)ds$ by parts noting that $\int u dv = uv - \int v du$. Let

$$u(s) = f(1)(s) \text{ and } dv(s) = ds$$

so that

$$du(s) = f(2)(s)ds \text{ and } v(s) = -(t-s)$$
where $t$ is a constant. Hence

$$\int_a^t f^{(1)}(s) \, ds = (t-a)f^{(1)}(a) + \int_a^t (t-s)f^{(2)}(s) \, ds$$

and so (1.2) becomes

$$f(t) = q_1(t) + \int_a^t (t-s)f^{(2)}(s) \, ds \quad (1.3)$$

where $q_1(t) = f(a) + f^{(1)}(a)(t-a)$. (This is actually a Taylor's series with integral remainder.)

To acquire constant limit of integration we can write (1.3) as

$$f(t) = q_1(t) + \int_a^b (t-s)f^{(2)}(s) \, ds \quad (1.4)$$

where $(h)_+$, the positive part of the function $h$, is defined by

$$(h)_+(t) = \begin{cases} h(t) & \text{if } h(t) \geq 0 \\ 0 & \text{if } h(t) < 0. \end{cases}$$

Now we consider the divided difference operator. Given a function $g$ and a set of points $\{\tau_1, \tau_{1+1}, \ldots, \tau_{1+m}\}$, the $m$-th divided difference of $g$ — denoted by $[\tau_1, \tau_{1+1}, \ldots, \tau_{1+m}]g(*)$ — is the leading coefficient of the polynomial of order $m+1$ which interpolates $g$ at $\tau_1, \tau_{1+1}, \ldots, \tau_{1+m}$ (and hence is a function of $\tau_1, \tau_{1+1}, \ldots, \tau_{1+m}$). The recursive relations

$$[\tau_p]g(*) = g(\tau_p)$$

$$[\tau_1, \tau_{1+1}, \ldots, \tau_{1+m}]g(*) = \frac{[\tau_{1+1}, \ldots, \tau_{1+m}]g(*) - [\tau_1, \ldots, \tau_{1+m-1}]g(*)}{\tau_{1+m} - \tau_1} \quad (1.5)$$

hold if $\tau_{1+m} \neq \tau_1$ (which we assume for our data). Presently we are interested in the case $m=2$. Equation (1.5) becomes (with $\tau_1 = t_1$)
\begin{align}
(t_{i+2} - t_i)[t_i, t_{i+1}, t_{i+2}]g(*) &= \frac{g(t_{i+2}) - g(t_{i+1})}{t_{i+2} - t_{i+1}} - \frac{g(t_{i+1}) - g(t_i)}{t_{i+1} - t_i} \\
\text{which is computable for } i = 1, 2, \ldots, n-2.
\end{align}

Notice that \([t_1, t_{i+1}, \ldots, t_{i+m}]p(*) = 0\) if \(p\) is a polynomial of order \(m\) or less (degree \(m-1\) or less). (From equation (1.6) we see that \((t_{i+2} - t_i)[t_i, t_{i+1}, t_{i+2}]g(*)\) measures a difference in slopes; the difference in slopes being zero if \(g\) is linear.)

Now we apply the (scaled) second-divided difference operator \((t_{i+2} - t_i)[t_i, t_{i+1}, t_{i+2}]\) to (1.4) and interchange the order of the integral and divided difference operators to obtain

\begin{align}
\tilde{d}_{i,2} &= \int_b^a [g(s)N_{i,1}(s)]ds \quad i = 1, 2, \ldots, n-2
\end{align}

where

\begin{align}
\tilde{d}_{i,2} &= (t_{i+2} - t_i)[t_i, t_{i+1}, t_{i+2}]f(*) \\
&= \frac{y_{i+2} - y_{i+1}}{t_{i+2} - t_{i+1}} - \frac{y_{i+1} - y_i}{t_{i+1} - t_i}, \quad (1.8)
\end{align}

\begin{align}
N_{i,2}(*) &= (t_{i+2} - t_i)[t_i, t_{i+1}, t_{i+2}](-s)_+ \\
&= \frac{(t_{i+2} - s)_+ - (t_{i+1} - s)_+}{t_{i+2} - t_{i+1}} - \frac{(t_{i+1} - s)_+ - (t_i - s)_+}{t_{i+1} - t_i} \quad (1.9)
\end{align}

and \(g = f(2)\). We call \(N_{i,2}\) the (normalized) linear B-spline (or B-spline of order 2) with knots \(t_1, t_{i+1}\) and \(t_{i+2}\). The graph of \(N_{i,2}\) is displayed in figure (1.2).
Figure (1.2): The Normalized Linear B-spline.
We have shown that if \( f \) is an interpolant in the Sobolev space \( (f \in A) \), then \( g = f^{(2)} \) satisfies (1.7). Let the set \( B \) consist of all functions which are in \( L_2[a,b] \) and which satisfy (1.7).

Now consider the problem

Find \( g_\star \in B \) such that \( \|g_\star\|_2 \leq \|g\|_2 \) for all \( g \in B \) \hspace{1cm} (1.10)

A unique solution exists since (1.10) is a minimal norm problem over a nonempty closed convex set in a Hilbert space. Furthermore, the solutions of problems (1.1) and (1.10) are related via \( g_\star = f_\star^{(2)} \).

Hence, to compute \( f_\star \) we can first calculate \( g_\star \) and then integrate \( g_\star \) twice. Since much of our emphasis will be on \( g_\star \), rather than \( f_\star \), we shall call \( g_\star \) the interpolant of minimal norm.

For brevity we denote the index \( m = n-2 \), the B-spline \( N_1 = N_{1,2} \), and the divided difference \( d_1 = d_{1,2} \). We also define the vector-valued function \( T : L_2[a,b] \rightarrow \mathbb{R}^m \) by

\[
(Tx)_i = \int_a^b x(t)N_i(t)dt \quad i=1,2,\ldots,m.
\]

To solve problem (1.10) we shall show that \( g_\star \), the interpolant of minimal norm, is the intersection of two specific sets—one an orthogonal complement of a subspace and the other a translate of a subspace—in \( L_2[a,b] \) via a variation of the Projection Theorem. If \( W \) is a closed subspace of a Hilbert space \( H \) and \( x \) is an arbitrary element of \( H \), then the Projection Theorem states that there exists a unique element \( w_0 \) in \( W \) satisfying

\[
\|x - w_0\| \leq \|x - w\| \quad \text{for all} \; w \in W
\]

and characterized by
\[(x - w_0, w) = 0\] for all \(w \in W\).

Hence \(x - w_0\) is in \(W^\perp\), the orthogonal complement of \(W\). The proof of the Projection Theorem can be found in any book dealing with Hilbert spaces (for example, [L, page 517]). The next proposition will serve as the actual form of the Projection theorem which we shall use.

**Proposition ([L, page 64]):** Let \(W\) be a closed subspace in a Hilbert space \(H\). For a fixed element \(x\) in \(H\) define \(V = x + W\). Then there exists a unique element \(x_0\) in \(V\) of minimal norm. Furthermore, \(x_0\) is in \(W^\perp\).

(The translate \(V\) is called an affine set or linear variety.) Notice that \(x_0\) is the intersection of the orthogonal complement of \(W\) and the translate \(V\) of \(W\). In fact, (1.11) reveals that \(x_0 = x - w_0\).

Define

\[W^\perp = \{z \in L_2[a,b] : Tz = 0\}\]

which is a closed subspace in \(L_2[a,b]\). Let \(g \in L_2[a,b]\) be any element such that \(Tg = d\). (Equivalently, let \(g\) be any element of \(B\).) Then \(B = g + W\) and \(B\) corresponds to the linear variety in the proposition. Hence \(g_0\) is the unique element in \(W^\perp\) satisfying \(Tg_0 = d\).

We consider the contents of \(W^\perp\). Any element which is orthogonal to each \(N_j\) is also orthogonal to any linear combination of the \(B\)-splines. Hence \(S = \text{span}(N_1, N_2, \ldots, N_m)\) is a subset of \(W^\perp\). We now show that \(W^\perp\) is a subset of \(S\) (and hence \(S = W^\perp\)) by contradiction. Assume that there exists an element \(y\) which is in \(W^\perp\) but not in \(S\). Since \(S\) is a closed subspace there exists (by the Projection Theorem)
an element \( s_0 \) in \( S \) such that

\[
\| y - s_0 \| \leq \| y - s \| \quad \text{for all } s \in S
\]

with \( y - s_0 \) in the orthogonal complement of \( S \). This implies

\( T(y - s_0) = 0 \) or \( (y - s_0) \in W \).

However \( y - s_0 \) is also in \( W^\perp \) since both \( y \) and \( s_0 \) are in \( W^\perp \). Therefore \( (y - s_0) = 0 \) and \( S = W^\perp \).

In summary, \( g_\alpha \) is characterized by

\[
g_\alpha = \sum_{i=1}^{m} \alpha_i N_i
\]

(since \( g_\alpha \) is in the span of the B-splines) where the coefficients \( \alpha_1, \alpha_2, \ldots, \alpha_m \) are chosen to satisfy

\[
( \sum_{j=1}^{m} \alpha_j N_j, N_i ) = d_i \quad i = 1, 2, \ldots, m
\]

(since \( Tg_\alpha = d \)). Equation (1.13), a system of \( m \) linear equations in \( m \) unknowns, can be written in matrix notation as

\[
A\mathbf{\alpha} = \mathbf{d}
\]

where the symmetric matrix \( A \) has entries \( A_{ij} = (N_i, N_j) \).

Because the B-splines are linearly independent, the matrix \( A \), a Grahm matrix, is nonsingular and hence a unique solution exists for any given \( d \). Furthermore, since \( N_1 \) has support \([t_1, t_{1+2}]\), the matrix \( A \) is tridiagonal. For any \( x \in \mathbb{R}^m \) we have
\[ x^T A x = \sum_{1=1}^{m} x_1 (A x)_1 \]
\[ = \sum_{1=1}^{m} x_1 (N_1 \sum_{j=1}^{m} N_j) \]
\[ = (\sum_{1=1}^{m} x_1 N_1) \sum_{j=1}^{m} N_j \]
\[ = \| \sum_{1=1}^{m} x_1 N_1 \|^2 \]
\[ \geq 0 \]

with equality holding if and only if \( x = 0 \). The matrix \( A \) is hence positive definite and (1.13) can be solved by Gauss elimination without pivoting, or, better still, by Cholesky decomposition.

We note also that

\[ \| g \| = \alpha^T A \alpha = \alpha^T d. \]

The entry \( A_{1j} \), the integral of the product of two piecewise linear polynomials, can be computed exactly by Simpson's rule applied on each subinterval \([t_k, t_{k+1}]\). Denoting \( \Delta t_k = t_{k+1} - t_k \) and \( z_k \) the midpoint of the interval \([t_k, t_{k+1}]\) we have for \( i=1,2,\ldots,m \)

\[ A_{11} = \left[ t_{1+1} \right]^{t_{1+2}} N_1(t)^2 dt + \left[ t_{1+1} \right]^{t_{1+2}} N_1(t)^2 dt \]
\[ = (\Delta t_{1+1}/6)[N_1(t_1)^2 + 4N_1(z_1)^2 + N_1(t_{1+1})^2] \]
\[ + (\Delta t_{1+2}/6)[N_1(t_{1+1})^2 + 4N_1(z_{1+1})^2 + N_1(t_{1+2})^2] \]
\[ = (t_{1+2} - t_1)/3. \]
We also compute for $i=1,2,\ldots,m-1$

\[ A_{i,1+1} = A_{i+1,1} \]

\[ = \int_{t_{i+1}}^{t_{i+2}} N_{i} N_{i+1}(t) \, dt \]

\[ = (t_{i+2} - t_{i+1})/6. \]

The solution $g_{x}$, being a linear combination of linear B-splines, is piecewise linear (and continuous) with knots $t_{1}$. After integrating $g_{x}$ twice and applying the interpolation conditions, we obtain $f_{x}$ which is piecewise cubic (with knots $t_{1}$) with two continuous derivatives.

Define $\beta \in \mathbb{R}^{n}$ via

\[ \beta_{i} = \begin{cases} 0 & i=1 \\ \alpha_{i-1} & i=2,3,\ldots,n-1 \\ 0 & i=n \end{cases} \]

and $\Delta \beta = \beta_{i+1} - \beta_{i}$. On $[t_{1},t_{1+1}]$ $f_{x}$ is defined by a unique cubic polynomial $p_{x}$ and hence $f_{x}$ can be determined by specifying the numbers $p_{x_{1}}^{(j)}(t_{1})$ for $i=1,2,\ldots,n-1$ and $j=0,1,2,3$. Then

\[ f_{x}(t) = \frac{p_{x_{1}}^{(2)}(t_{1})}{0!} + \frac{p_{x_{1}}^{(3)}(t_{1})(t-t_{1})}{1!} \]

\[ + \frac{p_{x_{1}}^{(2)}(t_{1})(t-t_{1})^{2}}{2!} + \frac{p_{x_{1}}^{(3)}(t_{1})(t-t_{1})^{3}}{3!} \]

(1.14)

for $t \in [t_{1},t_{1+1}]$. Of course, (1.14) can be more efficiently evaluated by using nested multiplication.
The polynomial $p_{*1}$ solves the differential equation

$$p_{*1}^{(2)}(t) = \beta_1 + (\Delta \beta_1 / \Delta t_1)(t-t_1)$$

(1.15)
on the interval $[t_1, t_1+1]$ with boundary conditions $p_{*1}(t_1) = y_1$ and $p_{*1}(t_1+1) = y_1+1$. Therefore

$$p_{*1}(t) = \frac{1}{2}(t-t_1)^2 + \frac{\Delta \beta_1}{6\Delta t_1}(t-t_1)^3 + c_1(t-t_1) + e_1$$

(1.16)

where the constants $c_1$ and $e_1$ are evaluated as $e_1 = y_1$ and

$$c_1 = \frac{\Delta y_1}{\Delta t_1} - \left( \frac{\beta_{1+1} + \Delta \beta_1}{2} \right) \Delta t_1$$

(1.17)

with $\Delta y_1 = y_{1+1} - y_1$. From (1.17) we obtain

$$p_{*1}^{(0)}(t_1) = y_1$$

$$p_{*1}^{(1)}(t_1) = c_1$$

(1.18)

$$p_{*1}^{(2)}(t_1) = \beta_1$$

$$p_{*1}^{(3)}(t_1) = \Delta \beta_1 / \Delta t_1$$

where $c_1$ is given by (1.17). A complete FORTRAN program for computing the natural cubic spline interpolant is given in Appendix A.

Figure (1.3) displays the natural cubic spline interpolant that is in contrast to the polynomial interpolant of figure (1.1).

We complete this chapter by posing (and solving) a generalization of problem (1.1). For $k$ fixed satisfying $2 \leq k \leq n$, let $A(k)$ be the set of interpolants (to the data) which are in $L^2_{2}(k)[a,b]$. We
Figure (1.3): The Natural Cubic Spline Interpolant.
consider the problem

Find $f_\# \in A(k)$ such that $\|f_\#^{(k)}\|_2 \leq \|f^{(k)}\|_2$ for all $f \in A(k)$

(1.19)

Let $f$ be an element of $A(k)$. Since (1.3) is valid for $f$, we can integrate by parts again (assuming $k > 2$) to obtain

$$f(t) = q_2(t) + \int_a^t \frac{(t-s)^2}{2!} f^{(3)}(s) ds$$

(1.20)

where

$$q_2(t) = f(a) + f^{(1)}(a)(t-a) + \frac{f^{(2)}(a)}{2!} (t-a)^2.$$ 

In general, after integrating by parts $k-1$ times we obtain

$$f(t) = q_{k-1}(t) + \int_a^t \frac{(t-s)^{k-1}}{(k-1)!} f^{(k)}(s) ds$$

(1.21)

or

$$f(t) = q_{k-1}(t) + \int_a^t \frac{(t-s)^{k-1}}{(k-1)!} f^{(k)}(s) ds.$$ 

(1.22)

Now we take the (scaled) $k$-th divided difference of (1.22) to obtain

$$d_{1,k} = \int_a^b g(s)N_{1,k}(s) ds \quad 1=1,2,\ldots,n-1$$

(1.23)
where

\[ d_{1,k} = (k-1)'(t_{1+k}-t_1)[t_1, \ldots, t_{1+k}]f(\cdot), \]  

\[ N_{1,k}(s) = (t_{1+k}-t_1)[t_1, \ldots, t_{1+k}](-s)^{k-1}_+ \]  

(1.24)

(1.25)

(the normalized B-spline of order k), and \( g = f(2) \).

Let \( B(k) \) denote the set of elements (in \( L_2[a,b] \)) which satisfy (1.23). Then the solution \( f_\# \) to (1.20) is related to the solution to the problem

\begin{equation}
\text{Find } g_\# \in B(k) \text{ such that } \|g_\#(k)\|_2 \leq \|g(k)\|_2 \text{ for all } g \in B(k) \tag{1.26}
\end{equation}

via \( g_\# = f_\#(k) \). Furthermore, for some \( \alpha \in \mathbb{R}^{n-k} \) we have

\[ g_\# = \sum_{j=1}^{n-k} \alpha_j N_{j,k}. \]

The coefficients \( \alpha_1, \alpha_2, \ldots, \alpha_{n-k} \) are chosen to solve the linear system of \( n-k \) equations in \( n-k \) unknowns represented by the matrix equation

\[ A \alpha = d \text{ where } A \text{ is symmetric and positive definite with entries } \]

\[ A_{i,j} = (N_{1,k}, N_{j,k}). \]

Since \( g_\# \) is a linear combination of piecewise polynomials of order \( k \), \( f_\# \) will be a piecewise polynomial of order \( 2k \). We call \( f_\# \) the natural spline interpolant of order \( 2k \).
2. A Minimal Norm Interpolation Problem
in the $L_p[a,b]$ Spaces

For $p$ such that $1 < p \leq \infty$ we define the set

$$G(p) = \left\{ g \in L_p[a,b]: \int_a^b g(t)\phi_i(t)dt = \int_a^b g_0(t)\phi_i(t)dt \text{ for } i=1,2,\ldots,n \right\}$$

(2.1)

where $\{\phi_i\}_{i=1}^n$ is a set of elements in $L_q[a,b]$, $q$ is conjugate to $p$ ($p+q = pq$ if $p \neq \infty$ and $q=1$ if $p = \infty$), and $g_0$ is a fixed element of $L_p[a,b]$. Consider the problem

Find $g_* \in G(p)$ such that $\|g_*\|_p \leq \|g\|_p$ for all $g \in G(p)$. (2.2)

In chapter 1 we solved (2.2) for the special case $p=2$; finding from a linear variety in a Hilbert space the element of minimal norm. The Projection Theorem came in handy to characterize $g_*$ as well as to guarantee uniqueness. However, for $p \neq 2 L_p[a,b]$ does not have the orthogonality properties of a Hilbert space and hence, we cannot use the Projection Theorem to solve (2.2). Instead we solve (2.2) in this chapter by utilizing the Hahn-Banach theorem to characterize $g_*$. Uniqueness follows in the case $1 < p < \infty$ by the strict convexity of the norm. This chapter, modeled after [deB(2)], motivates the use of the Hahn-Banach theorem in chapter 5.

Let $\lambda$ be the linear functional defined on the subspace

$$S = \text{span}(\phi_1, \ldots, \phi_n)$$
Any element of $G(p)$ (including $g_0$) will serve as an extension of $\lambda$ to a bounded linear functional defined on all of $L_q[a,b]$. Hence,

$$\|\lambda\|_s \leq \|g\|_p \text{ for all } g \in G(p).$$

Conversely, any extension of $\lambda$ to a bounded linear functional defined on all of $L_q[a,b]$, being identical to $\lambda$ on $S$, is represented by an element of $G(p)$.

The Hahn-Banach theorem guarantees the existence of an element $\hat{g} \in G(p)$ such that

$$\int_a^b f(t)\hat{g}(t)dt \leq \|\lambda\|_s \cdot \|f\|_q \text{ for all } f \in L_q[a,b].$$

This implies that $\|\hat{g}\| \leq \|\lambda\|_s$ which, taken along with (2.4), gives us $\|\hat{g}\| = \|\lambda\|_s$ and, therefore, a solution to (2.2). Now we characterize $\hat{g}$.

Let $\sum_{l=1}^n \alpha_i \phi_i$ be an element such that

$$\|\sum_{l=1}^n \alpha_i \phi_i\|_q = 1 \quad \text{and} \quad \lambda(\sum_{l=1}^n \alpha_i \phi_i) = \|\lambda\|_s.$$

(This element is unique if $1 < p < \infty$ since the norm is strictly convex.) Then
\[ \| \hat{g} \|_p = \| \lambda \|_s \]

\[ = \lambda \left( \sum_{i=1}^{n} \alpha_i \phi_i \right) \]

\[ = \int_a^b \left( \sum_{i=1}^{n} \alpha_i \phi_i \right)(t) \hat{g}(t) \, dt \]

\[ \leq \| \sum_{i=1}^{n} \alpha_i \phi_i \|_q \cdot \| \hat{g} \|_p \]

\[ = \| \hat{g} \|_p. \]

Therefore, equality holds throughout and we have

\[ \int_a^b \left( \sum_{i=1}^{n} \alpha_i \phi_i \right)(t) \hat{g}(t) \, dt = \| \sum_{i=1}^{n} \alpha_i \phi_i \|_q \cdot \| \hat{g} \|_p. \]

Since \( \hat{g} \) and \( \sum_{i=1}^{n} \alpha_i \phi_i \) are aligned, we must have

\[ \hat{g}(t) = \| \lambda \|_s \cdot \| \sum_{i=1}^{n} \alpha_i \phi_i \|_q \cdot \text{signum} \left( \sum_{i=1}^{n} \alpha_i \phi_i \right)(t). \]

We close this chapter by stating the interpolation problem that goes along with solving (2.2). Let \( p \) be a number such that \( 1 < p < \infty \), let \( k \) be an integer such that \( k \geq 2 \), and let \( f_0 \in L^p_k[a,b] \). Define the sets

\[ F = \{ f \in L^p_k[a,b]: f(t_i) = f_0(t_i) \quad i=1,2,\ldots,n \} \]
and

\[ G: = \{ g \in L^p[a, b]; \int_a^b g(t)N_{a, k}(t)\, dt = d_{1, k}^{n-k}, \quad i=1, 2, \ldots, n-k \} \]

Then the problems

Find \( f_\star \in F \) such that \( \| f_\star^{(k)} \|_p \leq \| f^{(k)} \|_p \) for all \( f \in F \)

and

Find \( f_\star \in G \) such that \( \| g_\star \|_p \leq \| g \|_p \) for all \( g \in G \)

are equivalent and

\[ g_\star(t) = f_\star^{(k)}(t) = \left| \sum_{i=1}^{n-k} \signum \left( \sum_{i=1}^i \beta_{1, 1, k} \right) \right|^{q-1} \sum_{i=1}^{n-k} \beta_{1, 1, k}(t). \]
3. The Convex Spline Interpolant

The data \( \{(t_i, y_i)\}_{i=1}^{n} \) are called convex if the point \((t_{i_2}, y_{i_2})\) lies on or beneath the line joining the points \((t_{i_1}, y_{i_1})\) and \((t_{i_3}, y_{i_3})\) whenever \(1 \leq i_1 < i_2 < i_3 \leq n\). Equivalently,

\[
[t_{i_1}, t_{i_2}, t_{i_3}] f(\cdot) = 0
\]

(where \(f\) is any interpolant to the data) or

\[
d_i = \frac{y_{i+2} - y_{i+1}}{t_{i+2} - t_{i+1}} - \frac{y_{i+1} - y_i}{t_{i+1} - t_i} \geq 0
\]

for \(i = 1, 2, \ldots, m (= n-2)\).

In this chapter we address the problem of finding, for convex data, the smoothest convex interpolant; that is, the convex interpolant having second derivative of minimal norm over all smooth convex interpolants. The natural cubic spline interpolant, the smoothest of all interpolants, regrettably does not always preserve the convexity of the data. In chapter 1 we showed that \(f^*\), the natural cubic spline interpolant, has second derivative

\[
f^*_*(2) = \sum_{j=1}^{m} \alpha_j N_j
\]

where the coefficients \(\alpha_1, \alpha_2, \ldots, \alpha_m\) satisfy (1.13). If any \(\alpha_i\) is negative, then \(f^*\) is actually concave on a subset of \([a, b]\).

Let \(\{(t_i, y_i)\}_{i=1}^{n}\) denote convex data and let \(A\) denote the set of convex interpolants in \(L_2(2)[a, b]\). We assume that \(A\) is nonempty.
There are convex data for which A is empty. For example, let
\[ y_1 = |t_1| \text{ and } t_1 = -2, t_2 = -1, t_3 = 0, t_4 = 1, \text{ and } t_5 = 2. \]
The only convex interpolant is \( f(x) = |x| \), which is not in \( L_2^{(2)}[-2,2] \).

Using the Peano kernel theorem as we did in Chapter 1, we can show that if \( f \in A \) then \( T(f^{(2)}) = d \) where \( T : L_2[a,b] \rightarrow R^m \) is given by 
\[
(Tg)_1 = (g, N_1). \]
Hence if
\[
B = \{ g \in L_2[a,b] : g \geq 0 \text{ and } Tg = d \},
\]
then problems

Find \( f_1 \in A \) such that \( \| f_1^{(2)} \|_2 \leq \| f^{(2)} \|_2 \) for all \( f \in A \) (3.1)

and

Find \( g_1 \in B \) such that \( \| g_1 \|_2 \leq \| g \|_2 \) for all \( g \in B \) (3.2)

are equivalent and the solutions are related via \( g_1 = f_1^{(2)} \). Since \( B \)

is a nonempty closed convex set, we consider (3.2) as finding the
distance from a point to a closed convex set in a Hilbert space.

**Proposition** ([L, page 69]): Let \( x \) be an element of a Hilbert space \( H \)
and let \( K \) be a nonempty closed convex subset of \( H \). Then there exists
a unique element \( k_o \in K \) such that 
\[
\| x - k_o \| \leq \| x - k \| \quad \text{for all } k \in K
\]
Furthermore, \( k_o \) is characterized by
\[
(x - k_o, k - k_o) \leq 0 \quad \text{for all } k \in K.
\]

Since we wish to find the element of minimal norm in \( B \), \( x \) corre-
sponds to the zero function and hence \( g_* \) is characterized by
\[
(g_*, g - g_*) \geq 0 \quad \text{for all } g \in B. \quad (3.3)
\]
Proposition ([MSSW, proposition 2.1]): If there exist coefficients $\alpha_1, \alpha_2, \ldots, \alpha_m$ satisfying

$$\begin{align*}
\int_a^b \left( \sum_{j=1}^m \alpha_j N_j \right) N(t) \, dt = d_i \quad i=1,2,\ldots,m,
\end{align*}$$

then $g^*_m = \left( \sum_{j=1}^m \alpha_j N_j \right)_+$. Furthermore, such coefficients exist if there exists $\hat{g} \in B$ such that $\{N_i\}_{i=1}^m$ are linearly independent over the support of $\hat{g}$.

Proof: Assume $\alpha_1, \alpha_2, \ldots, \alpha_m$ satisfy (3.4). Denote $s = \sum_{j=1}^m \alpha_j N_j$ and assume $g \in B$. Define $(h)_- = (-h)_+$ so that $h = (h)_+ - (h)_-$.

Then

$$((s)_+, g-(s)_+)$$

$$= (s + (s)_-, g - (s)_+)$$

$$= (s, g - (s)_+) + ((s)_-, g - (s)_+)$$

$$= ((s)_-, g) - ((s)_-, (s)_+)$$

$$= ((s)_-, g) \geq 0.$$ 

The last inequality is valid since both $(s)_-$ and $g$ are nonnegative functions. Hence $(s)_+$ satisfies (3.3).

We now show that we can find coefficients $\alpha_1, \alpha_2, \ldots, \alpha_m$ so that (3.4) holds by following the procedure employed in [MSSW].
We begin by considering the problem

\[
\inf_{a} \left\{ \sum_{j=1}^{m} \left( \frac{\alpha_{j}}{N_{j}} \right)^{2} (t) \frac{d}{dt} : \sum_{j=1}^{m} d_{j} = 1 \right\} \quad (3.5)
\]

and showing that if the infimum is attained at some \( \alpha \), then for some positive constant \( C \) the coefficients \( \alpha_{1}, \alpha_{2}, \ldots, \alpha_{m} \) satisfy (3.4).

If the infimum of (3.5) is attained at \( \alpha^{*} \), then \( \alpha^{*} \) is a critical point of the Lagrangian

\[
L(\alpha, \lambda) = \int_{a}^{b} \left( \sum_{j=1}^{m} \alpha_{j} N_{j} \right) \frac{d}{dt} (t) dt + \lambda \left( \sum_{j=1}^{m} \alpha_{j} N_{j} \right). \quad (3.6)
\]

At a minimum of \( L \) we must have

\[
0 = 2 \sum_{j=1}^{m} \alpha_{j} N_{j} \frac{d}{dt} (t) dt - \lambda d_{j} \quad i=1,2,\ldots,m \quad (3.7)
\]

and \( \alpha \cdot d = 1 \) for some \( \lambda \).

Now multiply (3.7) by \( \alpha_{i} \) and sum over \( i=1,2,\ldots,m \) to obtain

\[
2 \sum_{j=1}^{m} \left( \sum_{j=1}^{m} \alpha_{j} N_{j} \right) \left( \sum_{j=1}^{m} \alpha_{i} N_{j} \right)(t) dt - \lambda \sum_{j=1}^{m} \alpha_{i} d_{j} = 0
\]

or

\[
\lambda = 2 \sum_{j=1}^{m} \left( \sum_{j=1}^{m} \alpha_{j} N_{j} \right)^{2}(t) dt \geq 0. \quad (3.8)
\]

If \( \lambda > 0 \), then (3.7) reveals that
\[
\int_a^b \left( \sum_{j=1}^m \alpha_j N_j \right) g(t) \, dt = d \\
\alpha_j \geq 0, \quad j = 1, 2, \ldots, m
\]

(3.9)

where \( \alpha_j^* = \frac{\alpha_j}{\lambda} \). If \( \lambda = 0 \), then (3.8) reveals that

\[
\int_a^b \left( \sum_{j=1}^m \alpha_j N_j \right) g(t) \, dt = 0
\]

where \( \alpha \cdot d = 1 \). This implies that \( \sum_{j=1}^m \alpha_j N_j \leq 0 \). However, for any \( g \in B \) we have

\[
\int_a^b \left( \sum_{j=1}^m \alpha_j N_j \right) g(t) \, dt = \sum_{j=1}^m \alpha_j (N_j, g)
\]

\[
= \sum_{j=1}^m \alpha_j d_j
\]

\[
= 1
\]

which is impossible because \( g \) is nonnegative on \([a, b]\). We conclude that \( \lambda \) is strictly positive and, if the infimum in (3.5) is attained by some \( \alpha \), that (3.4) is solvable. We now show that the infimum is attained.

Let \( \{\alpha^{(k)}\}_{k=1}^\infty \) be a minimizing sequence. If \( \{\|\alpha^{(k)}\|_\infty\}_{k=1}^\infty \) is unbounded, then divide the objective function of (3.6) by \( \|\alpha\|_2 \) and
the constraint by \( \|\alpha\|_\infty \). There then exists \( \alpha \) such that
\[ \| \alpha \|_\alpha = 1, \]
\[ \alpha \cdot d = 0, \text{ and} \]
\[ \int_a^b \left( \sum_{j=1}^m \alpha_j N_j \right)^2 \delta(t) dt = 0. \]

We conclude that \( \sum_{j=1}^m \alpha_j N_j \) is nonpositive, but not identically zero. Since we have assumed there exists \( \hat{g} \in B \) such that the B-splines are linearly independent on the support of \( \hat{g} \),

\[ 0 = \sum_{j=1}^m \alpha_j d_j = \sum_{j=1}^m \alpha_j (\hat{g}, N_j) \]
\[ = (\hat{g}, \sum_{j=1}^m \alpha_j N_j) \]
\[ < 0 \]
which is a contradiction. Hence a minimizing sequence must be bounded and the infimum is attained via a convergent subsequence. This completes the proof of the proposition.

We note that the existence of \( \hat{g} \in B \), such that \( \{N_1\}_{j=1}^m \) are linearly independent over the support of \( \hat{g} \), in the previous proposition is guaranteed if \( d_j > 0 \) for each \( j \). Then each \( g \in B \) must be positive on some subinterval of \([t_1, t_{i+2}]\), the support of \( N_j \), for each \( j \).

Now we consider the implication of allowing \( d_k = 0 \) for some \( k \). As a specific example let \( t_1 = (1-1) \) for \( i=1,2,3,4 \) and let \( d = (1,0)^T \). If \( g_* \) is the positive part of a linear combination of B-splines, then there must exist numbers \( \alpha_1 \) and \( \alpha_2 \) satisfying
\[\int_0^3 (\alpha_1 N_1 + \alpha_2 N_2) N_1(t) dt = \begin{cases} 1 & \text{if } i=1 \\ 0 & \text{if } i=2 \end{cases} \quad (3.10)\]

which implies that \(\alpha_2 = -\infty\). This is equivalent to the solution being identically zero on \([1,3]\). In fact, any \(g \in B\) must be of the form \(g = g \chi_{[0,1]}\). It is shown in [MSSW, theorem 3.1] that the solution to (3.2) is

\[g_\# = \left( \sum_{j=1}^m \alpha_j N_j \right) \chi_\Gamma\]

for appropriate coefficients \(\alpha_1, \alpha_2, \ldots, \alpha_m\) where

\[\Gamma: = [a,b]/\{ \bigcup_{j=1}^m (t_j, t_{j+2}) : d_j = 0 \}.\]

Hence the solution to (3.2) with \(t_1 = (1-1)\) for \(i=1,2,3,4\) and \(d = (1,0)^T\) is

\[g_\# = 3N_1 \chi_{[0,1]}\]

Unless otherwise stated we assume \(d_1 > 0\) for each \(i\) for the remainder of this chapter.

Before we consider how to compute the coefficients \(\alpha_1, \alpha_2, \ldots, \alpha_m\) which satisfy (3.4), we give a procedure for integrating \(g_\#\).

Define \(\beta_1, \Delta \beta_1, \Delta t_1,\) and \(\Delta y_1\) as in chapter 1. We integrate \(g_\#\) on each subinterval \([t_i, t_{i+1}]\) separately, forming a piecewise polynomial, by solving the differential equation

\[p_{\#}(2)(t) = (\beta_1 + \frac{\Delta \beta_1}{\Delta t_1}(t-t_1))_+ \quad (3.11)\]

for \(t_1 \leq t \leq t_{i+1}\) with boundary conditions \(p_{\#}(t_1) = y_1\) and \(p_{\#}(t_{i+1}) = y_{i+1}\).
Two integrations gives us

\[ p_{x_1}(t) = \frac{\Delta t_{1}}{2\Delta \beta_{1}} \left( \beta_{1} + \frac{\Delta \beta_{1}}{\Delta t_{1}} (t-t_{1}) \right)^{2} + c_{1} \quad (3.12) \]

and

\[ p_{x_1}(t) = \frac{\Delta t_{1}}{6(\Delta \beta_{1})^{2}} (\beta_{1} + \Delta \beta_{1} (t-t_{1}))^{3} + c_{1} (t-t_{1}) + e_{1} \quad (3.13) \]

for constants \( c_{1} \) and \( e_{1} \). We proceed by cases.

Case 1 occurs when both \( \beta_{1} \) and \( \beta_{1+1} \) are nonnegative. The nonnegativity constraint is not active in this case and so (3.13) is equivalent to (1.16), although with modified constants \( c_{1} \) and \( e_{1} \). The values \( p_{x_1}(j)(t_{1}) \) for \( j = 0,1,2,3 \), are given by (1.18).

Case 2 occurs when \( \beta_{1} < 0 \) and \( \beta_{1+1} > 0 \). In this case \( p_{x_1} \) can be defined by two polynomials: a linear polynomial \( q_{11} \) defined on \([t_1, \tau_1] \) - where the nonnegativity constraint is active and hence the second derivative is zero - and a cubic polynomial defined on \([\tau_1, t_{1+1}] \) where

\[ \tau_1 = t_{1} - \beta_{1} \Delta t_{1} / \Delta \beta_{1} \quad (3.14) \]

Applying the boundary condition \( p_{x_1}(t_{1}) = y_{1} \) we obtain \( e_{1} = y_{1} \).

Applying \( p_{x_1}(t_{1+1}) = y_{1+1} \) we get an equation for \( c_{1} \):

\[ \frac{(\Delta t_{1})^{2}}{6(\Delta \beta_{1})^{2}} (\beta_{1+1})^{3} + c_{1} \Delta t_{1} + y_{1} = y_{1+1}. \]

Solving for \( c_{1} \) we have
\[ c_1 = \frac{\Delta y_1}{\Delta t_1} - \frac{(\beta_{1+1})^3 \Delta t_1}{2(\Delta \beta_1)^2} \]  

From (3.11), (3.12), and (3.13) we obtain

\[ q_{11}^{(1)}(t_1) = y_1 \]
\[ q_{11}^{(1)}(t_1) = c_1 \]
\[ q_{11}^{(2)}(t_1) = 0 \]
\[ q_{11}^{(3)}(t_1) = 0 \]

Case 3 occurs when \( \beta_1 > 0 \) and \( \beta_{1+1} < 0 \). In this case \( p_{n+1} \) is defined by a cubic polynomial \( q_{11} \) on \([t_1, \tau_1]\) and by a linear polynomial \( q_{12} \) on \([\tau_1, t_{1+1}]\) with \( \tau_1 \) defined by (3.14). These polynomials are determined by the values...
\[ q_{11}(t_1) = y_1 \]

\[ q^{(1)}_{11}(t_1) = c_1 + (\beta_1)^2 \Delta t_1 / (2\Delta \beta_1) \]

\[ q^{(2)}_{11}(t_1) = \beta_1 \]

(3.17)

\[ q^{(3)}_{11}(t_1) = \Delta \beta_1 / \Delta t_1 \]

\[ q_{12}(t_1) = c_1 (t_1 - t_1) + e_1 \]

\[ q^{(1)}_{12}(t_1) = c_1 \]

\[ q^{(2)}_{12}(t_1) = 0 \]

\[ q^{(3)}_{12}(t_1) = 0 \]

where \( c_1 \) and \( e_1 \) are given by

\[ c_1 = \frac{\Delta y_1}{\Delta t_1} - \frac{(\beta_1)^3 \Delta t_1}{2(\Delta \beta_1)^2} \]

and

\[ e_1 = y_1 - \frac{(\beta_1)^3 (\Delta t_1)^2}{6(\Delta \beta_1)^2} \]

Case 4 occurs when \( \beta_1 \) and \( \beta_{1+1} \) are both nonpositive. In this case we obtain a linear polynomial defined on \([t_1, t_{1+1}]\) and determined by
\[ p_{g_1}(t_1) = y_1 \]
\[ p_{g_1}^{(1)}(t_1) = \Delta y_1 / \Delta t_1 \]  
(3.18)
\[ p_{g_1}^{(2)}(t_1) = 0 \]
\[ p_{g_1}^{(3)}(t_1) = 0. \]

Since \( g_1 \) is piecewise linear and continuous (with knots at the \( t_i \)'s and \( r_i \)'s), \( f_1 \) will be piecewise cubic with two continuous derivatives (if \( d_1 > 0 \) for each \( i \)). We call \( f_1 \) the convex cubic spline interpolant.

Now we turn our attention to the task of numerically calculating the coefficients \( \alpha_1, \alpha_2, \ldots, \alpha_m \) which satisfy (3.4). We continue to assume that \( d_1 > 0 \) for each \( i \). Define \( F: \mathbb{R}^m \to \mathbb{R}^m \) by \( F = (F_1, F_2, \ldots, F_m)^T \) where

\[
F_1(\alpha) = \int_a^b \left( \sum_{j=1}^m \alpha_j N_j(t) \right) N_{j+1}(t) \, dt \quad i=1,2,\ldots,m. \quad (3.19)
\]

We wish to solve \( F(x) = d \).

One method is to use Jacobi iteration. An initial guess \( x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_m^{(0)})^T \) is chosen and a sequence \( \{x^{(k)}\}_{k=0}^{\infty} \) is generated by calculating \( x^{(k+1)} \), once \( x^{(k)} \) is known, by solving

\[
F_1(x_1^{(k)}, \ldots, x_{i-1}^{(k)}, x_i^{(k+1)}, x_{i+1}^{(k)}, \ldots, x_m^{(k)}) = d_i
\]
for $x_i^{(k+1)}$ for each $i$. A modification, the Gauss-Seidel iteration, involves calculating $x_i^{(k+1)}$, once $x_i^{(k)}$ is known, by solving

$$F_i(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i^{(k)}, x_{i+1}^{(k)}, \ldots, x_m^{(k)}) = d_i$$

for $x_i^{(k+1)}$ for $i=1,2,\ldots,m$ in succession. Both Jacobi and Gauss-Seidel iterations converge globally as proved in [IMS].

Now we consider Newton's method to solve $G(x) = F(x) - d = 0$. We pick a suitable initial guess $x^{(0)}$ and form a sequence $\{x^{(k)}\}_{k=0}^{\infty}$ by solving

$$\nabla G(x^{(k)})(x^{(k+1)} - x^{(k)}) = -G(x^{(k)})$$

for $x^{(k+1)}$ once $x^{(k)}$ is known. Since $\nabla G = \nabla F$, we can express (3.20) alternately as

$$\nabla F(x^{(k)})(x^{(k+1)} - x^{(k)}) = d - F(x^{(k)}).$$

The entries of the Jacobian matrix $\nabla F$ are

$$(\nabla F)_{ij}(\alpha) = \int_a^b (\sum_{k=1}^m \alpha_k N_k^0 N_{ij}^j(t) dt$$

where $(\sum_{k=1}^m \alpha_k N_k^0)$ is the characteristic function for the support of $N_k^0$.

We see that $\nabla F$ is symmetric and tridiagonal at each $\alpha$.

We now characterize those $\alpha$ for which $(\nabla F)(\alpha)$ is positive definite.

**Lemma (3.1):** The Jacobian $(\nabla F)(\alpha)$ is positive definite if and only if
\((\sum_{k=1}^{m} \alpha_{k} N_{k})^+\) does not vanish identically on any of the subintervals 

\([t_{l}, t_{l+2}]\) for \(l=1,2,\ldots,m\).

**Proof:** For any \(x \in k^{m}\) we have

\[
\sum_{l=1}^{m} x_{l} \sum_{j=1}^{m} (VF)_{l1}(\alpha)x_{j} = \int_{a}^{b} \left( \sum_{k=1}^{m} \alpha_{k} N_{k} \right)^{O} (\sum_{l=1}^{m} \alpha_{l} N_{l}) (t) dt \\
= \int_{a}^{b} \left( \sum_{k=1}^{m} \alpha_{k} N_{k} \right)^{O} (\sum_{l=1}^{m} \alpha_{l} N_{l})^{2} (t) dt
\]

\[\geq 0\]

If \((\sum_{k=1}^{m} \alpha_{k} N_{k})^+\) does not vanish identically on \([t_{l}, t_{l+2}]\) for each \(l\), then equality holds if and only if \(x_{l} = 0\) for each \(l\). If there exists some \(k\) such that \((\sum_{k=1}^{m} \alpha_{k} N_{k})^+\) is identically zero on \([t_{k}, t_{k+2}]\), then equality does hold for the nonzero vector \(x\) defined by \(x_{l} = \delta_{lk}\) for each \(l\). This completes the proof of the lemma.

From (3.20) we see that

\[
F_{l}(\alpha) = \sum_{j=1}^{m} \alpha_{j} \int_{a}^{b} \left( \sum_{k=1}^{m} \alpha_{k} N_{k} \right)^{O} N_{l} (t) dt \\
= \sum_{j=1}^{m} \alpha_{j} (VF)_{l1}(\alpha)
\]
so that \( F(a) = (\nabla F)(a) a \). Newton’s method – equation (3.22) - takes the form

\[
(\nabla F)(x^{(k)})x^{(k+1)} = d. \tag{3.23}
\]

**Theorem (3.2):** If \((\nabla F)(x^{(k)})\) is positive definite, then \((\nabla F)(x^{(k+1)})\) is positive definite for each \(k\) and, hence, Newton’s method – equation \(3.23\) - is always well-defined.

**Proof:** Having the known values \(x^{(k)}\), we wish to determine the values \(x^{(k+1)}\) satisfying

\[
\begin{align*}
\int_{S(k)} \left( \sum_{j=1}^{m} x^{(k+1)}_{j} N_{j}(t) \right) N_{1}(t) dt &= d_{1}, \quad 1, 2, \ldots, m \tag{3.24}
\end{align*}
\]

where \(S(k)\) is the support of \(\left( \sum_{j=1}^{m} x^{(k)}_{j} N_{j} \right)_{+}\). Since \((\nabla F)(x^{(k)})\) is positive definite, then \(S(k) \cup [t_{1}, t_{1+2}]\) contains an interval for each \(1\).

Since \(d_{1} > 0\), then \(\left( \sum_{j=1}^{m} x^{(k+1)}_{j} N_{j} \right)_{+}\) is positive on some subinterval of \([t_{1}, t_{1+2}]\). Hence, \((\nabla F)(x^{(k+1)})\) is positive definite. This completes the proof of the Theorem.

Note that if \(x^{(0)}\) has all positive components (for example, if \(x^{(0)}_{1} = 1\) for each \(1\), then \(S(0) = [a, b]\) and \(\sum_{j=1}^{m} x^{(1)}_{j} N_{j}\) is the second derivative of the natural cubic spline interpolant.

Now we assume that \(d_{k} = 0\) for some \(k\). In this case special care must be exercised since \(\{x^{(j)}_{k}\}_{j=0}^{\infty}\) may diverge to \(-\infty\), preventing any numerical convergence. We already know that \(d_{k} = 0\) implies that the
data points \((t_k, y_k), (t_{k+1}, y_{k+1}),\) and \((t_{k+2}, y_{k+2})\) are collinear and, hence, any convex interpolant must be linear on \([t_k, t_{k+2}]\). Equivalently, the second derivative of any convex interpolant must be zero on \([t_k, t_{k+2}]\). Hence \(g_k^*\) is of the form

\[
\sum_{j=1}^{m} \left( \sum_{j} \frac{N_j}{N_j(a,t_k)} \right) + \left( \sum_{j} \frac{N_j}{N_j(t_k+2,b)} \right).
\]

Since the value of \(x_k\) is immaterial and the \(k\)-th equation is automatically satisfied, the number of equations and unknowns each reduce by one. For computational convenience (3.23) can still be used with the following modifications: \((\nabla F)_k k = 1, (\nabla F)_{k, k+1} = 0,\) and \((\nabla F)_{k, k-1} = 0.\)

If \(d_k = 0,\) then the solution is discontinuous at \(t_k\) if \(x_{k-1} > 0\) and is discontinuous at \(t_{k+2}\) if \(x_{k+1} > 0.\) If the solution is discontinuous, then \(f^*_k\) will have only one continuous derivative.

A further problem is encountered when \(d_{k-1} = d_{k+1} = 0,\) but \(d_k \neq 0\) for some \(k.\) Any nonnegative function \(g\) which satisfies the \((k-1)\)-st and \((k+1)\)-st equations can not satisfy the \(k\)-th equation since \(g\) is identically zero on \([t_{k-1}, t_{k+1}]\) and on \([t_{k+1}, t_{k+2}]\). We conclude that there does not exist any convex interpolant in \(L_2^{(2)}[a,b]\) (and no solution to the problem as posed). However, we can find a convex interpolant whose second derivative is of the form
satisfying all but the k-th equation. We already know that this convex interpolant must be linear on \([t_{k-1}, t_{k+1}]\) and on \([t_{k+1}, t_{k+3}]\) and, hence, piecewise linear on \([t_{k-1}, t_{k+3}]\). If \(d_k\) is nonzero, then there will be a discontinuity in slope at \(t_{k+1}\). For the convenience of utilizing (3.23) we can set \(d_k\) to be zero to satisfy the k-th equation. The discontinuity in slope will show up after we integrate the solution to obtain the interpolant.

Figure (3.1) displays the natural cubic spline interpolant to the function

\[
f(t) = \frac{1}{(0.05+t)(1.05-t)}
\]

at the knots \(t_1 = 0, t_2 = 0.1, t_3 = 0.4, t_4 = 0.7, t_5 = 0.8,\) and \(t_6 = 1.0\). Figure (3.2) displays the convex spline interpolant to this function. Table (3.1) shows the convergence results for Jacobi, Gauss-Seidel, and Newton's method iterations taken from [IMS]. Note the quadratic convergence characteristic of Newton's method. These convergence results are typical.
Figure (3.1): The Natural Cubic Spline Interpolant.

Figure (3.2): The Convex Cubic Spline Interpolant
<table>
<thead>
<tr>
<th>Iteration Number</th>
<th>Jacobi</th>
<th>Gauss Seidel</th>
<th>Newton</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.46 x 10^2</td>
<td>.27 x 10^2</td>
<td>.19 x 10^2</td>
</tr>
<tr>
<td>2</td>
<td>.28 x 10^2</td>
<td>.11 x 10^2</td>
<td>.85 x 10^1</td>
</tr>
<tr>
<td>3</td>
<td>.75 x 10^1</td>
<td>.42 x 10^1</td>
<td>.29 x 10^1</td>
</tr>
<tr>
<td>4</td>
<td>.12 x 10^2</td>
<td>.18 x 10^1</td>
<td>.49 x 10^0</td>
</tr>
<tr>
<td>5</td>
<td>.26 x 10^1</td>
<td>.75 x 10^0</td>
<td>.14 x 10^{-1}</td>
</tr>
<tr>
<td>6</td>
<td>.49 x 10^1</td>
<td>.31 x 10^0</td>
<td>.11 x 10^{-4}</td>
</tr>
<tr>
<td>7</td>
<td>.10 x 10^1</td>
<td>.13 x 10^0</td>
<td>.71 x 10^{-11}</td>
</tr>
<tr>
<td>8</td>
<td>.21 x 10^1</td>
<td>.55 x 10^{-1}</td>
<td>.49 x 10^{-12}</td>
</tr>
<tr>
<td>9</td>
<td>.43 x 10^0</td>
<td>.23 x 10^{-1}</td>
<td>-----</td>
</tr>
<tr>
<td>10</td>
<td>.86 x 10^0</td>
<td>.96 x 10^{-2}</td>
<td>-----</td>
</tr>
<tr>
<td>20</td>
<td>.11 x 10^{-1}</td>
<td>.16 x 10^{-5}</td>
<td>-----</td>
</tr>
<tr>
<td>30</td>
<td>.14 x 10^{-3}</td>
<td>.26 x 10^{-9}</td>
<td>-----</td>
</tr>
<tr>
<td>40</td>
<td>.18 x 10^{-5}</td>
<td>.58 x 10^{-13}</td>
<td>-----</td>
</tr>
<tr>
<td>50</td>
<td>.24 x 10^{-7}</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>60</td>
<td>.30 x 10^{-9}</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>70</td>
<td>.39 x 10^{-11}</td>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>
4. The Shape-Preserving Spline Interpolant

We addressed in chapter 3 the problem of finding, for convex data, the smoothest convex interpolant. We begin this chapter by considering the problem of finding, for concave data, the smoothest concave interpolant. Then we continue the chapter by examining the problem of finding, for general data, the smoothest interpolant which is locally convex where the data are locally convex and is locally concave where the data are locally concave.

Let \{(t_i, y_i)\}_{i=1}^n denote concave data and let \(A\) denote the set of all concave interpolants in \(L_2(2)[a,b]\). Assume \(A\) is nonempty. Using the Peano kernel theorem as we did in chapter 1, we see that, if \(f \in A\), then

\[
\int_a^b f^{(2)}(t)N_1(t)\,dt = d, \quad l=1,2,\ldots,m(n-2)
\]

Equivalently, we have \(T(f^{(2)}) = d\).

Defining

\[
B = \{g \in L_2[a,b] : g \leq 0 \text{ and } Tg = d\},
\]

we conclude that the problems

Find \(f_* \in A\) such that \(\|f_*^{(2)}\|_2 \leq \|f^{(2)}\|_2\) for all \(f \in A\) (4.1)

(the problem of finding the smoothest concave interpolant) and

Find \(g_* \in B\) such that \(\|g_*\|_2 \leq \|g\|_2\) for all \(g \in B\)
are equivalent and the solutions are related via $g_\delta = f_\delta^{(2)}$.

Of course, the smoothest concave interpolant to the concave data $\{(t_1', y_1')\}_{1=1}^n$ is the negative of the smoothest convex interpolant to the convex data $\{(t_1, -y_1)\}_{1=1}^n$. We highlight this with the following proposition.

**Proposition [MSSW]:** If there exist coefficients $a_1, a_2, a_3, \ldots a_m$ satisfying
\[
\int_a^b \left( \sum_{j=1}^m a_j N_j \right) N_1(t) \, dt = d_1, \quad 1=1,2,\ldots,m \tag{4.3}
\]
then $g_\delta = - \left( \sum_{j=1}^m a_j N_j \right)$. Furthermore, such coefficients exist if there exists $\hat{g} \in B$ such that $\{N_1\}_{1=1}^m$ are linearly independent over the support of $\hat{g}$.

We note that the existence of $\hat{g} \in B$, such that $\{N_1\}_{1=1}^m$ are linearly independent over the support of $\hat{g}$, in the previous proposition is guaranteed if $d_1 < 0$ for each $1$. Then each $g \in B$ is negative on some subinterval of $[t_1', t_{1+2}]$, the support for $N_1$, for each $1$.

Now we consider the problem of finding, for general data, a smooth shape-preserving interpolant - a smooth interpolant which is locally convex where the data are locally convex and is locally concave where the data are locally concave. Assuming for the moment that $d_1$ is nonzero for each $1, we define the sets
and $$T_1 = \{ [t_1, t_{1+2}] : d_1 > 0 \} ,$$

$$T_2 = \{ [t_1, t_{1+2}] : d_1 > 0 \} ,$$

$$\Omega_1 = T_1 / T_2 ,$$

$$\Omega_2 = T_2 / T_1 ,$$

and $$\Omega_3 = [a, b] / (\Omega_1 \cup \Omega_2) .$$

Now we define the sets

$$A = \{ f \in L_2^{(2)}[a,b] : f^{(2)} x_{\Omega_1} \geq 0 , \ f^{(2)} x_{\Omega_2} \leq 0 , \ \text{and} \ f(t_1) = y_1 \ for \ i=1,2,\ldots,n \}$$

(which we assume is nonempty) and

$$B = \{ g \in L_2[a,b] : g x_{\Omega_1} \geq 0 , \ g x_{\Omega_2} \leq 0 , \ and \ Tg = d \} .$$

We conclude that the problems

Find $$f_* \in A$$ such that $$\| f_*^{(2)} \|_2 \leq \| f^{(2)} \|_2$$ for all $$f \in A \quad (4.4)$$

and

Find $$g_* \in B$$ such that $$\| g_* \|_2 \leq \| g \|_2$$ for all $$g \in B \quad (4.5)$$
arc equivalent and \( g^* = f^*_m(2) \).

The following proposition gives the solution to (4.5). We see that \( f^*_\Omega_1 \) has the character of the convex spline interpolant, \( f^*_\Omega_2 \) has the character of the concave spline interpolant, and \( f^*_\Omega_3 \) has the character of the natural spline interpolant.

**Proposition [MSSW]:** If there exists coefficients \( \alpha_1, \alpha_2, \ldots, \alpha_m \) satisfying

\[
\begin{aligned}
&\left\{ \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_1} + \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_2} \right. \\
&\left. + \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_3} \right\} N_1(t) dt = d_1 \quad 1 = 1, 2, \ldots, m
\end{aligned}
\]

(4.6)

then

\[
g^* = \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_1} - \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_2} + \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_3}.
\]

Furthermore, such coefficients exist if there exists \( \hat{g} \in B \) such that \( \{ N_1 \}_{1=1}^{m} \) are linearly independent over the support of \( \hat{g} \).

We note that the existence of \( \hat{g} \in B \), such that \( \{ N_1 \}_{1=1}^{m} \) are linearly independent over the support of \( \hat{g} \), in the previous proposition is guaranteed if \( d_1 \) is nonzero for each \( i \). Then each \( g \in B \) is nonzero on
some subinterval of \([t_i, t_{i+2}]\), the support of \(N_i\), for each \(i\).

We now solve (4.6). Define \(F : \mathbb{R}^m \to \mathbb{R}^m\) where \(F = (F_1, F_2, \ldots, F_m)^T\) is given

\[
F(x) = \begin{cases} 
\sum_{j=1}^{m} \int_{\Omega_1} (\sum_{j=1}^{m} x_j) N_j(t) dt \\
\sum_{j=1}^{m} \int_{\Omega_2} (\sum_{j=1}^{m} x_j) N_j(t) dt \\
\sum_{j=1}^{m} \int_{\Omega_3} (\sum_{j=1}^{m} x_j) N_j(t) dt \end{cases}
\]

\(1=1,2,\ldots,m\) (4.7)

We use Newton's method to solve \(F(x) = d\). Picking a suitable initial guess \(x^{(0)}\) we produce a sequence \{\(x^{(0)}, x^{(1)}, \ldots\)\} by solving

\[
(\nabla F)(x^{(k)})(x^{(k+1)} - x^{(k)}) = d - F(x^{(k)})
\]

(4.8)

for \(x^{(k+1)}\) once \(x^{(k)}\) is known. The Jacobian matrix has entries given by

\[
(\nabla F)_{ij}(x) = \int_a^b P(\alpha) N_j(t) N_i(t) dt
\]

where
From (4.9) we see that \( VF \) is symmetric and tridiagonal at each \( \alpha \).

We also note that

\[
\begin{align*}
P(\alpha) & = \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_1} + \left( \sum_{j=1}^{m} \alpha_j N_j \right) \chi_{\Omega_2} + \chi_{\Omega_3}. \\
& = \left( \sum_{j=1}^{m} N_j \right) \chi_{\Omega_1} + \left( \sum_{j=1}^{m} N_j \right) \chi_{\Omega_2} + \left( \sum_{j=1}^{m} N_j \right) \chi_{\Omega_3}
\end{align*}
\]  

so that \( F(\alpha) = (VF)(\alpha) \chi \) and, hence, (4.8) reduces to

\[
(VF)(\alpha^{(k)}) \chi^{(k+1)} = d. 
\]  

The following lemma (with proof similar to its counterpart in chapter 3) characterizes those \( \alpha \) for which \( (VF)(\alpha) \) is positive definite.

**Lemma (4.1):** The Jacobian \( (VF)(\alpha) \) is positive definite if and only if \( P(\alpha) \) does not vanish identically on any of the subintervals \([t_i, t_{i+2}]\) for \( i=1,2,\ldots,m \).
The following theorem is modeled after theorem (3.2).

**Theorem (4.2):** If \((VF)(x^{(0)})\) is positive definite, then Newton's method - equation (4.10) - is always well-defined.

Note that if \(x^{(0)}\) is given by \(x^{(0)}_i = \text{signum}(d_i)\) for each \(i\), then \(P(x^{(0)})\) is the characteristic function for the interval \([a,b]\) and \(\sum_{j=1}^{m} x_{j}^{(1)} N_{j}\) is the second derivative of the natural cubic spline interpolant.

If \(d_k = 0\) for some \(k\), then we already know that any shape-preserving interpolant must be linear on \([t_k, t_{k+2}]\). In fact any \(g \in B\) must satisfy

\[
g = g\{x[a, t_k] + x[t_{k+2}, b]\}.
\]

The solution in this case is of the form

\[
g_* = h\{x[a, t_k] + x[t_{k+2}, b]\}
\]

where

\[
h = \left( \sum_{j=1}^{m} \alpha_j N_j \right) x_{\Omega_1} - \left( \sum_{j=1}^{m} \alpha_j N_j \right) x_{\Omega_2} + \left( \sum_{j=1}^{m} \alpha_j N_j \right) x_{\Omega_3}.
\]
Since the value of $a_k$ is immaterial - the $k$-th equation $F_k(x) = d_k$ of (4.11) being automatically satisfied - the number of equations and unknowns reduce by one each. For computational convenience we can still use (4.10) by setting $(\nabla F)_{kk} = 1$, $(\nabla F)_{k,k+1} = 0$, and $(\nabla F)_{k,k-1} = 0$.

Once we solve $F(g) = d$ we proceed to integrate $g_*$ which is piecewise linear (but not necessarily continuous, even if $d_k$ is non-zero for each $k$) to obtain $f_*$ which is piecewise cubic. On the interval $[t_1, t_{1+1}]$, $f_*$ is given by the solution to the differential equation

$$p_1^{(2)}(t) = \beta_1 + \Delta \beta_1 / \Delta t_1 (t-t_1)$$

(4.12)

for $t_1 \leq t \leq t_{1+1}$ if $[t_1, t_{1+1}] \subset \Omega_3$, where

$$p_1^{(2)}(t) = (\beta_1 + (\Delta \beta_1 / \Delta t_1) (t-t_1))_{+}$$

(4.13)

for $t_1 \leq t \leq t_{1+1}$ if $[t_1, t_{1+1}] \subset \Omega_1$, or

$$p_1^{(2)}(t) = -(\beta_1 + (\Delta \beta_1 / \Delta t_1) (t-t_1))_{-}$$

(4.14)

for $t_1 \leq t \leq t_{1+1}$ if $[t_1, t_{1+1}] \subset \Omega_2$ with boundary conditions

$$p_1(t_1) = y_1 \quad \text{and} \quad p_1(t_{1+1}) = y_{1+1}.$$

The function $p_1$ is either a cubic polynomial or piecewise cubic
given by two polynomials $q_{11}$ and $q_{12}$ defined on separate subintervals
of $[t_1,t_{1+1}]$. The solution $p_1$ to (4.11) is given by (1.18). The
solution to (4.12) is, depending on signum ($\beta_1$) and signum ($\beta_{1+1}$),
given by (1.18), (3.16), (3.17), and (3.18). The solution to (4.13)
is determined by (1.18) if $\beta_1 \leq 0$ and $\beta_{1+1} \leq 0$, by (3.16) if $\beta_1 > 0$
and $\beta_{1+1} < 0$, by (3.17) if $\beta_1 < 0$ and $\beta_{1+1} > 0$, and by (3.18) if
$\beta_1 \geq 0$ and $\beta_{1+1} \geq 0$.

Figures (4.1), (4.3), (4.5), and (4.7) display the natural cubic
spline interpolants to the given data. Figures (4.2), (4.4), (4.6),
and (4.8) display the corresponding shape-preserving interpolants.
Tables (4.1), (4.2), (4.3), and (4.4) give convergence results for
Newton's method. Note the quadratic convergence characteristic of
Newton's method.

Appendix B lists a FORTRAN program for computing the shape-preser-
v ing cubic spline interpolant.
Figure (4.1): The Natural Cubic Spline Interpolant.

Figure (4.2): The Shape-Preserving Cubic Spline Interpolant.
Figure (4.3): The Natural Cubic Spline Interpolant.

Figure (4.4): The Shape-Preserving Cubic Spline Interpolant.
Figure (4.5): The Natural Cubic Spline Interpolant.

Figure (4.6): The Shape-Preserving Cubic Spline Interpolant.
Figure (4.7): The Natural Cubic Spline Interpolant.

Figure (4.8): The Shape-Preserving Cubic Spline Interpolant.
Table 4.1

<table>
<thead>
<tr>
<th>Iteration Number</th>
<th>$| F(x^{(n)}) - d |_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0.13 \times 10^1$</td>
</tr>
<tr>
<td>2</td>
<td>$0.67 \times 10^0$</td>
</tr>
<tr>
<td>3</td>
<td>$0.25 \times 10^0$</td>
</tr>
<tr>
<td>4</td>
<td>$0.42 \times 10^{-1}$</td>
</tr>
<tr>
<td>5</td>
<td>$0.12 \times 10^{-2}$</td>
</tr>
<tr>
<td>6</td>
<td>$0.88 \times 10^{-6}$</td>
</tr>
<tr>
<td>7</td>
<td>$0.58 \times 10^{-12}$</td>
</tr>
<tr>
<td>8</td>
<td>$0.64 \times 10^{-13}$</td>
</tr>
</tbody>
</table>
Table 4.2

<table>
<thead>
<tr>
<th>Iteration Number</th>
<th>$|F(x^{(n)}) - d|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0.12 \times 10^1$</td>
</tr>
<tr>
<td>2</td>
<td>$0.56 \times 10^0$</td>
</tr>
<tr>
<td>3</td>
<td>$0.121 \times 10^0$</td>
</tr>
<tr>
<td>4</td>
<td>$0.36 \times 10^{-1}$</td>
</tr>
<tr>
<td>5</td>
<td>$0.11 \times 10^{-2}$</td>
</tr>
<tr>
<td>6</td>
<td>$0.85 \times 10^{-6}$</td>
</tr>
<tr>
<td>7</td>
<td>$0.54 \times 10^{-12}$</td>
</tr>
<tr>
<td>8</td>
<td>$0.70 \times 10^{-13}$</td>
</tr>
<tr>
<td>Iteration Number</td>
<td>$| F(x^{(n)}) - d |_2$</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>1</td>
<td>0.24 x $10^1$</td>
</tr>
<tr>
<td>2</td>
<td>0.16 x $10^1$</td>
</tr>
<tr>
<td>3</td>
<td>0.12 x $10^1$</td>
</tr>
<tr>
<td>4</td>
<td>0.90 x $10^0$</td>
</tr>
<tr>
<td>5</td>
<td>0.53 x $10^0$</td>
</tr>
<tr>
<td>6</td>
<td>0.20 x $10^0$</td>
</tr>
<tr>
<td>7</td>
<td>0.26 x $10^{-1}$</td>
</tr>
<tr>
<td>8</td>
<td>0.42 x $10^{-3}$</td>
</tr>
<tr>
<td>9</td>
<td>0.97 x $10^{-7}$</td>
</tr>
<tr>
<td>10</td>
<td>0.37 x $10^{-12}$</td>
</tr>
<tr>
<td>11</td>
<td>0.21 x $10^{-12}$</td>
</tr>
</tbody>
</table>
Table 4.4

<table>
<thead>
<tr>
<th>Iteration Number</th>
<th>$|F(x^{(n)}) - d|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0.29 \times 10^1$</td>
</tr>
<tr>
<td>2</td>
<td>$0.13 \times 10^1$</td>
</tr>
<tr>
<td>3</td>
<td>$0.50 \times 10^0$</td>
</tr>
<tr>
<td>4</td>
<td>$0.11 \times 10^0$</td>
</tr>
<tr>
<td>5</td>
<td>$0.56 \times 10^{-2}$</td>
</tr>
<tr>
<td>6</td>
<td>$0.16 \times 10^{-4}$</td>
</tr>
<tr>
<td>7</td>
<td>$0.13 \times 10^{-9}$</td>
</tr>
<tr>
<td>8</td>
<td>$0.26 \times 10^{-12}$</td>
</tr>
</tbody>
</table>
5. Constrained Minimization in a Dual Space

Let $C$ be a convex cone in a normed dual space $X$ with predual $Y$.

Assume $y_1, y_2, \ldots, y_n$ are elements of $Y$ and define $T: X \rightarrow \mathbb{R}^n$ by

$$Tx = (x(y_1), x(y_2), \ldots, x(y_n))^T$$

Let $B = \{x \in C : Tx = d\}$ for a given vector $d$. Consider the problem

Find $x_* \in B$ such that $\|x_*\| \leq \|x\|$ for all $x \in B$ \hspace{1cm} (5.1)

of which (1.10), (3.2), and (4.5) are special cases. In this chapter we study existence and characterization of solutions to (5.1). The following lemma gives sufficient conditions for existence of a solution.

Lemma (5.1): If $B$ is nonempty, if $C$ is weak* closed, and if $Y$ is separable, then there exists a solution to problem (5.1).

Proof: Let $\gamma = \inf \{\|x\| : x \in C$ and $Tx = d\}$. Let $\{x_n\}$ be a sequence in $C$ such that

$$Tx_n = d \hspace{1cm} (5.2)$$

and

$$\|x_n\| \leq \gamma + 1/n \hspace{1cm} (5.3)$$
for each \( n \). Since \( Y \) is separable, by Alaoglu's theorem there exists a weak* convergent subsequence of \( \{x_n\} \) with weak* limit \( x \). Since \( C \)

is weak* closed we have \( x \in C \), from (5.2) we have \( Tx = d \), and from

(5.3) we have \( \|x\| \leq \gamma \) (and hence \( \|x\| = \gamma \)). This completes the

proof of the lemma.

Throughout this chapter we assume that \( B \) is nonempty, \( C \) is weak*

closed, and \( Y \) is separable. Since \( x_n = \theta \) if \( d = \theta \), we assume also

that \( d \neq \theta \). The following proposition gives us sufficient conditions

for \( C \) being weak* closed.

**Proposition (5.2):** If \( C \) is normed closed and if \( Y \) is a reflexive

space, then \( C \) is weak* closed.

**Proof:** Assume \( \{x_n\} \) is a sequence in \( C \) with weak* limit \( x \). We want

to show that \( x \) is in \( C \). We do this by contradiction. If \( x \) is not an

element of \( C \), then there exists an element \( y \) (an element of both the
dual and predual of \( X \)) which serves to separate \( x \) from \( C \) in the sense

that

\[
x_n(y) > K
\]

for each \( n \) and

\[
x(y) < K
\]

for some constant \( K \). This implies that

\[
\lim_{n \to \infty} x_n(y) \neq x(y)
\]
which is a contradiction. Therefore $x \in C$ and $C$ is weak* closed.

This completes the proof of the proposition.

For $\gamma > 0$ we define the convex set $G(\gamma) \subset \mathbb{R}^n$ by

$$G(\gamma) = \{Tx : x \in C \text{ and } \|x\| \leq \gamma\}.$$ 

We now show that $G(\gamma) = \gamma G(1)$ and $G(\gamma)$ is closed.

**Proposition (5.3):** For each $\gamma > 0$ we have $G(\gamma) = \gamma G(1)$.

**Proof:** By definition

$$G(\gamma) = \{Tx : x \in C \text{ and } \|x\| \leq \gamma\}$$

$$= \{Tx : \frac{x}{\gamma} \in C \text{ and } \|x/\gamma\| \leq 1\}$$

$$= \{T(x/\gamma) : \frac{x}{\gamma} \in C \text{ and } \|x/\gamma\| \leq 1\}$$

$$= \gamma\{Tw : w \in C \text{ and } \|w\| \leq 1\}$$

$$= \gamma G(1).$$

**Proposition (5.4):** The set $G(1)$ is closed.

**Proof:** Assume $\{z_n\}$ is a sequence in $G(1)$ which converges to $z$. We want to show that $z$ is an element of $G(1)$. Equivalently, we want to show that $x \in C$ exists such that $\|x\| \leq 1$ and $Tx = z$. 
For each n there exists $x_n \in C$ such that $\|x_n\| \leq 1$ and $T x_n = \gamma_n$.

By Alaoglu's theorem there exists a subsequence of $\|x_n\|$ which converges weak* to some $x \in C$. Hence $\|x\| \leq 1$ and $T x = \gamma$. This completes the proof of the proposition.

We define

$$
\gamma^* = \inf \{ \gamma : d \in G(\gamma) \}.
$$

(5.4)

Equivalently,

$$
\gamma^* = \inf \{ \gamma : \text{There exists } x \in C \text{ such that } T x = d \text{ and } \|x\| \leq \gamma \}
$$

$$
= \inf \{ \| x \| : x \in C \text{ and } T x = d \}.
$$

(5.5)

By lemma (5.1) we know that there exists $x_* \in C$ such that $\|x_*\| = \gamma^*$ and $T x_* = d$. We call $x_*$ an interpolant of minimal norm.

We now attempt to characterize $x_*$ via the Hahn-Banach theorem.

We begin by defining a functional $\rho : Y \to \mathbb{R}$ by

$$
\rho(y) = \sup \{ x(y) : x \in C \text{ and } \|x\| \leq 1 \}.
$$

Notice that if $C = X$ (the unconstrained problem), then $\rho$ is the norm on $Y$. In general, since we are taking the supremum over a subset of the closed unit ball $U$ in $X$, we have $\rho(y) \leq \|y\|$ for all $y \in Y$.

Since $\theta$ is an element of $C$, we have $\rho \geq 0$. In convex analysis $\rho$ is called the support functional of the convex set \{ $x \in C : \|x\| \leq 1$ \}. 

Since C is weak* closed, the supremum is attained at some element of \( \{ x \in C : \| x \| \leq 1 \} \); that is, for any \( y \in Y \) there exists an \( x \) (a function of \( y \)) such that \( x \in C, \| x \| \leq 1 \), and \( \rho(y) = x(y) \). In fact we have \( \| x \| = 1 \) unless \( x = 0 \). The following two propositions reveal that \( \rho \) is continuous, subadditive, and positive homogeneous.

**Lemma (5.5):** The functional \( \rho \) is continuous.

**Proof:** Assume \( y_1 \) and \( y_2 \) are elements of \( Y \) and define \( y = y_1 - y_2 \).

Let \( x \) be the element in \( \{ x \in C : \| x \| \leq 1 \} \) such that \( \rho(y_2) = x(y_2) \).

Since \( |x(y)| \leq \| y \| \), we have

\[
x(y_2) - \| y \| \leq x(y_2) + x(y)
\]

or

\[
x(y_2) - \| y \| \leq x(y_1).
\]

Therefore,

\[
\rho(y_2) - \| y \| \leq \rho(y_1).
\]

The elements \( y_1 \) and \( y_2 \) can be interchanged to obtain

\[
\rho(y_1) - \| y \| \leq \rho(y_2)
\]
and hence

\[ |\rho(y_1) - \rho(y_2)| \leq \|y_1 - y_2\|. \]

**Lemma (5.6):** The functional \( \rho \) is subadditive and positive homogeneous (hence convex).

**Proof:** Assume \( y_1 \) and \( y_2 \) are in \( Y \). To show that \( \rho \) is subadditive we must show that

\[ \rho(y_1 + y_2) \leq \rho(y_1) + \rho(y_2). \]

By definition

\[ \rho(y_1 + y_2) = \sup \{ x(y_1 + y_2) : x \in C \text{ and } \|x\| \leq 1 \} \]

\[ \leq \sup \{ x(y_1) : x \in C \text{ and } \|x\| \leq 1 \} \]

\[ + \sup \{ x(y_2) : x \in C \text{ and } \|x\| \leq 1 \} \]

\[ = \rho(y_1) + \rho(y_2). \]

Now assume \( \alpha > 0 \) and \( y \in Y \). To show that \( \rho \) is positive homogeneous we must show that

\[ \rho(\alpha y) = \alpha \rho(y). \]
By definition

\[
\rho(\alpha y) = \sup \{ x(\alpha y) : x \in C \text{ and } \|x\| \leq 1 \}
\]

\[
= \alpha \cdot \sup \{ x(y) : x \in C \text{ and } \|x\| \leq 1 \}
\]

\[
= \alpha \rho(y).
\]

This completes the proof of the lemma.

As an example we compute \(\rho\) for the case \(C = \{x \in L_p[a,b] : x \geq 0\}\) where \(1 < p < \infty\). For an arbitrary element \(g \in L_q[a,b]\), the predual of \(L_p[a,b]\) where \(p + q = pq\), we have for any \(f \in C\) with \(\|f\|_p \leq 1\) by the Minkowski inequality

\[
\int_a^b f(t) g(t) dt \leq \int_a^b f(t) g_+(t) dt
\]

\[
\leq \|f\|_p \cdot \|g_+\|_q
\]

\[
\leq \|g_+\|_q.
\]
Assuming $g_+ \neq 0$, let

$$f = (g)_+^{q-1} / \| (g)_+^{q-1} \|_p.$$ 

Then we have $f \in C$, $\|f\|_p = 1$, and

$$\int_a^b f(t)g(t)dt = \|g_+\|.$$ 

Hence

$$\rho(g) = \sup \left\{ \int_a^b f(t)g(t)dt : f \in C \text{ and } \|f\|_p \leq 1 \right\} = \|g_+\|_q.$$ 

If $g_+ = 0$, then $\rho(g) = 0$.

**Lemma (5.7):** For all $a \in \mathbb{R}^n$ we have

$$\sum_{i=1}^n a_i d_i \leq \gamma^* \rho \left( \sum_{i=1}^n a_i y_i \right). \quad (5.6)$$

**Proof:** Since $\gamma^* = \inf \{ \gamma : d \in G(\gamma) \}$, we have $d \in G(\gamma^* + \varepsilon)$ for any $\varepsilon > 0$. Hence for every positive integer $n$ there exists $x_m \in C$
such that \( Tx_m = d \) and \( \|x_m\| \leq \gamma + 1/m \). Therefore, for any \( \alpha \in \mathbb{R}^n \)

\[
\begin{align*}
\sum_{i=1}^{n} \alpha_i d_i & = \sum_{i=1}^{n} \alpha_i y_m(i) \\
& = x_m \left( \sum_{i=1}^{n} \alpha_i y_i \right) \\
& \leq \|x_m\| \rho \left( \sum_{i=1}^{n} \alpha_i y_i \right) \\
& \leq (\gamma + 1/m) \rho \left( \sum_{i=1}^{n} \alpha_i y_i \right).
\end{align*}
\]

Now let \( m \to \infty \) to obtain (5.6). This completes the proof of the lemma.

Since we know that \( G(\gamma) \) is closed from proposition (5.4), we could have used \( x_0 \) in place of \( x_m \) in the proof of lemma (5.7). The next lemma states that there exists a nonzero vector \( \beta \in \mathbb{R}^n \) such that equality holds in (5.6).

**Proposition (5.8):** There exists a vector \( \beta \in \mathbb{R}^n \) such that \( \|\beta\| = 1 \) and

\[
\beta \cdot d = \gamma \left( \sum_{i=1}^{n} \beta_i y_i \right),
\]

(5.7)

**Proof:** The vector \( d \) is an element of \( G(\gamma) \), but not an element of \( G(\gamma - \varepsilon) \) for any \( \varepsilon > 0 \). Hence the closed convex set \( G(\gamma - \varepsilon) \) and the
vector $d$ can be strictly separated by a hyperplane. This implies the existence of a nonzero vector $\beta(\varepsilon)$ such that

$$\beta(\varepsilon) \cdot y < \beta(\varepsilon) \cdot d$$

for all $y \in G(y^* - \varepsilon)$ and without loss of generality we may assume that $\|\beta(\varepsilon)\| = 1$. Equivalently, we have

$$\beta(\varepsilon) \cdot Tx < \beta(\varepsilon) \cdot d$$

and by the linearity of $T$

$$\sum_{i=1}^{n} x(\sum_{1}^{n} \beta_1(\varepsilon)y_1) < \beta(\varepsilon) \cdot d$$

for all $x \in C$ such that $\|x\| \leq y^* - \varepsilon$. Hence we obtain

$$(y^* - \varepsilon)\rho(\sum_{1}^{n} \beta_1(\varepsilon)y_1) < \beta(\varepsilon) \cdot d$$

We can take the limit as $\varepsilon \to 0$ to obtain a vector $\bar{\beta}$ such that $\|\bar{\beta}\| = 1$ and

$$y^* \rho(\sum_{1}^{n} \beta_1 y_1) \leq \bar{\beta} \cdot d.$$
We have the reverse inequality from lemma (5.7) and therefore

$$\beta \cdot d = \gamma \rho (\Sigma \beta y_1).$$

This completes the proof of the lemma.

Let $\lambda$ be a linear functional defined on the subspace

$$S: = \text{span}(y_1, y_2, \ldots, y_n)$$

by

$$\lambda (\Sigma \alpha y_1) = \Sigma \alpha d$$

so that (5.6) can now be written

$$\lambda(y) \leq \gamma \rho (y) \text{ for all } y \in S.$$

The Hahn-Banach theorem states that there exists an element $w$

in $X$ such that

$$w(y) = \lambda(y) \text{ for all } y \in S \quad (5.8)$$

and

$$w(y) \leq \gamma \rho (y) \text{ for all } y \in Y. \quad (5.9)$$
**Theorem (5.9):** The Hahn-Banach extension \( w \) is an interpolant of minimal norm.

**Proof:** From (5.8) we see that \( Tw = d \) so that \( w \) interpolates the data. To complete the proof we show that \( w \in C \) and \( \|w\| = \gamma^* \).

We show that \( w \) is in \( C \) by contradiction. Assume \( w \) is not an element of \( C \). Since \( C \) is weak* closed, there exists an element \( y_0 \) in \( Y \) which strictly separates \( w \) from \( C \) in the sense that

\[
    w(y_0) > x(y_0) \quad \text{for all } x \in C. \tag{5.10}
\]

Since \( C \) is a cone we have \( \lambda x \in C \) whenever \( \lambda > 0 \) and \( x \in C \). Hence (5.10) implies

\[
    0 \geq x(y_0) \quad \text{for all } x \in C \tag{5.11}
\]

(or \( \rho(y_0) = 0 \)) and

\[
    w(y_0) > 0. \tag{5.12}
\]

However, from (5.9) and (5.12) we have

\[
    0 < w(y_0) \leq \gamma^* \rho(y_0) = 0
\]

which is a contradiction. Hence \( w \) must be an element of \( C \).
Lastly, we show that $\|w\| = \gamma^*$. We already know that

$$\gamma^* \leq \|w\|$$

(5.13)

since $w \in B (w \in C$ and $Tw = d)$. Because $\rho$ is bounded above by the norm on $Y$, (5.9) yields

$$w(y) \leq \gamma^* \|y\| \quad \text{for all } y \in Y$$

and hence

$$\|w\| \leq \gamma^*.$$ \hspace{1cm} (5.14)

Taken together, (5.13) and (5.14) imply that $\|w\| = \gamma^*$. This completes the proof of the theorem.

Recall that for a given element $y_o$ in $Y$ there exists an element $x_o$ (a function of $y_o$) in $C$ such that $\rho(y_o) = x(y_o)$. Furthermore, either $\|x_o\| = 1$ or $x_o$ is the zero element. The following lemma will lead us to the conclusion that, if $\rho$ is differentiable at $y_o$, then $\rho'(y_o) = x_o$.

\textbf{Lemma (5.10):} Let $f$ be a functional defined on a normed linear space $Z$. If $f$ is differentiable at $x_o \in Z$ and if there exists a linear functional $\lambda$ such that
\[ f(z_o) + \lambda(z - z_o) \leq f(z) \tag{5.15} \]

for all \( z \) in some neighborhood of \( z_o \), then \( \lambda = (\forall f)(z_o) \).

**Proof:** Let \( z = z_o + tu \) where \( t > 0 \) and \( u \in Z \). Inequality (5.15) yields

\[
\lambda(u) \leq \frac{f(z_o + tu) - f(z_o)}{t} \tag{5.16}
\]

Since (5.16) holds for all \( t > 0 \) (and sufficiently small) and for all \( u \in Z \), we have \( \lambda \leq (\forall f)(z_o) \). Substituting \(-u\) for \( u \) in (5.16) yields

\[
\lambda(u) \geq \frac{f(z_o - tu) - f(z_o)}{t} \tag{5.17}
\]

for all \( t > 0 \) (and sufficiently small) and for all \( u \in Z \). Taken together, (5.16) and (5.17) imply \( \lambda = (\forall f)(z_o) \).

**Corollary (5.11):** If \( \rho \) is differentiable at \( y_o \in Y \), then \( \rho'(y_o) = x_o \).

**Proof:** Since \( \rho(y_o) = x_o(y_o) \) and \( x_o(y) \leq \rho(y) \) for all \( y \in Y \), we have

\[
\rho(y_o) + x_o(y - y_o) \leq \rho(y)
\]
for all $y \in Y$. By the previous lemma we have $p'(y_0) = x_0$. This completes the proof of the corollary.

Inequality (5.6) motivates the problem

$$\inf_{a} \{ p\left( \sum_{1}^{n} a_{1} y_{1} \right) : a \cdot d = 1 \}. \quad (5.18)$$

Notice that if $a$ is any vector satisfying $a \cdot d = 1$ and if $x$ is any element of $B$, then

$$1 = \sum_{1}^{n} a_{1} d_{1} = x\left( \sum_{1}^{n} a_{1} y_{1} \right)$$

and hence

$$\leq \|x\| p\left( \sum_{1}^{n} a_{1} y_{1} \right)$$

This implies that the infimum is positive (and, in fact, is bounded below by $(\gamma^*)^{-1}$. If the infimum is attained at some $a^* \in \mathbb{R}^n$ and if $p$

is differentiable at $\sum_{1}^{n} a_{1} y_{1}$, then we are led to a solution to (5.1) as
the next theorem reveals.

**Theorem (5.12):** If there exists \( \alpha^* \in \mathbb{R}^n \) such that \( \alpha^* \cdot d = 1 \) and

\[
\rho(\sum_{1=1}^{n} \alpha^* y_1) = \inf_{\alpha} \left\{ \rho(\sum_{1=1}^{n} \alpha y_1) : \alpha \cdot d = 1 \right\}
\]

and if \( \rho \) is differentiable at \( \sum_{1=1}^{n} \alpha^* y_1 \), then

\[
\gamma \rho'(\sum_{1=1}^{n} \alpha y_1)
\]

is an interpolant of minimal norm.

**Proof:** Problem (5.18) has Lagrangian

\[
L(\alpha, \lambda) = \rho(\sum_{1=1}^{n} \alpha y_1) - \lambda(\sum_{1=1}^{n} \alpha d - 1).
\]

If there exists a solution \( \alpha^* \) to (5.18), then there exists \( \lambda^* \) so that \( (\alpha^*, \lambda^*) \) is a stationary point of (5.19). Hence

\[
x(y_1) - \lambda^* d_1 = 0 \quad 1=1,2,\ldots,n
\]
where \( x = \rho'(\sum_{1}^{n} \alpha_{1}y_{1}) \), \( x \in \mathbb{C} \), \( \|x\| = 1 \), and \( \alpha \cdot 1 = 1 \).

We first show that \( \lambda > 0 \). Multiply (5.20) by \( \alpha_{1}^{*} \) and sum over \( 1 \) to obtain

\[
x(\sum_{1}^{n} \alpha_{1}^{*}y_{1}) = \lambda \sum_{1}^{n} \alpha_{d_{1}} = \lambda.
\]

Since \( x = \rho'(\sum_{1}^{n} \alpha_{1}y_{1}) \), we have

\[
x(\sum_{1}^{n} \alpha_{1}y_{1}) = \rho(\sum_{1}^{n} \alpha_{1}y_{1})
\]

so that

\[
\lambda = \rho(\sum_{1}^{n} \alpha_{1}y_{1}) \geq 0
\]
Actually, we know that since the infimum is positive, we have $\lambda^* > 0$.

We can also show this by contradiction. If $\lambda^* = 0$, then

$$x(\sum_{1=1}^{n} a^*_i y_i) \leq 0 \quad \text{for all } x \in C. \quad \text{(5.21)}$$

Let $s$ be any interpolant in $C$. (We know that there exists an interpolant in $C$ since $B$ is nonempty.) Then

$$s(\sum_{1=1}^{n} a^*_i y_i) = \sum_{1=1}^{n} a^*_i d_i = 1$$

which contradicts (5.21). Therefore, $\lambda^* > 0$.

Now we show that $\lambda^* \gamma^* = 1$. From (5.20) we see that $x/\lambda^*$ is an interpolant in $C$. Hence

$$\gamma^* \leq \|x\| / \lambda^* = 1/\lambda^*$$

or

$$\gamma^* \lambda^* \leq 1 \quad \text{(5.22)}$$

Let $w$ be an interpolant of minimal norm satisfying (5.9). Then
Equivalently, we have

\[ w(\sum_{i=1}^{n} \alpha_i y_i) \leq \gamma^* \rho(\sum_{i=1}^{n} \alpha_i y_i). \]

which leads to

\[ 1 \leq \gamma^* \lambda^*. \]

(5.23)

Taken together, (5.22) and (5.23) imply

\[ 1 = \gamma^* \lambda^*. \]

This concludes the proof of the theorem.

We consider now the problem of determining when the infimum is attained in (5.18). From proposition (5.8) we know that there exist a nonzero vector \( \beta \) such that

\[ 0 \leq \beta \cdot d = \gamma^* \rho(\sum_{i=1}^{n} \beta_i y_i). \]
If $\beta \cdot d > 0$, then the infimum is attained in (5.18) at $a^* = \beta / (\beta \cdot d)$.

**Proposition (5.13):** If $d$ is in the relative interior of 

\[ S = \{ r : r \in G(\gamma) \text{ for some } \gamma \}, \]

then there exists a vector $\beta$ such that

\[ 1 = \beta \cdot d = \gamma^* \left( \sum_{i=1}^{n} \beta_i y_i \right). \]

**Proof:** We prove by contradiction. Assume that every vector $\beta$ which satisfies

\[ \beta \cdot d = \gamma^* \left( \sum_{i=1}^{n} \beta_i y_i \right) \]

also satisfies $\beta \cdot d = 0$. Without loss of generality it can be assumed that there exists a nonzero vector $\beta$ such that

\[ 0 = \beta \cdot d = \gamma^* \left( \sum_{i=1}^{n} \beta_i y_i \right) \]

and

\[ \beta \cdot y \geq 0 \text{ for all } y \in G(\gamma^*). \]
In any relative neighborhood of \( d \) there is a vector \( z \) such that \( \beta \cdot z < 0 \). If \( z \) were an element of \( S \), then there would be an element \( r \) in \( G(\gamma^*) \) such that \( z = \alpha r \) for some \( \alpha > 0 \). However, we would then have

\[
\beta \cdot z = \alpha \beta \cdot r \geq 0
\]

which is a contradiction. Therefore \( z \) is not an element of \( S \) and \( d \) is not in the relative interior of \( S \). This completes the proof of the proposition.
References


Appendix A

A Program for Constructing the Natural Cubic Spline Interpolant to Given Data.
PROGRAM UNCON (INPUT, OUTPUT, TAPE5=INPUT, TAPE6=OUTPUT)

WE FORM THE NATURAL CUBIC SPLINE INTERPOLANT.

INTEGER N, M, J
REAL T(50), F(50), D(50), X(50), A(50), PP(4, 50)
REAL AA(50), BB(50), CC(50)

THE ARRAYS (T) AND (F) - EACH OF SIZE M, THE NUMBER
OF DATA POINTS - CONTAIN THE COMPONENTS OF THE DATA.
THE DATA FILE IS OF THE FOLLOWING FORM

WHERE WE ASSUME (T) HAS STRICTLY INCREASING COMPONENTS.

READ (3, *) M
READ (3, *) (T(I), F(I), I=1, M)
N = M - 2

THE ARRAY (D) CONSISTS OF THE SCALED
SECOND DIVIDED DIFFERENCES.

DO 100 I = 1, N
D(I) = ( F(I+2) - F(I+1) ) / ( T(I+2) - T(I+1) )
C = ( F(I+1) - F(I) ) / ( T(I+1) - T(I) )
CONTINUE

THE SECOND DERIVATIVE OF THE NATURAL CUBIC SPLINE
INTERPOLANT IS A LINEAR COMBINATION OF LINEAR B-SPLINES.
WE CALCULATE THE COEFFICIENTS.

AA(I) = 0.0
BB(I) = (T(3) - T(1))/3.0
CC(I) = (T(2) - T(2))/6.0
DO 200 I = 2, N - 1
AA(I) = (T(I+1) - T(I))/6.0
BB(I) = (T(I+2) - T(I))/3.0
CC(I) = (T(I+2) - T(I+1))/6.0
CONTINUE

200 AA(N) = (T(N+1) - T(N))/6.0
BB(N) = (T(N+2) - T(N))/3.0
CC(N) = 0.0
CALL TRID(AA, BB, CC, D, N)

A(1) = 0.0
A(M) = 0.0
DO 300 I = 2, N + 1
A(I) = D(I - 1)
CONTINUE

DO 400 K = 1, N + 1
IF = F(K + 1) - F(K)
DT = T(K + 1) - T(K)
DA = A(K + 1) - A(K)
PP(4, K) = DA / DT
PP(3, K) = A(K)
PP(2, K) = IF / DT - (A(K) / 2 + DA / 6) * DT
PP(1, K) = F(K)
CONTINUE

PP(4, M) = 0.0
PP(3, M) = 0.0
PP(2, M) = 0.0
PP(1, M) = F(M)

DO 500 h = 1, M
WRITE(6, 450) K, T(K), (PP(I, K), I = 1, 4)
CONTINUE

WE CREATE A DATA FILE FOR PLOTTING THE (JDER)-TH DERIVATIVE OF THE NATURAL CUBIC SPLINE INTERPOLANT BY EVALUATING IT AT (MM) EQUALLY SPACED POINTS, INCLUDING THE ENDPOINTS. WE ASSUME THAT (JDER) HAS VALUE 0, 1, 2, OR 3.

JDER = 0
MM = 201
CALL DATAFL(1, PP, M, MM, JDER)

STOP
SUBROUTINE DATAFL(TX,PP,LI,MM,JDER)

WE CREATE A DATA FILE FOR PLOTTING THE (JDER)-TH DERIVATIVE OF THE PIECEWISE CUBIC POLYNOMIAL. WE ASSUME (JDER) HAS VALUE 0, 1, 2, OR 3.

INTEGER LI,MM,JDER

REAL TX(100),PP(4,100)

LEFT= 1

MMONE= MM - 1

WRITE(4,*), MM

XE= (TX(LI)-TX(1))/FLOAT(MMONE)

DO 500 IP=1,MM

XT= TX(1) + XE*FLOAT(IP-1)

WE FIND THE INTERVAL IN WHICH THE POINT (XT) LIES.

IF ( LEFT .NE. LI ) THEN

DO 200 IS=LEFT,LI-1

IF ( XT .LT. TX(IS+1) ) GO TO 300

CONTINUE

CONTINUE

END IF

LEFT= IS

WE NOW COMPUTE THE VALUE OF THE POLYNOMIAL AT THE POINT (XT) BY USING MESTED MULTIPLICATION.

H= XT - TX(LEFT)

FAC= 4.0 - FLOAT(JDER)

YT= 0.0

DO 400 M=4,JDER+1,-1

YT= (YT/FAC)*H + PP(M,LEFT)

FAC= FAC - 1.0

CONTINUE

WRITE(4,450) XT,YT

FORMAT(F8.4,E18.9)

CONTINUE

RETURN

END
SUBROUTINE TRN(SUB, DIAG, SUP, B, N)
INTEGER N, I
REAL N, DIAG(N), SUB(N), SUP(N)
IF (N.LE.1) THEN
  B(I) = B(I)/DIAG(I)
  RETURN
END IF
DO 111 I = 2, N
  SUB(I) = SUB(I)/DIAG(I-1)
  DIAG(I) = DIAG(I) - SUB(I)*SUP(I-1)
  B(I) = B(I) - SUB(I)*B(I-1)
  CONTINUE
B(N) = B(N)/DIAG(N)
DO 222 I = N-1, 1, -1
  B(I) = (B(I) - SUP(I)*B(I+1))/DIAG(I)
  CONTINUE
RETURN
END
Appendix B

A Program for Constructing the Shape-Preserving Cubic Spline Interpolant to Given Data
PROGRAM MAIN, INPUT, OUTPUT, TAPE5=INPUT, TAPE6=OUTPUT)

WE COMPUTE A SHAPE-PRESERVING INTERPOLANT TO GIVEN DATA.

NOTE ON THE SIZE OF THE ARRAYS:

THE ARRAYS (T), (F), AND (A) MUST BE OF LENGTH AT LEAST M, THE NUMBER OF DATA POINTS. THE
ARRAY (TX) AND THE SECOND COMPONENT OF THE
ARRAY (PP) SHOULD BE OF LENGTH 2M. THE ARRAYS
(X), (Y), AND (D) MUST BE OF LENGTH AT LEAST M-2.
THE ARRAY (ID) MUST BE OF LENGTH AT LEAST M-1.

REAL T(50), F(50), X(50), Y(50), A(50)
REAL TX(100), PP(4, 100), TL, TF, cPS
INTEGER M, N, ITMAX, I, J, IFLAG, MM
COMMON D(50), ID(50)

THE ARRAYS (T) AND (F) - EACH OF SIZE M, THE NUMBER
OF DATA POINTS - CONTAIN THE COMPONENTS OF THE DATA.
THE DATA FILE IS OF THE FOLLOWING FORM

M
T(1), F(1)
T(2), F(2)
.
.
.
T(M), F(M)

WHERE WE ASSUME (T) HAS STRICTLY INCREASING COMPONENTS.

READ (*, *) M
READ(3, * ) (T(I), F(I), I=1, M)
N= M-2

ITMAX IS A (POSITIVE NUMBER USED TO TEST FOR
CONVERGENCE IN NEWTON'S METHOD - SUBROUTINE (ZERO).
ITMAX IS THE MAXIMUM NUMBER OF ITERATIONS
WHICH W. PERMIT FOR NEWTON'S METHOD TO CONVERGE.

EPS = 1.0E-8
ITMAX = 20
THE ARRAY (X) IS THE KNOT SEQUENCE (T) WITH THE
ENDPOINTS TL AND TR DELETED.
DO 120 I=1,N
X(I)= T(I+1)
CONTINUE

THE ARRAY (D) CONSISTS OF THE SCALED
SECOND DIVIDED DIFFERENCES.

IT IS IMPORTANT THAT WE IDENTIFY DIVIDED DIFFERENCES
WHICH ARE ZERO. THIS MEANS THAT WE MUST COMPARE TWO
FLOATING-POINT NUMBERS. TO DO THIS WE ASSUME D(K) IS
ZERO IF D(K) IS SMALL.

XEPS= 1.0
DO 130 J=1,20
XEPS= XEPS/10.
Z= 1.0 + XEPS
IF ( Z .EQ. 1.0 ) GO TO 135
YEPS= XEPS
CONTINUE
CONTINUE
YEPS= YEPS*1000.

DO 140 K=1,N
D(K)= ( F(K+2)-F(K+1) )/ ( T(K+2)-T(K+1) )
C = ( F(K+1)-F(K) )/ ( T(K+1)-T(K) )
IF ( ABS(D(K)) .LE. YEPS ) D(K)= 0.0
CONTINUE

THE INITIAL GUESS (Y) FOR NEWTON’S METHOD
WILL YIELD THE SECOND DERIVATIVE OF THE
NATURAL SPLINE SOLUTION, EXCEPT POSSIBLY
WHEN D(K)= 0.0 FOR SOME K.

DO 145 K=1,N
IF ( D(K) .GT. 0.0 ) THEN
Y(K)= 1.0
ELSE
Y(K)= -1.0
END IF
CONTINUE

WRITE(6,150)

WRITE(6,150)
FORMAT(6,160) (I(I), I=1,N)
FORMAT(5X,4E12.6)
WRITE(6,170)
FORMAT(//)

ID(K)= 1 INDICATES THAT THE INTERPOLATING FUNCTION
IS CONSTRAINED TO BE CONVEX ON [T(K),T(K+1)]
AND, HENCE, ITS SECOND DERIVATIVE IS CONSTRAINED
TO BE NONNEGATIVE ON THIS INTERVAL.

ID(K)= -1 INDICATES THAT THE INTERPOLATING FUNCTION
IS CONSTRAINED TO BE CONCAVE ON [T(K),T(K+1)]
AND, HENCE, ITS SECOND DERIVATIVE IS CONSTRAINED
TO BE NONPOSITIVE ON THIS INTERVAL.

ID(K)= 0 INDICATES THAT THE INTERPOLATING FUNCTION
IS UNCONSTRAINED ON [T(K),T(K+1)].

DO 180 I=1,N-1
180 ID(I+1)= 0
IF (D(I).GE.0.0 .AND. D(I+1).GE.0.0) ID(I+1)= 1
IF (D(I).LE.0.0 .AND. D(I+1).LE.0.0) ID(I+1)= -1
CONTINUE
IF (D(N).GE.0.0) THEN
ID(N+1)= 1
ELSE
ID(N+1)= -1
END IF

IF A NONZERO DATA VALUE D(I) LIES BETWEEN TWO
ZERO DATA VALUES D(I-1) AND D(I+1), THEN D(I)
IS TAKEN TO BE ZERO FOR COMPUTATIONAL PURPOSES.

DO 185 I=2,N-1
185 IF (D(I-1).EQ.0.0 .AND. D(I+1).EQ.0.0) D(I)= 0.0
CONTINUE

SUBROUTINE ZERO CALCULATES THE PIECEWISE
LINEAR SECOND DERIVATIVE OF THE SHAPE-
PRESERVING INTERPOLANT.

CALL ZERO(Y,X,N,ITMAX,EPS,IFLAG,TL,TR)
00153C  
00154  A(1) = 0.0
00155  A(M) = 0.0
00156  DO 190 I = 2, N + 1
00157  A(I) = Y(I-1)
00158  CONTINUE
00159C
00160  WRITE(6,200)
00161  200  FORMAT(/,' PIECEWISE LINEAR 2ND DERIVATIVE ' ,/)
00162  WRITE(6,210) (T(I),A(I), I = 1, M)
00163  210  FORMAT(SX,IS,' ( ',F14.6,' , ',F14.6,' ) ' )
00164  WRITE(6,220)
00165  220  FORMAT(/)
00166C
00167C  SUBROUTINE (POLY) INTEGRATES THE RESULT
00168C  FROM SUBROUTINE (ZERO).
00169C
00170  CALL POLY(A,T,PP,M,F,LI,TX)
00171C
00172  WRITE(6,230)
00173  230  FORMAT(/,' KNOTS AND COEFFICIENTS OF PIECEWISE CUBIC ' ,/)
00174  DO 250 I = 1, LI
00175  WRITE(6,240) I,TX(I),(PP(J,I), J = 1, 4)
00176  240  FORMAT(SX,IS,5F14.6)
00177  250  CONTINUE
00178  WRITE(6,260) IFLAG
00179  260  FORMAT(/,' ERROR CODE = ',15,/)  
00180  WRITE(6,270) ITMX
00181  270  FORMAT(/, ' NUMBER OF ITERATIONS = ',15,/)  
00182C
00183C  SUBROUTINE (DATAFL) IS USED TO CREATE A
00184C  DATA FILE FOR PLOTTING. WE EVALUATE THE
00185C  (JDER)-TH DERIVATIVE OF THE PIECEWISE CUBIC
00186C  POLYNOMIAL AT MM EQUALLY SPACED POINTS,
00187C  INCLUDING THE ENDPOINTS TL AND TR. WE
00188C  ASSUME (JDER) HAS VALUE 0, 1, 2, OR 3.
00189C
00190  MM = 201
00191  JDER = 0
00192  CALL DATAFL(TX,PP,LI,MM,JDER)
00193C
00194  STOP
00195  END
SUBROUTINE ZERO(A,X,N,ITMAX, EPS, IFLAG, TL, TR)

INTEGER N, ITMAX, L, LJ, L, IFLAG
REAL A(N), X(N), FXI50), AL, XL, AR, XR, DT, DA, T, W
REAL SUB(50), DIAG(50), SUP(50), H(50), SUM1, SUM2
REAL RATIO, GLEFT, GRIGH, EPS, FNORM1, TL, TR
COMMON D(50), ID(50)

INPUT PARAMETERS:
A...INITIAL ESTIMATE FOR NEWTON'S METHOD.
X...KNOT SEQUENCE WITH THE ENDPOINTS DELETED.
N...THE SIZE OF THE ARRAY (A): THE NUMBER OF UNKNOWNS.
ITMAX...MAXIMUM NUMBER OF ITERATIONS FOR NEWTON'S METHOD.
EPS...PARAMETER USED TO TEST FOR CONVERGENCE.
TL, TR...LEFT- AND RIGHT-ENDPOINTS OF THE INTERVAL RESPECTIVELY.

OUTPUT PARAMETERS:
A...THE CALCULATED ZERO IF CONVERGENCE OCCURRED.
ITMAX...NUMBER OF ITERATIONS REQUIRED FOR NEWTON'S METHOD TO CONVERGE.
IFLAG...IFLAG=1: CONVERGENCE INDICATED BY COMPARING THE L1 NORMS OF THE ITERATES
        IFLAG=2: NUMBER OF ITERATIONS EXCEEDED ITMAX.

PRINT 100
FORMAT (' ITERATION NUMBER AND RESIDUAL: ',/)
C ' QUADRATIC CONVERGENCE IS EXPECTED,'/)
DO 350 LJ=1, ITMAX

THE ARRAYS (SUB), (DIAG), AND (SUP) CONTAIN THE ELEMENTS OF THE TRIDIAGONAL POSITIVE-DEFINITE JACOBIAN MATRIX (J), EVALUATED AT THE VECTOR (A).
IT SHOULD BE NOTED THAT THE MATRIX EQUATION SOLVER, THE SUBROUTINE (TRIN), DOES NOT TAKE ADVANTAGE OF THE SYMMETRY OF (J). HENCE (SUB) AND (SUP) ARE BOTH NECESSARY. ALTHOUGH SUB(K)=SUP(K-1), EQUATIONS FOR BOTH ARRAYS ARE WRITTEN OUT IN FULL.
1= D(K)=0.0 FOR SOME K, THEN THE NUMBER
OF UNKNOWNS (AND EQUATIONS) REDUCE, IN ORDER
TO PERMIT THE COMPUTATION OF ONE JACOBIAN
MATRIX THE PROGRAM SETS SUB(k)=SUP(k-1)=0.0
AND DIAG(k)=1.0.

DO 125 k=1,N

IF (K.EQ.1) THEN
  AL = 0.0
  XL = TL
ELSE
  AL = A(K-1)
  XL = X(K-1)
END IF

IF (K.EQ.N) THEN
  AR = 0.0
  XR = TR
ELSE
  AR = A(K+1)
  XR = X(K+1)
END IF

IF (AL.GE.0.0 .AND. A(K).GE.0.0 ) J1 = 1
IF (AL.LT.0.0 .AND. A(K).GE.0.0 ) J1 = 2
IF (AL.GE.0.0 .AND. A(K).LT.0.0 ) J1 = 3
IF (AL.LE.0.0 .AND. A(K).LE.0.0 ) J1 = 4

IF (A(K).GE.0.0 .AND. AR.GE.0.0 ) J2 = 1
IF (A(K).LT.0.0 .AND. AR.GE.0.0 ) J2 = 2
IF (A(K).GE.0.0 .AND. AR.LT.0.0 ) J2 = 3
IF (A(K).LE.0.0 .AND. AR.LE.0.0 ) J2 = 4

DT = X(K)-XL
DA = A(K)-AL

IF (ID(K) .EQ. 1) THEN
  IF (H.NE.1) THEN
    IF (J1.EQ.1) THEN
      SUB(k)= DT/6.0
    ELSE IF (J1.EQ.2) THEN
      T = XL-(DT/DA)*AL
      W = 0.5*(X(K)+T)
SUB(K) = (X(K) - T) / 6.0 * ((T-XL)/DT) * ((X(K)-T)/DT)

C = 4.0 * ((W-XL)/DT) * ((X(K)-W)/DT)

GLEFT = (X(K) - T) / 6.0 * ((T-XL)/DT)**2

C = 4.0 * ((W-XL)/DT)**2 + 1.0

ELSE IF (J1.EQ.3) THEN

T = XL - (DT/DA)*XL

W = 0.5*( T+XL )

SUB(K) = (T-XL) / 6.0 * ( 4.0* ((W-XL)/DT) * ((X(K)-W)/DT)

C = + ((T-XL)/DT) * ((X(K)-T)/DT)

GLEFT = (T-XL) / 6.0 * ( 4.0* ((W-XL)/DT)**2

C = + ((T-XL)/DT)**2

ELSE IF (J1.EQ.4) THEN

SUB(K) = 0.0

GLEFT = 0.0

END IF

ELSE IF (K.EQ.1) THEN

SUB(I) = 0.0

GLEFT = 0.0

IF (J1.EQ.1) GLEFT = DT/3.0

END IF

ELSE IF ( ID(K) .EQ. 0 ) THEN

SUB(K) = DT/6.0

GLEFT = DT/3.0

ELSE IF ( ID(K) .EQ. -1 ) THEN

IF (K.NE.1) THEN

SUB(K) = DT/6.0

GLEFT = DT/3.0

ELSE IF (J1.EQ.3) THEN

T = XL - (DT/DA)*XL

W = 0.5*( X(K) + T )

SUB(K) = (X(K)-T) / 6.0 * ((T-XL)/DT) * ((X(K)-T)/DT)

C = 4.0* ((W-XL)/DT) * ((X(K)-W)/DT)

GLEFT = (X(K)-T) / 6.0 * ((T-XL)/DT)**2

C = 4.0* ((W-XL)/DT)**2 + 1.0

ELSE IF (J1.EQ.4) THEN

SUB(K) = DT/6.0

GLEFT = DT/3.0

ELSE IF (J1.EQ.3) THEN

SUB(K) = DT/6.0

GLEFT = DT/3.0

ELSE IF (J1.EQ.4) THEN

SUB(K) = DT/6.0

GLEFT = DT/3.0

ELSE IF (J1.EQ.3) THEN

SUB(K) = DT/6.0

GLEFT = DT/3.0
ELSE IF (J1.EQ.2) THEN

T = XL - (DT/DA)*XK

W = 0.5*(T + XL)

SUB(K) = (T - XL)/6.0 *( 4.0*((W - XL)/DT)*(X(K) - W)/DT)

C = ((T - XL)/DT)*((X(K) - T)/DT)

GLEFT = (T - XL)/6.0 *( 4.0*(((W - XL)/DT)**2)

ELSE

END IF

ELSE IF (J1.EQ.1) THEN

SUB(K) = 0.0

GLEFT = 0.0

END IF

ELSE IF (K.EQ.1) THEN

SUB(1) = 0.0

GLEFT = 0.0

IF (J1.EQ.4) GLEFT = DT/3.0

END IF

ELSE IF (K.NE.1) THEN

IF (D(K-1).EQ.0.0) THEN

SUB(K) = 0.0

GLEFT = 0.0

END IF

END IF

IF (K.NE.1) THEN

IF (J2.EQ.1) THEN

SUP(K) = DT/6.0

GRIGH = DT/3.0

ELSE

END IF

T = X(K) - (DT/DA)*A(K)

W = 0.5*(XR*T)

SUP(K) = (XR - T)/6.0 *( ((T - X(K))/DT)**2*(XR - T)/DT)
**FORTRAN Program Code**

```
00204 C + 4.0*((W-X(K))/DT)*(X-R)/DT)
00205 C = GRI=H = (X-R)/6.0 * ((X-R)/DT)**2
00206 C + 4.0*(((X-R)/DT)**2)
00207 C
00208 C ELSE IF (J2.EQ.3) THEN
00209 C
00210 T= X(K)-(DT/DA)*A(K)
00211 W= 0.5*( T+X(K) )
00212 C = SUP(K) = (T-X(K))/6.0 * ( 4.0*((W-X(K))/DT)*(X-R)/DT)
00213 C + ((T-X(K))/DT)**2
00214 GRI=H = (T-X(K))/6.0 * ( 1.0 + 4.0*(((X-R)/DT)**2)
00215 C + ((X-R)/DT)**2)
00216 C
00217 C ELSE IF (J2.EQ.4) THEN
00218 C
00219 C = SUP(K) = 0.0
00220 C = GRI=H = 0.0
00221 C
00222 C END IF
00223 C
00224 C ELSE IF (K.EQ.N) THEN
00225 C
00226 C = SUP(N) = 0.0
00227 C = GRI=H = 0.0
00228 IF (J2.EQ.1) GRI=H = DT/3.0
00229 C
00230 C END IF
00231 C
00232 C ELSE IF ( ID(K+1) .EQ. 0 ) THEN
00233 C
00234 C = SUP(K) = DT/6.0
00235 C = GRI=H = DT/3.0
00236 C
00237 C ELSE IF ( ID(K+1) .EQ. -1 ) THEN
00238 C
00239 C IF (K.EQ.N) THEN
00240 C
00241 C IF (.J2.EQ.4) THEN
00242 C
00243 C = SUP(K) = DT/6.0
00244 C = GRI=H = DT/3.0
00245 C
00246 ELSE IF (J2.EQ.3) THEN
00247 C
00248 C = T= X(K)-(DT/DA)*A(K)
00249 C = W= 0.5*( X+T )
00250 C = SUP(K) = (X-R)/6.0 * ((T-X(K))/DT)**2
00251 C + 4.0*(((W-X(K))/DT)**2)
00252 C = GRI=H = (X-R)/6.0 * ((X-R)/DT)**2
00253 C + 4.0*(((X-R)/DT)**2)
00254 C
```
ELSE IF (J2.EQ.2) THEN
  T = X(K)-(DT/DA)*A(K)
  W = 0.5*(T+X(K))
  SUP(K) = (T-X(K))/6.0 * (4.0*((W-X(K))/DT)**((XR-W)/DT))
  C = (T-X(K))/DT*(((XR-T)/DT)
  GRIGH = (T-X(K))/6.0 * (1.0 + 4.0*(((XR-W)/DT)**2))
ELSE IF (J2.EQ.1) THEN
  SUP(K) = 0.0
  GRIGH = 0.0
END IF
ELSE IF (K.EQ.N) THEN
  SUP(N) = 0.0
  GRIGH = 0.0
  IF (J2.EQ.4) GRIGH = DT/3.0
END IF
END IF
END IF
DIAG(K) = GLEFT+GRIGH
IF (D(K) .EQ. 0.0 ) THEN
  DIAG(K) = 1.0
  SUB(K) = 0.0
  SUP(K) = 0.0
END IF
CONTINUE
DO 150 L=1,N
H(L) = D(L)
CONTINUE
WE SOLVE THE MATRIX EQUATION JX=H, THE ARRAY (H)
BEING IDENTICAL TO THE ARRAY (D). THE SOLUTION
IS RETURNED IN THE ARRAY (H).

CALL TRID(SUB, DIAG, SUP, H, N)

SUM1 = 0.0

DO 200 L = 1, N
    A(L) = H(L)
    SUM1 = SUM1 + ABS(A(L))
CONTINUE

THE FUNCTION EVALUATION SUBROUTINE COMPUT MAY
BE DELETED. IN THIS CASE THE FOLLOWING EIGHT
LINES ARE TO BE DELETED AND THE ARRAY (FX)
CAN BE TAKEN FROM THE REAL STATEMENT AT THE
BEGINNING OF THIS SUBROUTINE.

CALL COMPUT(A, FX, N, X, TL, TR)

FNORM1 = 0.0

DO 250 L = 1, N
    FNORM1 = FNORM1 + FX(L) * FX(L)
CONTINUE

FNORM1 = SQRT(FNORM1)

WRITE(6, 300) LJ, FNORM1
FORMAT(I5, E15.6)

IF (LJ .NE. 1) THEN
    RATIO = ABS(SUM1 - SUM2)
    AB = EPS * SUM2
    IFLAG = 1
    IF (RATIO .LE. AB) GO TO 400
END IF

SUM2 = SUM1

CONTINUE

IFLAG = 2

CONTINUE

ITMAX = LJ

RETURN

END
SUBROUTINE COMPUT(A,FX,N,X,TL,IR)

SUBROUTINE (COMPUT), THE FUNCTION EVALUATING

SUBROUTINE, IS OPTIONAL.

REAL A(N),FX(N),F1,ALO,AHI,TLO,THI

REAL GLEF,GRIG,TS,X(N)

INTEGER N,K,J1,J2

COMMON D(50),IDC(50)

DO 100 K=1,N

IF ( D(K) .NE. 0.0 ) THEN

IF (K.EQ.1) THEN

ALO= 0.0

TLO= TL

ELSE

ALO= A(K-1)

TLO= X(K-1)

END IF

IF (K.EQ.N) THEN

AHI= 0.0

THI= TR

ELSE

AHI= A(K+1)

THI= X(K+1)

END IF

IF (ALO.GE.0.0 .AND. A(K).GE.0.0) J1= 1

IF (ALO.LT.0.0 .AND. A(K).GE.0.0) J1= 2

IF (ALO.GE.0.0 .AND. A(K).LT.0.0) J1= 3

IF (ALO.LT.0.0 .AND. A(K).LT.0.0) J1= 4

IF (A(K).GE.0.0 .AND. AHI.GE.0.0) J2= 1

IF (A(K).LT.0.0 .AND. AHI.GE.0.0) J2= 2

IF (A(K).GE.0.0 .AND. AHI.LT.0.0) J2= 3

IF (A(K).LT.0.0 .AND. AHI.LT.0.0) J2= 4

DT= X(K)-TLO

IF ( ID(K) .EQ. 1 ) THEN

IF (J1.EQ.1) THEN

GLEF= DT*( 2.0*A(K) + ALO )/6.0

00001  00002C  00003C  00004C  00005C  00006C  00007  00008  00009  00010  00011  00012C  00013  00014C  00015  00016  00017  00018  00019  00020  00021  00022C  00023  00024  00025  00026  00027  00028  00029  00030C  00031C  00032C  00033  00034  00035  00036  00037C  00038  00039  00040  00041  00042C  00043  00044C  00045  00046C  00047  00048C  00049  00050C
ELSE IF (J1.EQ.2) THEN
  TS = TLO - ALO*DT/(A(K)-ALO)
  F1 = (TS+X(K))*0.5 - TLO)/DT
  GLEF = (X(K)-TS)*A(K)*((2.0*F1 + 1.0)/6.0)
ELSE IF (J1.EQ.3) THEN
  TS = TLO - ALO*DT/(A(K)-ALO)
  GLEF = (TS-TLO)**2*ALO/(6.0*DT)
ELSE IF (J1.EQ.4) THEN
  GLEF = 0.0
END IF
END IF
ELSE IF (ID(K) .EQ. 0) THEN
  GLEF = DT*(2.0*A(K) + ALO)/6.0
ELSE IF (ID(K) .EQ. -1) THEN
  IF (J1.EQ.4) THEN
    TS = TLO - ALO*DT/(A(K)-ALO)
    F1 = (TS+X(K))*0.5 - TLO)/DT
    GLEF = (X(K)-TS)*A(K)*((2.0*F1 + 1.0)/6.0)
  ELSE IF (J1.EQ.3) THEN
    TS = TLO - ALO*DT/(A(K)-ALO)
    GLEF = (TS-TLO)**2*ALO/(6.0*DT)
  ELSE IF (J1.EQ.2) THEN
    TS = TLO - ALO*DT/(A(K)-ALO)
    GLEF = (TS-TLO)**2*ALO/(6.0*DT)
  ELSE IF (J1.EQ.1) THEN
    GLEF = 0.0
  END IF
END IF
END IF
END IF
DT = THI-X(K)
ELSE IF (ID(K+1) .EQ. 1) THEN
  IF (J2.EQ.1) THEN
    0
IF (J2.EQ.2) THEN
GRIG = DT*( 2.0*A(K) + AHI )/6.0
ELSE IF (J2.EQ.3) THEN
GRIG = ((THI-TS)**2)*AHI/(6.0*DT)
ELSE IF (J2.EQ.4) THEN
GRIG = 0.0
END IF
END IF

IF (J2.EQ.2) THEN
TS = X(K) - A(K)*DT/( AHI-A(K) )
ELSE IF (J2.EQ.3) THEN
TS = X(K) - A(K)*DT/( AHI-A(K) )
ELSE IF (J2.EQ.4) THEN
GRIG = 0.0
END IF

IF (J2.EQ.4) THEN
F1 = ( THI-0.5*( TS*X(K) ) )/DT
GRIG = ( TS-X(K) )*A(K)*( 1.0 + 2.0*F1 )/6.0
ELSE
F1 = ( THI-0.5*( TS*X(K) ) )/DT
GRIG = ( TS-X(K) )*A(K)*( 1.0 + 2.0*F1 )/6.0
END IF

IF (J2.EQ.4) THEN
TS = X(K) - A(K)*DT/( AHI-A(K) )
ELSE
GRIG = DT*( 2.0*A(K) + AHI )/6.0
END IF

IF (J2.EQ.4) THEN
GRIG = DT*( 2.0*A(K) + AHI )/6.0
ELSE
GRIG = DT*( 2.0*A(K) + AHI )/6.0
END IF

IF (J2.EQ.2) THEN
GRIG = DT*( 2.0*A(K) + AHI )/6.0
ELSE
GRIG = DT*( 2.0*A(K) + AHI )/6.0
END IF

IF (J2.EQ.2) THEN
GRIG = DT*( 2.0*A(K) + AHI )/6.0
ELSE
GRIG = DT*( 2.0*A(K) + AHI )/6.0
END IF

IF (K.NE.1) THEN
END IF
00153 IF ( D(K-1) .EQ. 0.0 ) GLEF= 0.0
00154 END IF
00155C
00156 IF (K.NE.N) THEN
00157 IF ( D(K+1) .EQ. 0.0 ) GRIG= 0.0
00158 END IF
00159C
00160 FX(K)= GLEF + GRIG - D(K)
00161C
00162 ELSE IF ( D(K) .EQ. 0.0 ) THEN
00163 FX(K)= 0.0
00164 END IF
00165C
00166 100 CONTINUE
00167 RETURN
00168 END
SUBROUTINE POLY(A,T,PP,M,F,L,TX)

SUBROUTINE POLY INTEGRATES BACK TWICE THE
POSITIVE PART OF THE PIECEWISE LINEAR SECOND
DERIVATIVE WHERE THE DATA SUGGESTS THAT THE
INTERPOLATING CURVE SHOULD BE CONVEX, THE
NEGATIVE PART OF THE PIECEWISE LINEAR SECOND
DERIVATIVE WHERE THE DATA SUGGESTS THAT THE
INTERPOLATING CURVE SHOULD BE CONCAVE, AND
THE REMAINING PORTION OF THE PIECEWISE LINEAR
SECOND DERIVATIVE ON THE TRANSITION INTERVALS.

THE INTEGRATION YIELDS A PIECEWISE CUBIC
POLYNOMIAL WITH KNOTS GIVEN BY THE SEQUENCE
(TX). THIS CUBIC POLYNOMIAL INTERPOLATES THE
DATA AND ITS COEFFICIENTS ARE DENOTED BY THE
NUMBERS PP(J,I) - THE VALUE OF THE (J-1)ST
DERIVATIVE OF THE FUNCTION EVALUATED AT TX(I),
FOR X SUCH THAT TX(I).GE.X.LT.TX(I+1) THE VALUE
OF THE CUBIC POLYNOMIAL IS

PP(1,I) + PP(2,I) * ( X-TX(I) )
+ (1/2)PP(3,I) * ( X-TX(I) )**2
+ (1/6)PP(4,I) * ( X-TX(I) )**3

INTEGER M,J,L,LI
REAL A(50),T(50),PP(4,100),F(50),TX(100),TAU
REAL DF,DT,DA,C,E
COMMON D(50),ID(50)
LI= 1
MN1= M-1
IOD 100 L=1,MN1
DF= F(L+1)-F(L)
DT= T(L+1)-T(L)
DA= A(L+1)-A(L)

JP= 0
IF (L.EQ.1) THEN
IF ( D(1) .EQ. 0.0 ) JP= 1
ELSE IF (L.EQ.MN1) THEN
IF ( D(M-2) .EQ. 0.0 ) JP= 1
ELSE
C= D(L-1)*D(L)
IF ( C .EQ. 0.0 ) JP= 1
END IF
IF (JP.EQ.1) THEN
PP(4,LI) = 0.0
PP(3,LI) = 0.0
PP(2,LI) = DF/DT
PP(1,LI) = F(L)
TX(LI) = T(L)
LI = LI+1
ELSE IF (JP.EQ.0) THEN
IF (A(L).GE.0.0 .AND. A(L+1).GE.0.0) J = 1
IF (A(L).LT.0.0 .AND. A(L+1).GT.0.0) J = 2
IF (A(L).LE.0.0 .AND. A(L+1).LE.0.0) J = 3
IF (ID(L) .EQ. 1) THEN
IF (J.EQ.1) THEN
C = DF/DT - (DA/6.0 + A(L)/2.0)*DT
PP(4,LI) = DA/DT
PP(3,LI) = A(L)
PP(2,LI) = C
PP(1,LI) = F(L)
TX(LI) = T(L)
LI = LI+1
ELSE IF (J.EQ.2) THEN
TAU = T(L) - A(L)*DT/DA
C = DF/DT - (A(L)**3)*DT/(6.0*DA*DA)
PP(4,LI) = 0.0
PP(3,LI) = 0.0
PP(2,LI) = C
PP(1,LI) = F(L)
PP(4,LI+1) = DA/DT
PP(3,LI+1) = 0.0
PP(2,LI+1) = C
PP(1,LI+1) = C*(TAU-T(L)) + F(L)
TX(LI) = T(L)
TX(LI+1) = TAU
LI = LI+2
ELSE IF (J.EQ.3) THEN
TAU = T(L) - A(L)*DT/DA
E = F(L) - (A(L)**3)*DT**2/(6.0*DA*DA)
C = DF/DT + (A(L)**3)*DT/(6.0*DA*DA)
PP(4,LI) = DA/DT
PP(3,LI) = A(L)
PP(2,LI) = C + A(L)*A(L)*DT*0.5/DA
PP(1,LI) = F(L)
PP(4,LI+1) = 0.0
PP(3,LI+1) = 0.0
PP(2,LI+1) = C
PP(1,LI+1) = C*(TAU-T(L)) + E
TX(LI) = T(L)
TX(LI+1) = TAU
LI = LI+2
ELSE IF (J.EQ.4) THEN
PP(4,LI) = 0.0
PP(3,LI) = 0.0
PP(2,LI) = DF/DT
PP(1,LI) = F(L)
TX(LI) = T(L)
LI = LI+1
END IF
ELSE IF (ID(L) .EQ. 0) THEN
C = DF/DT - (DA/6.0 + A(L)/2.0)*DT
PP(4,LI) = DA/DT
PP(3,LI) = A(L)
PP(2,LI) = C
PP(1,LI) = F(L)
TX(LI) = T(L)
LI = LI+1
ELSE IF (ID(L) .EQ. -1) THEN
C = DF/DT - (DA/6.0 + A(L)/2.0)*DT
PP(4,LI) = DA/DT
PP(3,LI) = A(L)
PP(2,LI) = C
PP(1,LI) = F(L)
TX(LI) = T(L)
LI = LI+1
ELSE IF (J.EQ.4) THEN
C = DF/DT - (DA/6.0 + A(L)/2.0)*DT
PP(4,LI) = DA/DT
PP(3,LI) = A(L)
PP(2,LI) = C
PP(1,LI) = F(L)
TX(LI) = T(L)
LI = LI+1
ELSE IF (J.EQ.3) THEN
TAU = T(L) - A(L)*DT/DA
C = DF/DT - (A(L+1)**3)*DT/(6.0*DA**DA)
PP(4,LI) = 0.0
PP(3,LI) = 0.0
PP(2,LI) = C
PP(1,LI) = F(L)
PP(4,LI+1) = DA/DT
PP(3,LI+1) = 0.0
PP(2,LI+1) = C
PP(1,LI+1) = C*(TAU-T(L)) + F(L)
TX(LI) = T(L)
TX(LI+1) = TAU
LI = LI+2

ELSE IF (J.EQ.2) THEN

TAU = T(L) - A(L)*DT/DA
E = F(L) - (A(L)**3)*DT*DT/(6.0*DA*DA)
C = DF/DT + (A(L)**3)*DT/(6.0*DA*DA)
PP(4,LI) = DA/DT
PP(3,LI) = A(L)
PP(2,LI) = C + A(L)*A(L)*DT*0.5/DA
PP(1,LI) = F(L)
PP(3,LI+1) = 0.0
PP(2,LI+1) = C
PP(1,LI+1) = C*(TAU-T(L)) + E
TX(LI) = T(L)
TX(LI+1) = TAU
LI = LI+2

ELSE IF (J.EQ.1) THEN

PP(4,LI) = 0.0
PP(3,LI) = 0.0
PP(2,LI) = DF/DT
PP(1,LI) = F(L)
TX(LI) = T(L)
LI = LI+1

END IF

END IF

END IF

END IF
Subroutines TRID and DATAFL are listed in Appendix A.
In computational fluid dynamics and in CAD/CAM, a physical boundary is usually known only discretely and most often must be approximated. An acceptable approximation preserves the salient features of the data such as convexity and concavity. In this dissertation, a smooth interpolant which is locally concave where the data are concave and is locally convex where the data are convex is described. The interpolant is found by posing and solving a minimization problem whose solution is a piecewise cubic polynomial. The problem is solved indirectly by using the Peano Kernel theorem to recast it into an equivalent minimization problem having the second derivative of the interpolant as the solution. This approach leads to the solution of a nonlinear system of equations. It is shown that Newton's method is an exceptionally attractive and efficient method for solving the nonlinear system of equations. Examples of shape-preserving interpolants as well as convergence results obtained by using Newton's method are also shown. A FORTRAN program to compute these interpolants is listed. The problem of computing the interpolant of minimal norm from a convex cone in a normal dual space is also discussed. An extension of de Boor's work on minimal norm unconstrained interpolation is presented.