Shape preserving interpolation by quadratic splines

Aatos LAHTINEN
Department of Mathematics, University of Helsinki, Hallituskatu 15, SF-00100 Helsinki, Finland

Received 13 April 1988
Revised 28 February 1989

Abstract: An interpolating quadratic spline was constructed which preserves the shape of data. The spline contains parameters which can be used to match it to additional information.

Keywords: Quadratic spline, preserving the shape.

1. Introduction

In many situations the values of a function are known only at some fixed set of points and in addition something may be known about the shape of the function. On the basis of this information one has then to form an approximation to the function.

There are several ways of doing this. McAllister and Roulier [3,4] have studied the existence of an interpolating quadratic spline with a given shape. They have even presented an algorithm for construction of such a spline in [5]. Their method is not, however, as flexible as is desirable.

This article considers a flexible construction of a shape preserving interpolating quadratic spline based on the work of Schumaker [8]. The basis of the construction is the Hermite interpolation with regard to the function values and to the first derivatives. It is shown how the monotony of the interpolating quadratic spline depends on the derivatives. This, combined with the convexity results of Schumaker [8], gives the shape of the spline as a function of the derivatives and additional breakpoints. A method to choose the derivatives is then presented and it is shown that it produces a shape preserving spline. The derivatives and additional breakpoints remain as parameters which can be chosen within certain limits. This freedom can be used for example to match the spline with some additional information about the data.

2. The problem

Let \([a, b]\) be an interval containing points \((x_i)_{i=1}^n\) such that \(a = x_1 < x_2 < \cdots < x_{n-1} < x_n = b\) and let \((y_i)_{i=1}^n\) be real numbers. We use the notation \(\Delta x_i = x_{i+1} - x_i\) for difference and \(\delta_i = \Delta y_i / \Delta x_i\) for the first divided difference.

The point set $D = \{(x_i, y_i)\}_{i=1}^n$ is called increasing (respectively decreasing) if the set $(y_i)_{i=1}^n$ is increasing (resp. decreasing) and $D$ is called convex (resp. concave) if $(\delta_i)_{i=1}^{n-1}$ is increasing (resp. decreasing).

For the motivation of these terms suppose that there exists a function $g$ with $g(x_i) = y_i$, $i = 1, \ldots, n$. If $g$ is increasing (resp. decreasing) on $[a, b]$, then $D$ is also increasing (resp. decreasing) and if $g$ is convex (resp. concave), then $D$ is also convex (resp. concave). The common name of an increasing and decreasing point set is a monotone. A piecewise monotone point set is defined in a natural way. Similarly the point set $D$ is said to be convex (resp. concave) on a subinterval $[x_i, x_{i+1}] \subset [a, b]$ if $(\delta_i)_{i=1}^{i+1}$ is increasing (resp. decreasing).

A function $f$ interpolating at the set $D$ is shape preserving or preserves the shape of $D$ if it is piecewise monotone in the same way as $D$ and if it is convex and concave in the same way as $D$. Now we can formulate our problem:

**Let points $a = x_1 < x_2 < \cdots < x_n = b$ and values $(y_i)_{i=1}^n$ be given. Find a continuously differentiable function $f : [a, b] \to \mathbb{R}$ with $f(x_i) = y_i$, $i = 1, \ldots, n$, which preserves the shape of $((x_i, y_i))_{i=1}^n$.**

3. Hermite interpolation

We will try to find the solution of our problem among quadratic splines with $C^1$-regularity. A common interpolating quadratic spline has one or two degrees of freedom which globally effect the spline. Passow [7] has shown that this is not sufficient for reproducing the shape of the data.

The decisive idea is the use of splines which have more breakpoints than interpolating points. The additional breakpoints are not determined beforehand but are added to suitable places when necessary. In this construction the derivatives $(\delta_i)_{i=1}^{i+1}$ at interpolating points are free parameters which effect the spline only locally. This means that on the interval $[x_i, x_{i+1}]$ the spline depends only on the derivatives $m_i$ and $m_{i+1}$.

McAllister and Roulier [4] have presented a way to construct this kind of quadratic spline. We will follow here the idea of Schumaker [8] which is based on the Hermite interpolation problem.

**The Hermite Interpolation Problem (HIP).** Let an interval $[x_i, x_{i+1}]$ and numbers $y_i, y_{i+1}, m_i, m_{i+1}$ be given. Find a quadratic spline $s$ with the fewest number of breakpoints such that $s(x_j) = y_j$, $s'(x_j) = m_j$, $j = i, i + 1$.

**The solution of the HIP:** This problem has always a solution. The number of breakpoints depends on the values $m_i, m_{i+1}$ and on the divided difference $\delta_i = \Delta y_i / \Delta x_i$.

(a) If $m_i + m_{i+1} = 2\delta_i$, then the solution is given by a parabola $s$,

$$
s(x) = y_i + m_i(x - x_i) + \frac{\Lambda m_i}{2\Lambda x_i} (x - x_i)^2.
$$

(b) If $m_i + m_{i+1} \neq 2\delta_i$, then the solution cannot be a parabola. The solution is given by a quadratic spline with one breakpoint $\xi_i$ which can be chosen anywhere on the open interval $]x_i, x_{i+1}[$. The presentation of the spline $s$ is

$$
s(x) = \begin{cases} y_i + m_i(x - x_i) + \frac{\mu_i - m_i}{2(\xi_i - x_i)} (x - x_i)^2, & x_i \leq x < \xi_i, \\
\delta_i + \mu_i(x - \xi_i) + \frac{m_{i+1} - \mu_i}{2(x_{i+1} - \xi_i)} (x - \xi_i)^2, & \xi_i \leq x < x_{i+1}, \end{cases}
$$
where

\[
\mu_i = 2\delta_i - \frac{\xi_i - x_i}{\Delta x_i} m_i - \frac{x_{i+1} - \xi_i}{\Delta x_i} m_{i+1}, \quad d_i - y_i + \frac{1}{2}(m_i + \mu_i)(\xi_i - x_i).
\]

Note that \( s(\xi_i) = d_i \) and \( s'(\xi_i) = \mu_i \). The Hermite Interpolation Problem makes it possible to construct an interpolating quadratic spline to a data set \( D = ((x_i, y_i)) \) for every given set of breakpoint derivatives \( (m_i) \) by adding new breakpoints between certain interpolating points. The additional breakpoints can be freely chosen within the interval between two original breakpoints. Thus both derivatives \( m_i \) and breakpoints \( \xi_i \) are free parameters.

4. The monotony

The Hermite interpolation provides us with an interpolation method which has several degrees of freedom, namely the derivatives \( m_i \) at the breakpoints and the additional breakpoints \( \xi_i \). We have now to examine whether these parameters can be chosen so that the interpolating quadratic spline preserves the shape of the data. We will start with the conservation of the monotony.

It is sufficient to consider monotony locally, i.e., separately on each interval \([x_i, x_{i+1}]\). Suppose at first that there is no additional breakpoint at \([x_i, x_{i+1}]\). Then it is easy to see that the solution of the HIP is monotone on \([x_i, x_{i+1}]\) if and only if the derivatives \( m_i \) and \( m_{i+1} \) have the same sign.

If there is an additional breakpoint at the interval \([x_i, x_{i+1}]\), then the condition that the derivatives \( m_i \) and \( m_{i+1} \) and the divided difference \( \delta_i \) all have the same sign is clearly a necessary one. It is not, however, sufficient as the next proposition shows (cf. [8]).

Proposition 1. If \( m_i + m_{i+1} = 2\delta_i \), then the solution \( s \) of the HIP is monotone on \([x_i, x_{i+1}]\) if and only if \( m_i \) and \( m_{i+1} \) have the same sign.

If \( m_i + m_{i+1} \neq 2\delta_i \) and \( m_i, m_{i+1} \) and \( \delta_i \) have the same sign, then the solution \( s \) of the HIP is monotone on \([x_i, x_{i+1}]\) if and only if

\[
\min(|m_i|, |m_{i+1}|) < 2|\delta_i|, \quad \text{if } m_i \neq m_{i+1},
\]

and the additional breakpoint \( \xi_i \) is chosen so that

\[
\begin{align*}
\max(x_i, c_i) < \xi_i < x_{i+1}, & \quad \text{if } |m_i| < |m_{i+1}|, \\
x_i < \xi_i < x_{i+1}, & \quad \text{if } |m_i| = |m_{i+1}|, \\
x_i < \xi_i < \min(x_{i+1}, c_i), & \quad \text{if } |m_i| > |m_{i+1}|.
\end{align*}
\]

Here the point \( c_i \) has the expression

\[
c_i = x_i + \Delta x_i \frac{m_{i+1} - 2\delta_i}{\Delta m_i}.
\]

Proof. The first part of the proposition has already been mentioned. Therefore we can propose that an extra breakpoint \( \xi_i \) is needed. It is easy to see that in this case the solution \( s \) of the HIP
is monotone on \([x_i, x_{i+1}]\) if and only if \(m_i, m_{i+1}\) and \(s'(\xi_i)\) have the same sign. By using the expression of the solution of the HIP for \(s'(\xi_i)\), this condition can be written
\[
2|\delta_i| \geq \frac{1}{\Delta x_i}\left(\frac{(\xi_i - x_i)m_i + (x_{i+1} - \xi_i)m_{i+1}}{m_i + m_{i+1}}\right).
\] (3)
As all the other quantities are known, this gives a condition for placing the breakpoint \(\xi_i\). On the other hand, we must have
\[
x_i < \xi_i < x_{i+1}.
\] (4)
Both conditions cannot be valid unless the sizes of the derivatives \(m_i\) and \(m_{i+1}\) are restricted.

If \(|m_i| = |m_{i+1}|\), then it can be seen that (3) is valid for every \(\xi_i\) fulfilling (4), if and only if \(|m_i| \leq 2|\delta_i|\).

Suppose then that \(|m_i| < |m_{i+1}|\). Now (3) is equivalent to the condition
\[
\xi_i \geq x_i + \Delta x_i \frac{m_{i+1} - 2\delta_i}{\Delta m_i} = c_i.
\]
By condition (4) we must demand that
\[
c_i < x_{i+1}.
\] (5)
It is easy to see that (5) is valid if the derivatives \(m_i\) and \(m_{i+1}\) in the expression of \(c_i\) are restricted by (1). Thus both (3) and (4) are fulfilled for breakpoints \(\xi_i \in \text{max}(x_i, c_i), x_{i+1}\). This proves the proposition for values \(|m_i| < |m_{i+1}|\). The case \(|m_i| > |m_{i+1}|\) is treated similarly.

In some cases the interval (2) where the breakpoint \(\xi_i\) must be chosen is very short. This may produce a corner to the graph of the spline \(s\). This can be avoided by demanding that the length of the interval (2) is always at least \(\epsilon \Delta x_i\), where \(\epsilon, 0 < \epsilon \leq 1\), is a given parameter. In this case the result of Proposition 1 is valid if condition (1) is replaced by
\[
(1 - \epsilon) \min(|m_i|, |m_{i+1}|) + \epsilon \max(|m_i|, |m_{i+1}|) \leq 2|\delta_i|.
\] (6)
Notice that condition (1) restricts only the smaller derivative whereas (6) restricts both derivatives.

The initial steps in the proof of this modification of Proposition 1 are the same as in the proof of Proposition 1. So in the case \(|m_i| < |m_{i+1}|\) we must have conditions (3) and (4). Further (3) is valid if \(\xi_i \geq c_i\). Because we have now a minimal length \(\epsilon \Delta x_i\) for the interval where \(\xi_i\) is to be chosen, we demand instead of (5) that
\[
c_i \leq x_{i+1} - \epsilon \Delta x_i.
\]
It is easy to see that this is valid if the derivatives \(m_i\) and \(m_{i+1}\) in the expression of \(c_i\) are restricted by (6). This proves the modification in the case \(|m_i| < |m_{i+1}|\). The case \(|m_i| > |m_{i+1}|\) is treated similarly. If \(|m_i| = |m_{i+1}|\), then the additional condition is trivially valid.

Proposition 1 says that all derivatives \(m_j\) with the necessary sign don’t produce a monotone solution of the HIP. On the other hand, the derivatives \(m_i\) are free parameters. If they have been given so that condition (1) fails for some value \(j\), then the derivatives \(m_j\) and \(m_{j+1}\) can be changed so that a monotone spline is achieved.
If, however, the derivatives $m_i$ are chosen according to some additional information, e.g., according to the shape of the function, then it may not be desirable to change them too much. One possible compromise is given in the following proposition.

**Proposition 2.** Let $x_i$, $x_{i+1}$, $y_i$, $y_{i+1}$, $m_i$, $m_{i+1}$ be the initial values of the HIP so that $m_i$, $m_{i+1}$ and $\delta_i$ have the same sign and

$$\min(|m_i|, |m_{i+1}|) > 2|\delta_i|.$$  

Then the replacement of the derivatives $m_i$ and $m_{i+1}$ by values

$$m'_i = \frac{\alpha_i |\delta_i| m_i}{|\beta_i m_i + (1 - \beta_i) m_{i+1}|}, \quad m'_{i+1} = \frac{\alpha_i |\delta_i| m_{i+1}}{|\beta_i m_i + (1 - \beta_i) m_{i+1}|},$$

where $0 < \alpha_i < 2$ and $0 < \beta_i < 1$ gives a monotone solution of the HIP on the interval $[x_i, x_{i+1}]$. If an additional breakpoint $\xi_i$ is needed, it must be chosen

$$\xi_i = (1 - \beta_i) x_i + \beta_i x_{i+1}.$$  

**Proof.** Now $m'_i$, $m'_{i+1}$ and $\delta_i$ have the same sign. If $m'_i + m'_{i+1} = 2\delta_i$, the statement follows from the first part of Proposition 1. If $m'_i + m'_{i+1} \neq 2\delta_i$, then the solution of the HIP is monotone if the additional breakpoint $\xi_i$ is chosen so that $x_i < \xi_i < x_{i+1}$ and

$$2|\delta_i| \geq \frac{1}{\Delta x_i} \left( ((\xi_i - x_i) |m'_i| + (x_{i+1} - \xi_i) |m'_{i+1}|) \right).$$

It can be seen that the derivatives $m'_i$, $m'_{i+1}$ and the breakpoint $\xi_i$ fulfill these conditions by direct calculation. \(\square\)

Note that one can place the breakpoint $\xi_i$ on the interval $[x_i, x_{i+1}]$ by fixing the parameter $\beta_i$ and then determining the magnitude of derivatives with the parameter $\alpha_i$.

If the quadratic spline is constructed by starting on the interval $[x_1, x_2]$, then the use of Proposition 2 means that the expression of the spline has to be recomputed on the interval $[x_{i-1}, x_i]$. If this is not convenient, then the derivative $m_i$ must be intact and only $m_{i+1}$ is changed. One possibility is to choose simply

$$m'_{i+1} = \alpha_i \delta_i, \quad 0 < \alpha_i < 2,$$

where the number $\alpha_i$ can regulate the magnitude of $m'_{i+1}$.

5. The convexity

Even when the monotony of the solution of the HIP has been assured there is still freedom in the choice of derivatives $m_i$ and additional breakpoints $\xi_i$. This freedom can be used in trying to get the solution of the HIP to preserve also the convexity and concavity of the data. It is sufficient to consider this locally, i.e., separately for each interval $[x_i, x_{i+1}]$. Schumaker [8] has presented the following result.
Proposition 3. Let $s$ be a solution of the HIP on the interval $[x_i, x_{i+1}]$ in the case where $m_i$, $m_{i+1}$ and $\delta_i$ have the same sign.

(a) If $m_i + m_{i+1} = 2\delta_i$, then $s$ is convex, when $\Delta m_i > 0$ and concave, when $\Delta m_i < 0$.
(b) If $m_i + m_{i+1} \neq 2\delta_i$ and
\[
(m_{i+1} - \delta_i)(m_i - \delta_i) < 0,
\]
then $s$ is convex, when $\Delta m_i > 0$ and concave, when $\Delta m_i < 0$ supposing that the additional breakpoint $\xi_i$ is chosen so that
\[
\begin{aligned}
x_i < \xi_i \leq x_i + 2\Delta x_i \frac{m_{i+1} - \delta_i}{\Delta m_i}, \quad & \text{when } |m_{i+1} - \delta_i| < |m_i - \delta_i|, \\
x_{i+1} - 2\Delta x_i \frac{\delta_i - m_i}{\Delta m_i} \leq \xi_i < x_{i+1}, \quad & \text{when } |m_{i+1} - \delta_i| > |m_i - \delta_i|.
\end{aligned}
\]
(7)

(c) If $m_i + m_{i+1} \neq 2\delta_i$ and
\[
(m_{i+1} - \delta_i)(m_i - \delta_i) > 0,
\]
then $s$ has a point of inflection on the interval $[x_i, x_{i+1}]$.

6. The choice of derivatives

The Propositions 1 and 3 together give the shape of a solution of the HIP. Proposition 3 is similar to Proposition 1 in the sense that both restrict derivatives and the positions of additional breakpoints. The restrictions are, however, different in the case where $m_i + m_{i+1} \neq 2\delta_i$. On the other hand, it is easy to see that the restrictions in part (b) of Proposition 3 imply the validity of restrictions of Proposition 1. This leads to the following result.

Proposition 4. Let the derivatives $m_i$, $m_{i+1}$ and the breakpoint $\xi_i$ be chosen so that they fulfill the conditions of part (b) of Proposition 3. Then the corresponding solution of the HIP is monotone on the interval $[x_i, x_{i+1}]$.

The fact that $m_i$, $m_{i+1}$ and $\xi_i$ fulfill the conditions of part (c) of Proposition 3 does not convey any information about monotony. We must state the conditions of Proposition 1 for monotony in this case.

The results of the propositions depend on the choice of derivatives $m_i$ at the interpolating points. Therefore it is necessary to find out which way the derivatives must be chosen so that the resulting solution of the HIP preserves the shape of the data.

At any point $x_i$ what we know about the change of the data are only the slopes $\delta_{i-1}$ and $\delta_i$. Therefore we take the derivatives at breakpoints $(x_i)_{i}$ as follows:

\[
m_i = \begin{cases} 
\delta_i + a_i(\delta_{i-1} - \delta_i), & 0 \leq a_i \leq 1, \quad \text{if } \delta_{i-1}\delta_i > 0, \\
0, & \text{if } \delta_{i-1}\delta_i \leq 0,
\end{cases}
\]
(8a)

when $i = 2, \ldots, n - 1$ and

\[
m_1 = a_1\delta_1, \quad 0 \leq a_1, \\
m_n = a_n\delta_{n-1}, \quad 0 \leq a_n.
\]
(8b)
We generally choose $m_i$ to be a convex combination of the slopes $\delta_{i-1}$ and $\delta_i$. If, however, the slopes have a different sign, or at least one of them is zero, we take $m_i = 0$. This latter case is necessary for preserving the monotony. It also means that if for instance $\delta_i = 0$, then we have $m_i = m_{i+1} = 0$ and the solution $s$ of the HIP is constant on the interval $[x_i, x_{i+1}]$.

In the derivatives the numbers $a_i$ are parameters. We are interested in examining whether this choice of derivatives preserves the shape of the data for some parameter values.

7. Shape preserving interpolation

We consider from now on the solution $s$ of the HIP in the case where the derivatives are chosen according to (8). We will start the examination of how to choose the parameter values $a_i$ by studying the monotony.

It is sufficient to prove the monotony locally. Therefore we consider the situation on the interval $[x_i, x_{i+1}]$ supposing at first that $2 \leq i \leq n - 2$. Proposition 1 indicates when the solution $s$ of the HIP is monotone on $[x_i, x_{i+1}]$. The essential condition is (1) which restricts the derivatives. If it is fulfilled, then the possible additional breakpoint can always be chosen according to (2).

If the slopes $\delta_{i-1}$, $\delta_i$ and $\delta_{i+1}$ don't have an equal sign, then $s$ preserves the monotony of the data for every choice

$$0 \leq a_i \leq 1, \quad 0 \leq a_{i+1} \leq 1.$$  \hfill (9)

If $\delta_{i-1}$, $\delta_i$ and $\delta_{i+1}$ have an equal sign, then

$$\min(|\delta_{j-1}|, |\delta_j|) \leq |m_j| \leq \max(|\delta_{j-1}|, |\delta_j|), \quad j = i, i + 1.$$  \hfill (10)

Therefore, by (1), the spline $s$ preserves the monotony of the data according to the condition (9) if

$$\min(|\delta_{i-1}|, |\delta_{i+1}|) < 2|\delta_i|.$$  \hfill (11a)

If $\delta_{i-1}$, $\delta_i$ and $\delta_{i+1}$ have an equal sign, but (10) is not valid, then $s$ preserves the monotony of the data only if either

$$0 \leq a_i < \frac{\delta_i}{\delta_{i-1} - \delta_i}, \quad 0 \leq a_{i+1} < 1,$$  \hfill (11a)

or

$$0 \leq a_i \leq 1, \quad 1 - \frac{\delta_i}{\delta_{i+1} - \delta_i} < a_{i+1} \leq 1.$$  \hfill (11b)

Thus for an interval $[x_i, x_{i+1}]$ which is not an end interval we can always find values $a_i$, $a_{i+1}$ and, when necessary, additional breakpoint $\xi_i$ such that the HIP-solution $s$ preserves the monotony of the data on $[x_i, x_{i+1}]$.

The same is true for end intervals. Consider the interval $[x_1, x_2]$. If $\delta_1$ and $\delta_2$ have a different sign or if $|\delta_2| < 2|\delta_1|$, then $s$ preserves the monotony of the data if

$$a_1 \in \mathbb{R}_+, \quad 0 \leq a_2 \leq 1.$$  \hfill (12)
If \( \delta_1 \) and \( \delta_2 \) have an equal sign and \( |\delta_2| \geq 2 |\delta_1| \), then \( s \) preserves the monotony of the data if either

\[
a_1 \in \mathbb{R}_+, \quad 1 - \frac{\delta_1}{\delta_2 - \delta_1} < a_2 \leq 1, \tag{13a}
\]
or

\[
0 \leq a_1 < 2, \quad 0 \leq a_2 \leq 1. \tag{13b}
\]

Similar results are valid for the preservation of monotony of the data on the interval \([x_{n-1}, x_n]\).

We collect the monotony results.

**Proposition 5.** Let \( s \) be on the interval \([x_i, x_{i+1}]\) a solution of the HIP, where the derivatives \( m_i \) and \( m_{i+1} \) are as in (8). Then there exist intervals \( A_j, j = i, i + 1, \) and \( B_i \) such that \( s \) preserves the monotony of the data for every \( a_i \in A_i, a_{i+1} \in A_{i+1} \), and, if an additional breakpoint \( \xi_i \) is needed, for every \( \xi_i \in B_i \).

After the preserving of the monotony is settled we can turn our attention to the issue of the convexity and concavity. First suppose that the data is convex on an interval \([x_i, x_{i+1}]\), \( 2 \leq i \leq n - 2 \). Then \( \delta_{i-1} < \delta_i < \delta_{i+1} \). If these slopes have an equal sign and

\[
0 < a_i < 1, \quad 0 < a_{i+1} < 1, \tag{14}
\]

then we get \( \delta_{i-1} < m_i < \delta_i < m_{i+1} < \delta_{i+1} \). This shows, by Proposition 3, that the solution \( s \) of the HIP is convex on \([x_i, x_{i+1}]\). A similar result is valid for concavity. By considering also the other sign combinations of the slopes we get the following result.

**Proposition 6.** Suppose that the data is convex (resp. concave) on an interval \([x_i, x_{i+1}]\), \( 2 \leq i \leq n - 2 \), and that \( s \) is a solution of the HIP with derivatives (8) so that (14) is valid. If \( \delta_i \neq 0 \), then \( s \) is convex (resp. concave) on \([x_i, x_{i+1}]\), when the additional breakpoint, if necessary, is chosen as in (7). If \( \delta_i = 0 \), then \( s \) is constant on \([x_i, x_{i+1}]\).

Local convexity on each interval \([x_i, x_{i+1}]\) implies global convexity on \([a, b]\). If the data is globally convex, then Proposition 6 ensures the convexity of the solution of the HIP on \([x_2, x_{n-1}]\). The same is true for concavity. It thus remains to see that \( s \) does not behave differently at end intervals.

It is quite apparent that if \( a_2 \in ]0, 1[ \) and \( a_1 \) is chosen so that

\[
\begin{cases}
(a_1 - 1)\delta_1(\delta_2 - \delta_1) < 0, & \text{when } \delta_1(\delta_2 - \delta_1) \neq 0, \\
1, & \text{when } \delta_1(\delta_2 - \delta_1) = 0,
\end{cases} \tag{15}
\]

then \( s \) has the wanted behaviour on \([x_1, x_2]\). The same is true on \([x_{n-1}, x_n]\), if \( 0 < a_{n-1} < 1 \) and \( a_n \) is such that

\[
\begin{cases}
(a_n - 1)\delta_{n-1}(\delta_{n-1} - \delta_{n-2}) < 0, & \text{when } \delta_{n-1}(\delta_{n-1} - \delta_{n-2}) \neq 0, \\
1, & \text{when } \delta_{n-1} - \delta_{n-2} = 0.
\end{cases} \tag{16}
\]

Suppose that the data is globally convex (resp. concave) and \( s \) is the solution of the HIP. Then \( s \) is convex (resp. concave) on \([a, b]\) if \( 0 < a_i < 1, 2 \leq i \leq n - 1 \) and \( a_1, a_n \) are as in (15), (16) and the necessary additional breakpoints \( \xi_i \) are as in (7).
If the data is not globally convex or globally concave, then we have to see what happens to the HIP-solution \( s \) between the intervals of convexity and intervals of concavity. We start with an auxiliary result.

**Proposition 7.** Let \( s \) be a solution of the HIP on the interval \([x_i, x_{i+1}]\) with the derivative values \( m_i \) and \( m_{i+1} \). If \( m_i < \delta_i > m_{i+1} \) or \( m_i > \delta_i < m_{i+1} \), then \( s \) has an additional breakpoint \( \xi_i \) and one point of inflection which is \( \xi_i \). If \( m_i < \delta_i > m_{i-1} \), then \( s \) is convex on \([x_i, \xi_i]\) and concave on \([\xi_i, x_{i+1}]\). If \( m_i > \delta_i < m_{i+1} \), then \( s \) is concave on \([x_i, \xi_i]\) and convex on \([\xi_i, x_{i+1}]\).

**Proof.** Suppose that \( m_i < \delta_i > m_{i+1} \). The expression of \( s \) implies now that

\[
\frac{s'(\xi_i)}{m_i} = 2(\delta_i - \max(m_i, m_{i+1})) > 0,
\]

\[
m_{i+1} - s'(\xi_i) \leq 2(\max(m_i, m_{i+1}) - \delta_i) < 0.
\]

This implies the statement by Proposition 3. The case \( m_i > \delta_i < m_{i+1} \) is similar. \( \Box \)

Suppose now that the data is convex on \([x_{i-1}, x_i]\) and concave on \([x_i, x_{i+1}]\). This means that \( \delta_{i-2} < \delta_i < \delta_{i+1} > \delta_{i+2} \). Let \( s \) again be the solution of the HIP with derivatives (8). Proposition 7 shows that for every \( < a_i < 1 < a_{i+1} \), the spline is either constant on \([x_i, x_{i+1}]\) or \( s \) has a breakpoint \( \xi_i \) on \([x_i, x_{i+1}]\) and \( s \) is convex on \([x_{i-1}, \xi_i]\) and concave on \([\xi_i, x_{i+1}]\). A similar result is valid if the data is concave on \([x_{i-1}, x_i]\) and convex on \([x_{i+1}, x_{i+2}]\).

This means that if the data changes from convex to concave or vice versa, then the HIP-solution \( s \) changes similarly from convex to concave or vice versa without any extra points of inflection.

In all previous considerations we have supposed that \( \delta_i \neq \delta_{i+1} \) for all values \( i \). Finally we will examine what happens if two adjacent slopes have the same value. Let \( s \) be the solution of the HIP with derivatives (8) where parameters \( a_i \) and additional breakpoints \( \xi_i \) have been chosen according to previous considerations.

First suppose that the equality of slopes happens between two areas of convexity so that \( \delta_{i-1} < \delta_i = \delta_{i+1} < \delta_{i+2} \). If \( \delta_i = 0 \), then clearly \( s \) is constant on the interval \([x_i, x_{i+1}]\). Otherwise \( s \) has breakpoints \( \xi_i \in [x_i, x_{i+1}] \) and \( \xi_{i+1} \in [x_{i+1}, x_{i+2}] \). They are also points of inflection so that \( s \) is concave on \([\xi_i, \xi_{i+1}]\) and convex elsewhere. A similar result is valid if the equality of slopes takes place between two areas of concavity.

Suppose then that this equality takes place between an interval of convexity and an interval of concavity so that \( \delta_{i-1} < \delta_i = \delta_{i+1} > \delta_{i+2} \). If \( \delta_i = 0 \), then again \( s \) is constant on the interval \([x_i, x_{i+1}]\). If \( \delta_i \neq 0 \), then it is easy to see that \( s \) has two additional breakpoints \( \xi_i \in [x_i, x_{i+1}] \) and \( \xi_{i+1} \in [x_{i+1}, x_{i+2}] \) which are also points of inflection. In this case, however, the point \( x_{i+1} \) is also a point of inflection. Thus \( s \) changes the direction of curvature three times between convexity and concavity on the interval \([x_i, x_{i+2}]\).

### 8. The solution

By using the results of preceding sections we can now answer to the question of the solvability of our problem.
Theorem 8. Let points \( a = x_1 < x_2 < \cdots < x_n = b \) and values \((y_i)_i\) be given. Choose the values \((m_i)_i\) as in (8). Then there exist parameter intervals \((A_i)_i\) and \((B_i)_i\) such that the solution \(s\) of the HIP preserves the shape of data for every \(a_i \in A_i\) and for every necessary additional breakpoint \(\xi_i \in B_i\).

Because all the results leading to this theorem are constructive, it is easy to form an algorithm which constructs the shape preserving solution of the HIP. In fact, such an algorithm is presented in [2]. There it is used for tasks connected with forest inventory. On each interval \([x_i, x_{i+1}]\) there is freedom in the choice of the derivative parameter \(a_i\) and on the additional breakpoint \(\xi_i\) (if this is needed). This makes the construction flexible. The remaining freedom can be used for example to get the solution of the HIP to simulate some additional information known about the situation which produced the data. Such information may be, for instance, the norm of the function.

References