PRE-PRINT
UD 2014-01

On the non-differentiability of Hyman Monotonicity constraint

Michele Maio[*]
and
Jorrit-Jaap de Jong[†]

uGly Duckling B.V.

Burgemeester le Fevre de Montignylaan 30,
3055LG Rotterdam, the Netherlands

October 21, 2014

Abstract

In this paper we describe some new features of Hyman's monotonicity constraint, which is implemented into various cubic spline interpolation methods. We consider the problem of understanding how sensitive such methods are to small changes of the input y-values and, in particular, how relevant Hyman's constraint is with respect to such changes. We find that many things cancel out and that eventually Hyman's constraint can be safely omitted when the monotone-preserving cubic spline is used. We also find that consistency requires including some specific boundary conditions that become relevant for special values of the parameter space.

Keywords: Yield curve, interpolation, monotone preserving cubic splines.

[*] m.maio@uglyduckling.nl
[†] j.jong@uglyduckling.nl

Contents

1 Introduction

In this paper we consider one-dimensional interpolation methods that preserve the trend of the input data points. Many standard papers and books exist on interpolation, and an overview of the most common methods can be found for example in the Numerical Recipes book [1], or in the papers by Fritsch and Carlson [2], Hagan and West [3], or in specialized articles such as the one by Lehmann, Gonner and Spitzer [4]. Preserving the trend of the data points, or in short preserving the monotonicity, is an important issue in any interpolation procedure. In fact, often unwanted and unphysical oscillating behaviour between consecutive points is introduced into the interpolating solution as a spurious effect of an ineffective method. In order to avoid such spurious contributions, it is crucial to guarantee that the interpolant follows the behaviour of the data points.

The main monotonicity-preserving algorithm was introduced by Hyman [5] and can be implemented in principle on almost any cubic-spline-like interpolation method. In fact, a large class of cubic splines is defined by specifying the vector of derivatives at the input points. This allows us to have various types of splines, such as the Bessel-Hermite splines, where the vector of derivatives at each point is computed by using the slope of the parabola passing through one point before, the point itself, and one point after (the first three values for the derivative at the first point, and the last three values for the derivative at the last point), or the monotone preserving spline where Hyman's constraint is used on top of the Fritsch-Butland prescription [2, 6, 7]. Even if in principle one can implement Hyman's constraint on the Bessel-Hermite splines, in this paper we will focus on the monotone preserving splines only. A good reference with a survey of other types of splines, together with some of other popular methods, is Hagan and West [3][1]. We will use this reference often throughout this paper and borrow their notation.

In many practical applications, we are interested in the behavior of the specific interpolation method under small changes of the input data. In particular we focus on changing the y-values of the input data. Changing the input points, even if of a small amount, does change the interpolated value corresponding to x-values away from the input points. Using some notation that will be explained later, the quantity that will be responsible for describing such a behaviour in our case is simply

$$\frac{\partial f(x)}{\partial f_j} \tag{1}$$

which is the derivative of the interpolant with respect to the y-values of the input data. Mathematically, this quantity is a measure of the quality of the interpolation method in terms of sensitivity and stability.

In order to provide an example where such a problem becomes relevant, let us consider a financial application. When dealing with interest rated derivatives, quantities such as the PV01 (present value of one basis point[2]) or the DV01 (dollar value of one basis point) are used for risk management purposes. In these cases, the input y-values are the interest rates r_i given for specific maturities t_i (the x-values), and the associated continuous curve $r(t)$ is known as the yield curve. The yield curve is the very basis for computing prices of financial products, and it is crucial to have a proper interpolation method to estimate it. Quantities such as the PV01 are defined exactly in terms of the derivative

$$\frac{\partial r(t)}{\partial r_j}$$

of the yield curve with respect to the input rate r_j. This derivative also enters the calculation of the hedging positions for such instruments, hence this is a crucial quantity for many financial institutions. This situation is discussed in many details in [10].

This kind of calculation is often done numerically, mainly because the numerical version is easy to implement while the analytic derivation is cumbersome. However, the numerical version is only an approximation of the analytic solution, and when large numbers are involved the differences between the two approaches become important. Moreover, by working out all the formulas, we were able to find our results.

Our main result is that Hyman's constraint does not contribute to any calculation in the monotone-preserving cubic spline framework. This is important for various reasons. From the practical point of view, this makes calculations much easier and much shorter and the constraint redundant in this framework. From the conceptual point of view, Hyman's recipe includes non-differentiable functions (such as the max and min functions) whose discontinuous derivatives enter the calculation.

This paper is organized as follows. In section 2, we define our set up and fix our notation, which follows [3]. Section 3 contains the main result. We show explicitly how Hyman's monotonicity constraint is redundant in the monotone-preserving cubic spline method. Moreover, we will see that an additional boundary condition must be included in

[1] The papers [3] and [8] have become quite famous within the mathematical finance community, since their author have introduced a new interpolation method which is deeply intertwined with interest rates, forward rates, and discount factors. As noted e.g. in [9], this method sometimes produces discontinuous forward rates.

[2] 1 basis point = 0.01%.

order to get the correct answer. In section 4, we carry out some numerical checks and find full agreement between the numerical approximation and the analytic formulas. Section 5 contains a summary of the work done and some conclusions. Finally in Appendix A, we add a few remarks about the min and max functions.

2 Set up and notation

In this section we define our set up and fix our notation for the monotone preserving cubic splines. The main reference are [3] and [8]. Suppose we are given a mesh of data points $\{t_1, t_2, \ldots, t_n\}$ (we will think of the x-values as times) and corresponding values $\{f_1, f_2, \ldots, f_n\}$ for a generic but unknown function $f(t)$. Cubic splines are generically defined by piecewise cubic polynomial that pass through consecutive points:

$$f(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3\,, \tag{2}$$

with $t \in [t_i, t_{i+1}]$ and $i = 1, \ldots, n$. We will use the following definitions:

$$h_i = t_{i+1} - t_i \tag{3}$$

$$m_i = \frac{f_{i+1} - f_i}{h_i}\,, \tag{4}$$

with $i = 1, \ldots, n-1$. The coefficients a_i, b_i, c_i, and d_i, depends on the details of the method, and they are related to the values of $f(t)$ and its derivatives at the node points. In general,

$$a_i = f(t_i) \equiv f_i\,, \qquad b_i = f'(t_i)\,, \qquad \text{etc.} \tag{5}$$

where the prime denotes the derivative of the interpolating function $f(t)$ w.r.t. its argument t. In particular, the a_i coefficients are always determined by the input points. Moreover, given a_i and b_i, we can express c_i and d_i as follows:

$$c_i = \frac{3m_i - b_{i+1} - 2b_i}{h_i} \tag{6}$$

$$d_i = \frac{b_{i+1} + b_i - 2m_i}{h_i^2}\,. \tag{7}$$

We can use (2) to compute the derivative

$$\frac{\partial f(t)}{\partial f_j} = \frac{\partial a_i}{\partial f_j} + \frac{\partial b_i}{\partial f_j}(t - t_i) + \frac{\partial c_i}{\partial f_j}(t - t_i)^2 + \frac{\partial d_i}{\partial f_j}(t - t_i)^3\,. \tag{8}$$

We have

$$\frac{\partial a_i}{\partial f_j} = \delta_i^j \tag{9}$$

$$\frac{\partial m_i}{\partial f_j} = \frac{1}{h_i}\delta_{i+1}^j - \frac{1}{h_i}\delta_i^j \tag{10}$$

$$\frac{\partial c_i}{\partial f_j} = \frac{1}{h_i}\left(3\frac{\partial m_i}{\partial f_j} - \frac{\partial b_{i+1}}{\partial f_j} - 2\frac{\partial b_i}{\partial f_j}\right) \tag{11}$$

$$\frac{\partial d_i}{\partial f_j} = \frac{1}{h_i^2}\left(\frac{\partial b_{i+1}}{\partial f_j} + \frac{\partial b_i}{\partial f_j} - 2\frac{\partial m_i}{\partial f_j}\right)\,, \tag{12}$$

which depends on $\frac{\partial b_i}{\partial f_j}$ and $\frac{\partial b_{i+1}}{\partial f_j}$. Here δ_i^j is the Kronecker delta, which is equal to one if $i = j$ and zero otherwise.

Due to our previous formulas, once the derivatives at the points, or equivalently the b_i coefficients, are specified, everything else is fixed. In particular, if we are interested in computing $\frac{\partial f(t)}{\partial f_j}$, then all the work will be in the calculation of the derivatives of b_i w.r.t. f_j. By locality of the method, we will also expect that if j is far away from i then this derivative will vanish. We will soon see that this is indeed the case. This calculation is however tricky if we use monotone preserving splines (or any other method which enforces monotonicity a-la-Hyman [5]), where the b_i's are non-differentiable functions of the f_j's (which involve the min and max functions). Let us consider this case in more detail.

2.1 Monotone preserving cubic splines

Let us start by recalling the formulas for the b_i's in the monotone preserving cubic spline method as defined in the Hagan-West paper [3]. First of all

$$b_1 = 0\,, \qquad b_n = 0\,. \tag{13}$$

If the curve is not monotone at t_i, i.e. $m_{i-1} \cdot m_i \leq 0$, then

$$b_i = 0 \qquad (\text{if } m_{i-1} \cdot m_i \leq 0)\,, \tag{14}$$

so that it will have a turning point there.

Note that we have included the case where either m_{i-1} or m_i vanishes in the non-monotone case. We could have included it in the monotone case as well, but we would have found $b_i = 0$ anyway.

The monotone case is $m_{i-1} \cdot m_i > 0$. In this situation we define

$$\beta_i = \frac{3m_{i-1} \cdot m_i}{\max(m_{i-1}, m_i) + 2\min(m_{i-1}, m_i)} \tag{15}$$

and

$$b_i = \begin{cases} \min\left(\max(0, \beta_i), 3\min(m_{i-1}, m_i)\right) & \text{if } m_{i-1},\, m_i > 0 \\ \max\left(\min(0, \beta_i), 3\max(m_{i-1}, m_i)\right) & \text{if } m_{i-1},\, m_i < 0 \end{cases} \tag{16}$$

The former choice is made when the curve is increasing, the latter when decreasing. Equation (16) represents the monotonicity constraint introduced by Hyman [5] and based on the Fritsch-Butland algorithm [2, 6, 7]. Going back to our problem of computing $\frac{db_i}{df_j}$, we now face a problem: the min and max functions are not differentiable (an explicit and pedagogical example is given in appendix A). We will consider this calculation now explicitly and we will find that this problem is solved by some cancellations that happen thanks to the very definition of the Hyman monotonicity constraint.

3 The main results

In this section we carry out explicitly the calculation $\frac{\partial b_i}{\partial f_j}$. Later on we will show that similar remarks apply to the b_i coefficients. In the case of non-monotonic trend the result is automatically zero and the β_i's are not even defined. So we will focus on case of monotonic trend from now on. The main findings are the following:

- the only non-zero derivatives are

$$\frac{\partial b_i}{\partial f_{i-1}}\,, \qquad \frac{\partial b_i}{\partial f_i}\,, \qquad \frac{\partial b_i}{\partial f_{i+1}}\,, \tag{17}$$

  and all the others vanish if $j \neq i-1,\, i,\, i+1$;

- if m_i is equal to m_{i-1}, then when computing the derivative of (15) we must average the derivatives of m_i and m_{i-1}

$$\frac{\partial}{\partial f_j}\max(m_{i-1}, m_i) = \frac{1}{2}\left(\frac{\partial m_i}{\partial f_j} + \frac{\partial m_{i-1}}{\partial f_j}\right) \tag{18a}$$

$$\frac{\partial}{\partial f_j}\min(m_{i-1}, m_i) = \frac{1}{2}\left(\frac{\partial m_i}{\partial f_j} + \frac{\partial m_{i-1}}{\partial f_j}\right) \tag{18b}$$

  if $m_i = m_{i-1}$;

- the monotonicity constraint does not enter the calculation of the derivative, which turns out to be given only by the derivative of (15)

$$\frac{\partial b_i}{\partial f_j} = \frac{\partial \beta_i}{\partial f_j}\,, \tag{19}$$

  for any i and j.

The actual statement to prove is the last one. In fact, once we have showed that it holds true, we can easily convince ourselves of the first statement which is then simple to prove by explicit calculation. The second statement is nothing else than a boundary condition that we need to implement in order to get the correct answer out of an ill-defined situation, while the first one is just about the locality of the method. The remaining of this section is dedicated to proving our last statement.

3.1 The proof of the main results

In order to be as clear as possible let us make this calculation in small steps corresponding to the various building blocks. The stepping stones of the full calculation are the expressions (20)-(23) for $\frac{\partial \beta_i}{\partial f_j}$ and Theorem 2 which are given later.

3.1.1 Before the monotonicity constraint

The first step is to compute the the derivatives of the max and min that appear in the denominator of (15)

$$\frac{\partial}{\partial f_j}\max(m_{i-1}, m_i) \;=\; \begin{cases} \frac{\partial m_{i-1}}{\partial f_j} & \text{if } m_{i-1} > m_i \\ \frac{\partial m_i}{\partial f_j} & \text{if } m_{i-1} < m_i \end{cases}$$

$$\frac{\partial}{\partial f_j}\min(m_{i-1}, m_i) \;=\; \begin{cases} \frac{\partial m_{i-1}}{\partial f_j} & \text{if } m_{i-1} < m_i \\ \frac{\partial m_i}{\partial f_j} & \text{if } m_{i-1} > m_i \end{cases}$$

It is now clear that if $j \neq i-1,\ i,\ i+1$ these derivatives above all vanish. These expressions can be used to compute $\frac{\partial \beta_i}{\partial f_j}$. If we denote the denominator by

$$L \equiv \max(m_{i-1}, m_i) + 2\min(m_{i-1}, m_i)\,,$$

by explicit calculation we find:

- $\frac{\partial \beta_i}{\partial f_i}$

$$\frac{\partial \beta_i}{\partial f_i} = k \cdot \frac{3}{L^2} \cdot \left(\frac{(\max(m_{i-1}, m_i))^2}{h_{\min}} - \frac{2\,(\min(m_{i-1}, m_i))^2}{h_{\max}} \right) ,$$

where

$$k\,, h_{\min}\,, h_{max} = \begin{cases} k = -1\,, h_{\min} = h_i\,, h_{\max} = h_{i-1} & \text{if } m_{i-1} > m_i \\ k = 1\,, h_{\min} = h_{i-1}\,, h_{\max} = h_i & \text{if } m_{i-1} < m_i\,, \end{cases}$$

or explicitly

$$\frac{\partial \beta_i}{\partial f_i} = \frac{3}{L^2} \cdot \begin{cases} \frac{2m_i^2}{h_{i-1}} - \frac{m_{i-1}^2}{h_i} & \text{if } m_{i-1} > m_i \\ \frac{m_i^2}{h_{i-1}} - \frac{2m_{i-1}^2}{h_i} & \text{if } m_{i-1} < m_i\,. \end{cases} \tag{20}$$

- $\frac{\partial \beta_i}{\partial f_{i-1}}$

$$\frac{\partial \beta_i}{\partial f_{i-1}} = -k \cdot \frac{3}{L^2} \cdot \frac{m_i^2}{h_{i-1}} ,$$

where

$$k = \begin{cases} 2 & \text{if } m_{i-1} > m_i \\ 1 & \text{if } m_{i-1} < m_i\,, \end{cases}$$

or explicitly

$$\frac{\partial \beta_i}{\partial f_{i-1}} = -\frac{3}{L^2} \cdot \frac{m_i^2}{h_{i-1}} \cdot \begin{cases} 2 & \text{if } m_{i-1} > m_i \\ 1 & \text{if } m_{i-1} < m_i\,. \end{cases} \tag{21}$$

- $\frac{\partial \beta_i}{\partial f_{i+1}}$

$$\frac{\partial \beta_i}{\partial f_{i+1}} = k \cdot \frac{3}{L^2} \cdot \frac{m_{i-1}^2}{h_i} ,$$

where

$$k = \begin{cases} 1 & \text{if } m_{i-1} > m_i \\ 2 & \text{if } m_{i-1} < m_i\,, \end{cases}$$

or explicitly

$$\frac{\partial \beta_i}{\partial f_{i+1}} = \frac{3}{L^2} \cdot \frac{m_{i-1}^2}{h_i} \cdot \begin{cases} 1 & \text{if } m_{i-1} > m_i \\ 2 & \text{if } m_{i-1} < m_i\,. \end{cases} \tag{22}$$

- $\frac{\partial \beta_i}{\partial f_j}$

$$\frac{\partial \beta_i}{\partial f_j} = 0\,, \qquad \text{if} \quad j \neq i-1, i, i+1\,. \tag{23}$$

In any case, if $m_{i-1} = m_i$, then $\frac{\partial \beta_i}{\partial f_j}$ will be the average of the answers corresponding to the two options $m_{i-1} < m_i$ and $m_{i-1} > m_i$.

3.1.2 After the monotonicity constraint

We can now move to the monotonicity constraint. What we need first is

$$\begin{aligned}
\frac{\partial}{\partial f_j}\max(0,\beta_i) &= \begin{cases} 0 & \text{if } \beta_i < 0 \Leftrightarrow m_{i-1}, m_i < 0 \\ \frac{\partial \beta_i}{\partial f_j} & \text{if } \beta_i > 0 \Leftrightarrow m_{i-1}, m_i > 0 \end{cases} \\
\frac{\partial}{\partial f_j}\min(0,\beta_i) &= \begin{cases} 0 & \text{if } \beta_i > 0 \Leftrightarrow m_{i-1}, m_i > 0 \\ \frac{\partial \beta_i}{\partial f_j} & \text{if } \beta_i < 0 \Leftrightarrow m_{i-1}, m_i < 0 \end{cases}
\end{aligned}$$

The final step consists of putting all the information together to compute $\frac{\partial b_i}{\partial f_j}$.

Suppose first that the trend is increasing, i.e. $m_{i-1}, m_i, \beta_i > 0$. Using (16) we find:

$$\begin{aligned}
\frac{\partial b_i}{\partial f_j} &= \frac{\partial}{\partial f_j}\min\left(\max(0,\beta_i), 3\min(m_{i-1},m_i)\right) \\
&= \begin{cases} \frac{\partial}{\partial f_j}\max(0,\beta_i) & \text{if } \max(0,\beta_i) < 3\min(m_{i-1},m_i) \\ 3\frac{\partial}{\partial f_j}\min(m_{i-1},m_i) & \text{if } \max(0,\beta_i) > 3\min(m_{i-1},m_i) \end{cases} \\
&= \begin{cases} \frac{\partial \beta_i}{\partial f_j} & \text{if } \beta_i < 3\min(m_{i-1},m_i) & \text{(i)} \\ u(i,j) & \text{if } \beta_i > 3\min(m_{i-1},m_i) & \text{(ii)} \end{cases}
\end{aligned} \tag{24}$$

where in the last step we have used the fact that β_i is positive and $u(i,j)$ is a function that depends on the specific values of i and j. Even if the function $u(i,j)$ is known and is trivially given by our previous intermediate formulas, we will not write it explicitly because -as we will see in a moment- it never contributes to the final answer.

Let us now suppose that the trend is decreasing instead, i.e. $m_{i-1}, m_i, \beta_i < 0$. By (16) we have:

$$\begin{aligned}
\frac{\partial b_i}{\partial f_j} &= \frac{\partial}{\partial f_j}\max\left(\min(0,\beta_i), 3\max(m_{i-1},m_i)\right) \\
&= \begin{cases} \frac{\partial}{\partial f_j}\min(0,\beta_i) & \text{if } \min(0,\beta_i) > 3\max(m_{i-1},m_i) \\ 3\frac{\partial}{\partial f_j}\max(m_{i-1},m_i) & \text{if } \min(0,\beta_i) < 3\max(m_{i-1},m_i) \end{cases} \\
&= \begin{cases} \frac{\partial \beta_i}{\partial f_j} & \text{if } \beta_i > 3\max(m_{i-1},m_i) & \text{(iii)} \\ v(i,j) & \text{if } \beta_i < 3\max(m_{i-1},m_i) & \text{(iv)} \end{cases}
\end{aligned} \tag{25}$$

where in the last step we have used the fact that β_i is negative and $v(i,j)$ is a known function of i and j which will not contribute to the final answer.

Formulas (24) and (25) depend on which region of the parameter space the parameters belong to. However, it is possible to show, and we will do it in the next subsection, that some cancellations occur when checking the boundaries between the regions i) and ii) as well as between the regions iii) and iv). As a consequence, the final result simplifies and assumes the more general and better-looking expression

$$\frac{\partial b_i}{\partial f_j} = \frac{\partial \beta_i}{\partial f_j}, \tag{26}$$

which is always valid. This is the main statement of Theorem 2, which we will prove in the following subsection. This formula, together with (20)-(23), completely solves our problem.

3.2 The theorem

In this section we want to show that the conditions that define region ii) in (24) and region iv) in (25) are never satisfied independently of the values of m_{i-1} and m_i. We will start by proving the following theorem.

Theorem 1. *Given β_i as defined in (15), the following statements hold both true*

- *if m_{i-1}, $m_i > 0$, then $\beta_i < 3\min(m_{i-1}, m_i)$;*
- *if m_{i-1}, $m_i < 0$, then $\beta_i > 3\max(m_{i-1}, m_i)$.*

Proof. The proof is straightforward.

Consider the monotonically increasing case first, m_{i-1}, $m_i > 0$. By working out the inequality

$$\beta_i < 3\min(m_{i-1}, m_i)$$

and using the positivity of the denominator in (15), we find:

$$\begin{aligned} \frac{3m_{i-1}m_i}{\max(m_{i-1}, m_i) + 2\min(m_{i-1}, m_i)} &< 3\min(m_{i-1}, m_i)\,; \\ 0 &< 2\left(\min(m_{i-1}, m_i)\right)^2\,, \end{aligned} \tag{27}$$

where we have used the fact that $\max(m_{i-1}, m_i) \cdot \min(m_{i-1}, m_i) = m_{i-1}m_i$. Eq. (27) is clearly always true.

Similarly, for the monotonically decreasing case m_{i-1}, $m_i < 0$, by working out the inequality

$$\beta_i > 3\max(m_{i-1}, m_i)$$

and using the negativity of the denominator in (15), we find:

$$\begin{aligned} \frac{3m_{i-1}m_i}{\max(m_{i-1}, m_i) + 2\min(m_{i-1}, m_i)} &> 3\max(m_{i-1}, m_i)\,; \\ 0 &< \left(\max(m_{i-1}, m_i)\right)^2 + m_{i-1}m_i\,, \end{aligned} \tag{28}$$

where we have used again the fact that $\max(m_{i-1}, m_i) \cdot \min(m_{i-1}, m_i) = m_{i-1}m_i$. Eq. (28) is clearly always true. □

As we wanted to show, the main consequence of Theorem 1 is that regions ii) in (24) and region iv) in (25) are never reached by any value of the parameters. This result is important enough to give it its own theorem:

Theorem 2 (Main Result)**.** *In the monotone-preserving cubic spline interpolation method that uses Hyman monotonicity constraint (15)-(16), if the data trend is locally monotonic at node i, then we have*

$$\frac{\partial b_i}{\partial f_j} = \frac{\partial \beta_i}{\partial f_j} \tag{29}$$

else

$$\frac{\partial b_i}{\partial f_j} = 0\,. \tag{30}$$

Proof. The proof follows immediately from equations (24) and (25), and Theorem 1. □

3.3 Alternative derivation

So far we have worked at the level of derivatives and proved that, with the β_i's defined as in (15), the derivative of the β_i's with respect to the input nodes are equal to the derivatives of the b_i's. However, we can get to the same result if we work directly at the level of the b_i and β_i coefficients instead of their derivatives[3]. The reasoning is similar to the previous derivation, but shifted back one step earlier. In such a way we can make our proof more straightforward.

The starting point is the following theorem:

Theorem 3. *In the monotone-preserving cubic spline interpolation method that uses Hyman monotonicity constraint (15)-(16), if the data trend is locally monotonic at node i, then we have*

$$b_i = \beta_i \tag{31}$$

else

$$b_i = 0\,. \tag{32}$$

Proof. This follows from a few observations, on the same lines of Theorem 2:

1. when the trend is increasing at node i, then $\beta_i > 0$, while when the trend is decreasing then $\beta_i < 0$;
2. the four regions are still defined as before;
3. region ii) and iv) are still empty.

Let's use these observation into the constraint (16). We have:

$$\begin{aligned} b_i &= \begin{cases} \min\left(\max(0,\beta_i), 3\min(m_{i-1}, m_i)\right) & \text{if } m_{i-1},\, m_i > 0 \\ \max\left(\min(0,\beta_i), 3\max(m_{i-1}, m_i)\right) & \text{if } m_{i-1},\, m_i < 0 \end{cases} && (33)\\ &= \begin{cases} \min\left(\beta_i, 3\min(m_{i-1}, m_i)\right) & \text{if } m_{i-1},\, m_i > 0 \\ \max\left(\beta_i, 3\max(m_{i-1}, m_i)\right) & \text{if } m_{i-1},\, m_i < 0 \end{cases} && (34)\\ &= \begin{cases} \beta_i & \text{if } m_{i-1},\, m_i > 0 \\ \beta_i & \text{if } m_{i-1},\, m_i < 0 \end{cases} && (35)\\ &= \beta_i\,. && (36) \end{aligned}$$

In going from (33) to (34) we have used observation 1, while from (34) to (35) we have used observations 2 and 3 as well as the definitions of the four regions. Hence, in the case of local monotonicity we are left with

$$b_i = \beta_i\,. \tag{37}$$

□

Therefore, the whole Hyman constraint (16) is completely superfluous in the monotone-preserving spline and the identity between the derivatives follows trivially:

$$\frac{\partial b_i}{\partial f_j} = \frac{\partial \beta_i}{\partial f_j}\,. \tag{38}$$

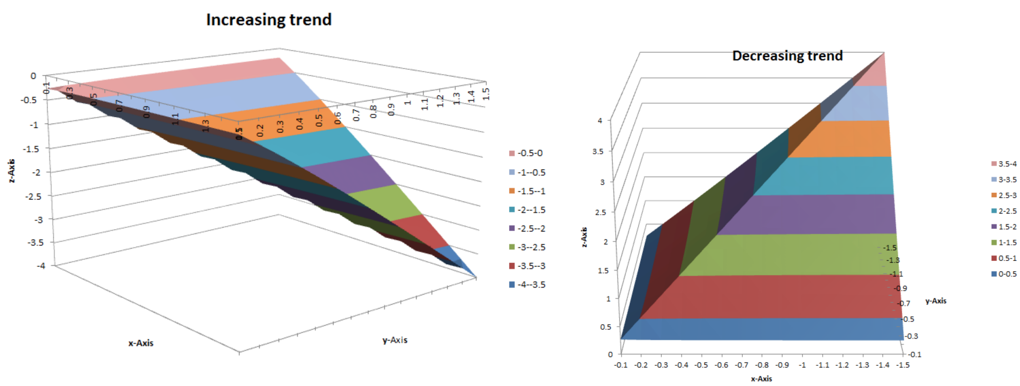
[3]We thank Dmytro Makogon for pointing out this to us.

As a general remark, the Hyman constraint (16) is trivial when the β_i's are defined by the Fritsch-Butland algorithm [7] as in (15). This is what happens for example in the notation of Hagan and West [3]. The reason for this identity is the fact that the β_i's as defined by the Fritsch-Butland algorithm already guarantee that the resulting curve will be monotonic. However, one can still implement the Hyman constraint (16) to ensure monotonicity in a non-trivial way when different first derivatives β_i at the node points are chosen, and in that case the identities (37) and (38) will not hold anymore.

4 Numerical checks

In this section we would like to present some evidence on the correctness of our approach. Recall from section 3 that our main statements are about locality, boundary conditions, and actual calculation of the derivative $\frac{\partial f}{\partial f_j}$. We will address each of them here.

First of all, let us show numerically that the boundary conditions for region ii) and iv) are never satisfied. Figure 1 show the behaviour of the function $z(x, y)$ defined by



(a) Increasing Trend (b) Decreasing Trend

Figure 1:
(a) This plot shows that $z(x, y) = \beta(x, y) - 3\min(x, y) < 0$, hence condition $i)$ in (24) is always verified.
(b) Similarly this plot shows that $z(x, y) = \beta(x, y) - 3\max(x, y) > 0$, hence condition $iii)$ in (25) is always verified.

$$z(x, y) = \beta(x, y) - 3\min(x, y) \tag{39}$$

in case of increasing trend (sub-figure 1a, x and y are positive), and

$$z(x, y) = \beta(x, y) - 3\max(x, y) \tag{40}$$

in case of decreasing trend (sub-figure 1b, x and y are negative). Here we have defined

$$\beta(x, y) = \frac{3xy}{\max(x, y) + 2\min(x, y)}, \tag{41}$$

and $(x,\, y)$ play the role of $(m_{i-1},\, m_i)$. Moreover, everything is symmetric under the exchange of x and y, according to the original Butland's idea [2, 6, 7]. As we can see from the plots, when x and y are positive (increasing trend) $z(x, y)$ is always negative, or equivalently $\beta(x, y) < 3\min(x, y)$, while when x and y are negative (decreasing trend) $z(x, y)$ is always positive, or equivalently $\beta(x, y) > 3\max(x, y)$. Hence, the numerical check confirms what was already proved in Theorem 1 and 2.

Let us now move on to the check of the method itself. We have implemented the monotone preserving cubic spline method into our library by using numerical methods and analytic formulas and compared the results with the numerical calculation.

From the analytic side, we use Theorem 2 and formulas (20)-(23) to determine the value of the derivative $\frac{\partial f}{\partial f_j}$.

From the numerical side, we compute the derivative $\frac{\partial f}{\partial f_j}$ by following standard procedures:

- we bump the input data point f_j by a small amount $+h$,
- we bump the input data point f_j by a small amount $-h$,
- we define a new interpolator object initialized with input rates $(f_1, \ldots, f_j+h, \ldots, f_n)$, which allows us to compute the interpolated function for any value of t, say $f(t; f_j + h)$
- we define a new interpolator object initialized with input rates $(f_1, \ldots, f_j-h, \ldots, f_n)$, which allows us to compute the interpolated function for any value of t, say $f(t; f_j - h)$
- we compute the incremental ratio

$$\frac{\partial f}{\partial f_j}(t) = \frac{f(t; f_j + h) - f(t; f_j - h)}{2h} .$$

This procedure gives an error on the numerical approximation of order $O(h^2)$.

The two operations are independent of each other and the agreement is within a very high accuracy.

In order to deeply understand the nature of such an agreement between the numerical and the analytic derivative, we need to stress some crucial points:

- the agreement holds either with or without the Hyman monotonicity constraint (16): this represents a direct numerical check of Theorem 2, and tells us indeed that Hyman constraint does not play any role in the computation of $\frac{\partial f}{\partial f_j}$;
- the agreement only works if we include the contribution (18) coming from the boundary conditions when $m_{i-1} = m_i$;
- the agreement shows that the method is local as expected.

Finally, as far as Theorem 3 is concerned, it is pretty straightforward to test whether the Hyman constraint within the monotone-preserving spline is redundant or not, since one just has to comment out the relevant lines of code. Clearly this test is also positive.

5 Summary and conclusions

In this paper we have consider monotone preserving interpolation methods and shown that Hyman's monotonicity constraint does not contribute within the framework of the monotone-preserving cubic spline.

However this is generically not true anymore when the Hyman constraint is applied on any other spline in order to construct a monotonic curve. In fact, in this case Theorem 1 does not need to hold and consequently regions ii) and iv) will in general contribute. In this case one will need to use to complete formulas for all the four regions.

This result is important for practical as well as conceptual reasons. Moreover this quantity is important in many areas (e.g. in finance for pricing and for risk management of interest rate derivatives). Often such a calculation is done numerically, but it is worth the effort to get it correctly. As it turns out, Hyman's constraint is not relevant when we use the monotone-preserving spline. This result has been derived in the body of the paper (in particular, in Theorem 1, Theorem 2 and Theorem 3) and checked against numerical tests.

Acknowledgements

We would like to thank Patrick Hagan for email exchanges, the ABN AMRO Model Validation team for providing us with feedback and Dmytro Makogon in particular for his useful comments and remarks. We are also grateful with the Lawrence Livermore National Laboratory for providing us with copies of the their reports.

References

[1] W. H. Press , S. A. Teukolsky , W. T. Vetterling , B. P. Flannery (2007), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Third Edition, 1256 pp., ISBN-10: 0521880688 http://www.nr.com/.

[2] F. N. Fritsch, R. E. Carlson (1980), *Monotone Piecewise Cubic Interpolation*, SIAM J. Numer. Anal., 17 (1980) 238-246 http://epubs.siam.org/doi/abs/10.1137/0717021.

[3] P. S. Hagan, G. West (2006), *Interpolation Methods for Curve Construction*, Applied Mathematical Finance, 13: 2, 89 - 129 http://dx.doi.org/10.1080/13504860500396032, http://finmod.co.za/Hagan_West_curves_AMF.pdf.

[4] T. M. Lehmann, C. Gonner, K. Spitzer (1999), *Survey: Interpolation Methodsin Medical Image Processing*, IEEE Transactions on Medical Imaging, Vol. 18, No. 11, 1049 - 1075 http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=816070&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D816070.

[5] J. M. Hyman (1983), *Accurate Monotonicity Preserving Cubic Interpolation*, SIAM J. Sci. Stat. Comput. Vol. 4, No. 4, 645 - 654 http://www.researchgate.net/publication/230676072_Accurate_monotonicity_preserving_cubic_interpolation/file/d912f50c0a8c9585cd.pdf.

[6] J. Butland (1980), *A Method of Interpolating Reasonable-Shaped Curves Through Any Data*, Proceedings of Computer Graphics 80, Online Publications Ltd, 409-422.

[7] F. N. Fritsch, J. Butland (1980), *An Improved Monotone Piecewise Cubic Interpolation Algorithm*, Lawrence Livermore National Laboratory preprint UCRL-85104.

[8] P. S. Hagan, G. West (2008), *Methods for Constructing a Yield Curve*, Wilmott Magazine, 70 - 81 http://www.math.ku.dk/~rolf/HaganWest.pdf.

[9] P. F. du Preez, E. Mare (2013), *Interpolating Yield Curve Data in a Manner that Ensures Positive and Continuous Forward Curves*, SAJEMS NS 16 (2013) No 4:395-406 http://sajems.org/index.php/sajems/article/view/388.

[10] M. Maio, J. de Jong (2014), *Geometry of Interest Rate Risk*, pre-print nr UD2014-02, http://uglyduckling.nl/library_files/PRE-PRINT-UD-2014-02.pdf.

Appendix

A Min and max are not differentiable

In this appendix we consider an explicit example to show that the min and max functions are not differentiable. We take an example from interest rates in finance. Consider figure 2. It plots two arbitrary yield curves representing the zero rates and the forward rates for some given data[4]. Forward rates $f(t)$ are related to zero rate $r(t)$ by the simple formula

$$f(t) = \frac{d}{dt}\left(t \cdot r(t)\right) , \tag{42}$$

where t represents the time to maturity. Figure 3 plots the min (in red) and max (in blue) of the same functions. Observe that at the points t^* when the two curves are equal,

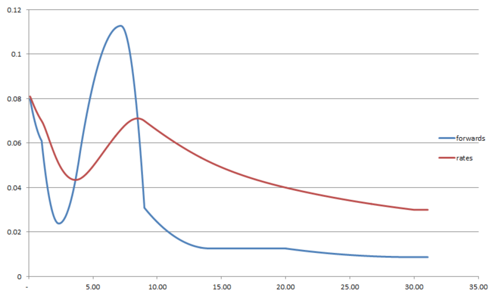


Figure 2: Rate and forward functions

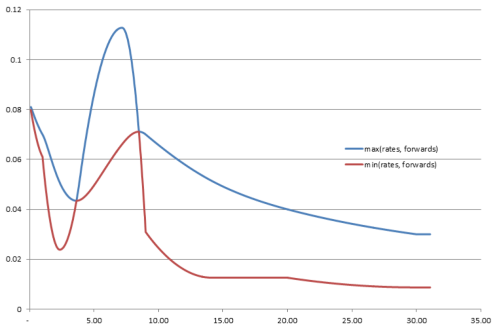


Figure 3: Min and max functions

i.e. $r(t^*) = f(t^*)$, the min and max functions are not differentiable:

$$\frac{d}{dt}\max(r(t), f(t)) = \begin{cases} r'(t) & \text{if } r(t) > f(t) \\ f'(t) & \text{if } r(t) < f(t) \end{cases} \tag{43}$$

and

$$\frac{d}{dt}\min(r(t), f(t)) = \begin{cases} r'(t) & \text{if } r(t) < f(t) \\ f'(t) & \text{if } r(t) > f(t) \end{cases} \tag{44}$$

This is of course always true: in general, for arbitrary functions $f_1(x)$ and $f_2(x)$, the derivatives of $\max(f_1, f_2)$ and $\min(f_1, f_2)$ have a discontinuity at the points x^* where $f_1(x^*) = f_2(x^*)$.

[4]The data points are as in [3].