

FITTING CONSTRAINED CONTINUOUS SPLINE CURVES

¹V. P. Kong, ²B. H. Ong

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

e-mail: ¹kongvp@hotmail.com, ²bhong@cs.usm.my

Abstract. *Fitting a curve through a set of planar data which represents a positive quantity requires that the curve stays above the horizontal axis. The more general problem of designing parametric and non-parametric curves which do not cross the given constraint boundaries is considered. Several methods will be presented.*

1 Introduction

There are a multitude of techniques for producing interpolating curves to a given set of planar data. One is often concerned with certain properties inherent in the data or qualitative mathematical properties. For example, the shape properties like monotonicity, convexity, positivity, and some degree of parametric or geometric smoothness like the continuous variation of tangent lines or curvatures are among the list of concern.

The need of positivity preserving interpolation could arise. When the data represent a non-negative quantity like the probability density or material concentration, negative values of an interpolant to these data would be physically meaningless, and so a non-negative interpolant is required. The problem of positivity preserving interpolation or more generally range restricted interpolation, where the movement of the interpolant is restricted by constraint boundaries, has been investigated in a number of papers. We shall consider the different techniques and approaches to this problem. In principle, two different approaches have been used to address this problem, the direct approach and the variational approach. In the direct approach, a suitable class of functions depending on a finite number of parameters is chosen. Conditions on these parameters are derived so that the preservation of positivity is ensured. Then an interpolant which satisfies these conditions for positivity is computed from the class of functions. In the second approach, the interpolant is determined as the solution of a convex constrained minimization problem with respect to a subset of an infinite dimensional function space.

Let $\{(x_i, y_i) : 0 \leq i \leq n\}$ be given data points, where $x_0 < x_1 < \dots < x_n$ and the ordinates are nonnegative, i.e. $y_i \geq 0$, $0 \leq i \leq n$. Are there smooth interpolants f which are nonnegative on $[x_0, x_n]$ with $f(x_i) = y_i$, $0 \leq i \leq n$? In the more general context, given data points lying on one side of given constraint curves, we are interested in exploring several constructions of a smooth parametric or non-parametric interpolating curve which lies on the same side of the constraint curves as the data, where the constraints may be straight lines or quadratic curves.

2 Variational methods

(Schmidt & Hess, 1988) considered the space of cubic C^1 splines and proved that these splines make the problem of positivity preserving interpolation always solvable. In fact there exists an infinite number of feasible spline interpolants. They derived a necessary and sufficient criterion for the positivity of a cubic polynomial on a given interval as described in the theorem below.

Theorem 1. Suppose that $x_0 < x_1 < \dots < x_n$ and $y_i \geq 0$, $0 \leq i \leq n$. Let $f(x)$ be the cubic C^1 spline on $[x_0, x_n]$ with $f(x_i) = y_i$, $0 \leq i \leq n$. Suppose also that $f'(x_i) = m_i$, $0 \leq i \leq n$. Then the polynomial f is nonnegative on $[x_0, x_n]$ if and only if

$$(m_i, m_{i+1}) \in W_i, \quad 0 \leq i \leq n-1,$$

$$\text{where } W_i = \left\{ (a_i, b_i) : a_i \geq -\frac{3y_i}{h_i}, b_i \leq \frac{3y_{i+1}}{h_i} \right\} \cup$$

$$\left\{ (a_i, b_i) : 36y_i y_{i+1} (a_i^2 + a_i b_i + b_i^2 - 3\Delta_i (a_i + b_i) + 3\Delta_i^2) \right. \\ \left. + 3(y_{i+1} a_i - y_i b_i) (2h_i a_i b_i - 3y_{i+1} a_i + 3y_i b_i) + 4h_i (y_{i+1} a_i^3 - y_i b_i^3) - h_i^2 a_i^2 b_i^2 \geq 0 \right\},$$

$$h_i = x_{i+1} - x_i \text{ and } \Delta_i = \frac{y_{i+1} - y_i}{h_i}.$$

Theorem 1 gives rise to simpler sufficient conditions for positivity. One of which is described in the following corollary.

Corollary 1. With the cubic function f as described in Theorem 1, then f is nonnegative on $[x_0, x_n]$ if $(m_i, m_{i+1}) \in S_i$, $0 \leq i \leq n-1$,

$$\text{where } S_i = \left\{ (a_i, b_i) : a_i \geq -\frac{3y_i}{h_i}, b_i \leq \frac{3y_{i+1}}{h_i} \right\}.$$

As there are an infinite number of nonnegative cubic C^1 spline interpolants, for selecting one of them the curvature $\phi(f) = \sum_{i=0}^{n-1} \omega_i \int_{x_i}^{x_{i+1}} f''(x)^2 dx$ is minimized, where the weights ω_i may be 1 or $1/(1+\Delta_i^2)^3$. This leads to a programming problem

$$\begin{aligned} & \text{minimize } \sum_{i=0}^{n-1} F_i(m_i, m_{i+1}) \\ & \text{subject to } (m_i, m_{i+1}) \in V_i \end{aligned}$$

where $F_i(m_i, m_{i+1}) = \frac{4\omega_i}{h_i} \left\{ (m_i - \Delta_i)^2 + (m_i - \Delta_i)(m_{i+1} - \Delta_i) + (m_{i+1} - \Delta_i)^2 \right\}$. Two cases of V_i have been considered, i.e. $V_i = S_i$ and $V_i = W_i$. This optimization problem which is solved numerically via dualization.

(Opfer & Oberle, 1988) investigated interpolating splines which are restricted in their movements by the presence of obstacles. It is well known that the interpolating cubic splines defined on the grid

$$a = t_0 < t_1 < \dots < t_n = b$$

minimize the convex functional

$$\psi(f) = \int_a^b f''(t)^2 dt$$

uniquely in the set of interpolating C^2 splines on $[a, b]$. Opfer and Oberle derived this well known result for a larger set,

$$W_2^2 = \{f \in C^1[a, b] : f' \text{ absolutely continuous, } f'' \in L^2[a, b]\}.$$

Given data values $y_i > 0$, $i = 0, 1, \dots, n$, a minimizer of the above functional is sought in the set where the second derivative is not necessarily continuous,

$$M = \{f \in W_2^2 : f(t_i) = y_i, i = 0, 1, \dots, n, f(t) \geq 0, \forall t \in [a, b]\}.$$

Theorem 2 The minimizer g of the functional ψ satisfies the following properties :

(i) g is a natural cubic spline on a grid

$$a = \tau_1 < \tau_2 < \dots < \tau_N = b, \quad N \geq n$$

where the τ 's consist of all the given t 's and possibly some new knots which are the isolated zeros of g .

(ii) Between two neighboring t 's, there are at most two additional knots. If there are precisely two additional knots, then g vanishes between these knots.

(iii) If τ is an additional knot (i.e. not one of the given t 's) then

$$g'''(\tau^+) - g'''(\tau^-) \geq 0.$$

Based on the above necessary conditions developed, the spline interpolation problem with obstacles is formulated as an optimal control problem with an inequality constraint. An algorithm is constructed for the computation of cubic splines which are restricted by constant upper and lower bounds. This method is illustrated with an example. Figure 1 shows the cubic interpolant without restrictions while Figure 2 shows the

optimal cubic interpolant g with the restriction $-0.1 \leq g(t) \leq 1.1$, $t \in [-1, 1]$. The given data points are marked by "•" and the additional points inserted by the algorithm are marked by "◦".

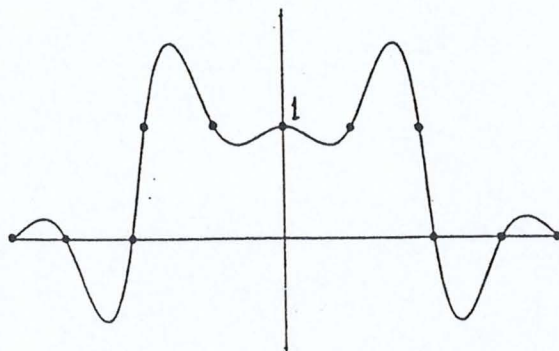


Figure 1. The unrestricted interpolating cubic spline

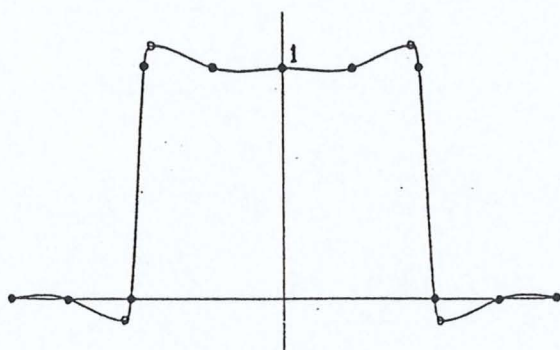


Figure 2. The restricted interpolating cubic spline $-0.1 \leq g(t) \leq 1.1$

Exponential splines with tension parameters are used in [Wever, 1988] instead of cubic splines. The exponential spline is the solution of a minimal bending energy functional

$$\sigma(y) = \int_{x_0}^{x_n} (y''(x)^2 + \Lambda(x)^2 y'(x)^2) dx \quad (2.1)$$

with $y(x_i) = y_i$, $i = 0, 1, \dots, n$, $\Lambda(x) = \lambda_i$ for $x_i \leq x \leq x_{i+1}$, $i = 0, 1, \dots, n-1$. The C^2 property of the exponential spline gives it the following representation in each interval

$$E(x) = y_{i+1}t + y_i(1-t) + \frac{d_{i+1}}{\lambda_i^2} \left\{ \frac{\sinh \mu_i t}{\sinh \mu_i} - t \right\} + \frac{d_i}{\lambda_i^2} \left\{ \frac{\sinh \mu_i (1-t)}{\sinh \mu_i} - (1-t) \right\}, \quad i = 0, 1, \dots, n-1 \quad (2.2)$$

with $x_i \leq x \leq x_{i+1}$, $h_i = x_{i+1} - x_i$, $t = \frac{x - x_i}{h_i}$, $\mu_i = \lambda_i h_i$, $y_i = E(x_i)$ and the unknown second order derivative $d_i = E''(x_i)$. The continuity of the first order derivative leads to a symmetric, positive definite tridiagonal system for the second order derivatives d_i . To get the optimal solution for the positivity preserving interpolation problem to the data $\{(x_i, y_i) : 0 \leq i \leq n\}$ with $y_i \geq 0$, the following problem must be solved :

minimize the functional (2.1)

over the set of C^2 functions on $[x_0, x_n]$ of the form (2.2) with $\lambda_i \in \mathbb{R}$, $i = 0, 1, \dots, n-1$

subject to $y(x) \geq 0$ on $[x_0, x_n]$.

This variational problem can be transformed into an optimization problem for the tension parameters λ_i . The resulting constrained non-linear optimization problem is solved by a Sequential Quadratic Program.

3. Direct methods

(Butt & Brodlie, 1993) observed that the condition at an end point in Corollary 1, for example, $m_1 \geq \frac{3y_1}{h_1}$ could be satisfied by reducing the value of h_1 and this motivates the idea of inserting an additional knot, sufficiently close to the end point to satisfy the condition. They applied this corollary to give a simple algorithm for generating a C^1 piecewise cubic Hermite interpolating spline curve which preserves positivity. Each piece of the Hermite interpolant is tested if the sufficient condition is satisfied. If not, one or two knots are inserted so that the resulting cubic spline preserves positivity. This gives an algorithm which is local in nature and it does not require the modification of the slope data. The algorithm is extended to cover the more general constraint that the constraint being a linear function of the independent variable instead of the zero function.

The methods described so far generates non-parametric curves. In (Goodman *et al.*, 1991) the problem of constructing a G^2 parametric interpolating rational spline curve which lies on the same side of a given set of constraint lines as the data has been considered. The initial interpolating curve is generated by constructing a rational cubic between each pair of adjacent data points by using the G^2 shape preserving interpolation scheme of [Goodman, 1989]. The interpolating curve is shape preserving in the sense that it has the minimal number of inflections consistent with the data. Each segment of the interpolating curve is a rational cubic of the form

$$r(t) = \frac{\alpha(1-t)^3 A + 3t(1-t)^2 B + 3t^2(1-t)C + \beta t^3 D}{\alpha(1-t)^3 + 3t(1-t)^2 + 3t^2(1-t) + \beta t^3}, \quad t \in [0, 1] \quad (3.1)$$

where the weights $\alpha, \beta > 0$ and $A, B, C, D \in \mathbb{R}^2$. The weights α and β are the parameters used for fulfilling the positivity as well as the G^2 continuity conditions. Central to the argument of this approach is the following theorem which gives the necessary and sufficient conditions for the curve segment r in (3.1) to cross over or just touch a given line.

Theorem 3: Let $r(t)$ be given by (3.1) and L a given line. Suppose that A, D lie on one side of the line L while B and / or C lie on the opposite side of L .

If $r(t)$ crosses over the line L [respectively touches the line at only one point], then

$$c^2 > 3\beta bd, \quad b^2 > 3\alpha ac, \quad (3.2)$$

and

$$b^2 c^2 + 18\alpha \beta abcd - 4(\alpha ac^3 + \beta b^3 d) - 27\alpha^2 \beta^2 a^2 d^2 > 0 \quad [\text{respectively } = 0], \quad (3.3)$$

where a, b, c, d are signed distances of the control points A, B, C, D from the line L , with points on the same side of the line as A and D having positive distances and negative otherwise. Moreover if (3.3) holds, then (3.2) also holds and $r(t)$ crosses over the line [respectively touches the line at only one point].

When the rational cubic segment in (3.1) crosses a given line, two methods of modification are presented for modifying the curve segment. The first method modifies the curve by increasing one or both of the weights of the rational cubic according to Theorem 3 so that the new curve segment just touches the line. Observe that increasing the weights, which is equivalent to decreasing the magnitudes of the tangent vectors at the corresponding data points, pulls the curve towards the line segment AD . The adjacent segments of the modified segment has to be adjusted in order to restore G^2 continuity at the knot. The second approach modifies the curve segment by the addition of a knot, i.e. the curve segment to be modified is replaced by two segments with the addition of a point that lies on the constraint line.

The constraint boundaries which have been considered above are linear curves. Non linear constraint curves was first considered in (Ong & Unsworth, 1992). The parametric G^2 scheme of (Goodman *et al.*, 1991) was adapted to generate non-parametric constrained C^1 curves with straight lines and quadratic curves as constraints. Figure 3 shows two quadratic constraint curves which are marked by (i). The unconstrained interpolating curve which is marked by (ii) has crossed the constraints but the constrained interpolating curve which is marked by (iii) stays within the region enclosed by the quadratic constraint curves. The latter is obtained by modifying the initial weights along with the addition of a knot denoted by a "x" in the figure.

(Meek *et al.*, 2003) further extends the method of Goodman *et al.* to allow a polyline as a more general constraint where all the data points need not lie on one side of the infinite line through each of its edges. This is made possible through an extension of Theorem 1. Moreover data points may also lie on the constraint

polyline. In Figure 4 the unconstrained interpolating curve drawn in gray crosses the constraint polyline which is drawn in dotted lines. The constrained interpolating curve drawn in black does not cross the constraint polyline.

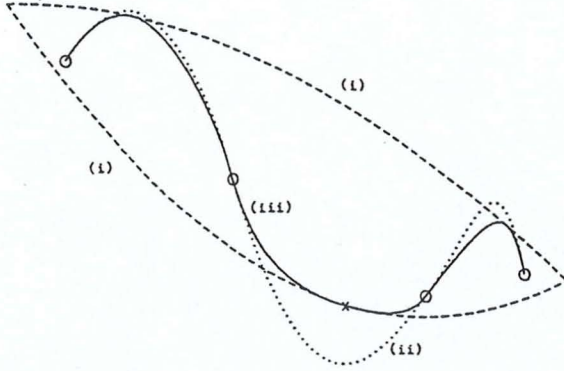


Figure 3. The interpolating curve with quadratic constraint curves

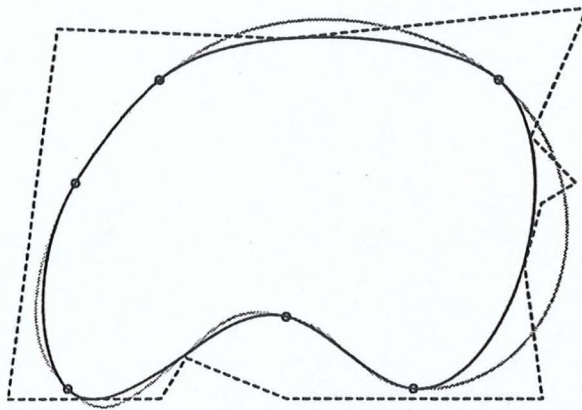


Figure 4. The interpolating curve with constraint polyline

Zhang and Cheng (2000) presented the construction of a C^1 parametric cubic interpolating spline curve which preserves the convexity of the data and lies on the same side of a given polygon as the data points. The constraint that the interpolant lies on one side of the given polygon is satisfied by varying both the magnitude and direction of the tangent vectors at the data points. The main task of this method is to choose appropriate tangent vectors. The tangent vector is usually defined as a weighted combination of adjacent chord vectors, i.e. the tangent vector T_i at data point P_i uses three consecutive data points P_{i-1}, P_i, P_{i+1} and is defined as

$$T_i = (1 - \alpha_i) \Delta P_i + \alpha_i \Delta P_{i+1}, \quad \alpha \in [0, 1],$$

where $\Delta P_i = \frac{P_{i+1} - P_i}{h_i}$, $h_i = t_{i+1} - t_i$ and t_j are the knots associated with P_j , $i - 1 \leq j \leq i + 1$. Zhang and

Cheng used scaled chord vectors. The tangent vector D_i at P_i is defined as

$$D_i = (1 - \alpha_i) \mu_i \Delta P_i + \alpha_i \nu_i \Delta P_{i+1},$$

where $0 \leq \mu_i, \nu_i \leq 1$ are scalars to be determined. The introduction of two degrees of freedom, μ_i and ν_i , provides flexibility in constructing the curve so that it has a shape suggested by the given data and lies on the same side of the given polygon as the data points. The number of interpolating curves that have the shape suggested by the data points and satisfy the given constraints could be infinite. Two techniques are presented to select the final interpolating curve. The first one selects the final curve by modifying the initial Bessel tangents

whenever necessary; and the second one achieves that by minimizing the maximum curvature of the constructed interpolant via an iterative numerical approach.

4. Conclusion

We have considered the problem of constructing interpolating curves that avoid constraint boundaries which may be straight lines, quadratic curves or polylines. Several methods and techniques have been presented.

Acknowledgements

The financial support of the Fundamental Research Grant Scheme is gratefully acknowledged.

References

- [1] Butt S. & Brodlie K.W., Preserving positivity using piecewise cubic interpolation, *Computer Graphics* 17 (1993) 55-64.
- [2] Randall L. Dougherty, Alan Edelman & James M. Hyman, Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation, *Mathematics of Computation*, Vol. 52, No. 186 (1989) 471-494.
- [3] Goodman T. N. T., Shape preserving interpolation by parametric rational cubic splines, in: *International Series of Numerical Mathematics*, Vol. 86 (Birkhauser Verlag, Basel, 1988) 149-158.
- [4] Goodman T. N. T., Shape preserving representations, in: Tom Lyche & Larry L. Schumaker, eds., *Mathematical Methods in Computer Aided Geometric Design* (Academic Press, Boston, 1989) 333-351.
- [5] Goodman T. N. T., Ong B. H. & Unsworth K., Constrained interpolation using rational cubic splines, in: G. Farin, ed., *Nurbs for Curve and Surface Design* (SIAM, Philadelphia, 1991) 59-74.
- [6] Aatos Lahtinen, Positive Hermite interpolation by quadratic splines, *SIAM J. Math. Anal.*, Vol 24, No. 1 (1993) 223-233.
- [7] Meek D. S., Ong B. H. & Walton D. J., Constrained interpolation with rational cubics. To appear in *Computer Aided Geometric Design*.
- [8] Ong B. H. & K. Unsworth K., On non-parametric constrained interpolation, in: Tom Lyche & Larry L. Schumaker, eds., *Mathematical Methods in Computer Aided Geometric Design II* (Academic Press, Inc., 1992) 419-430.
- [9] Opfer G. & Oberle H. J., The derivation of cubic splines with obstacles by methods of optimization and optimal control, *Numer. Math.* 52 (1988) 17-31.
- [10] Schmidt J.W. & Hess W., Positivity of cubic polynomials on intervals and positive spline interpolation, *Bit* 28 (1988) 340-362.
- [11] U. Wever, Non-negative exponential splines. *Computer Aided Design*, Vol. 20, No.1 (1988) 11-16.
- [12] Zhang Caiming & Cheng Fuhua, Constrained shape-preserving curve interpolation with minimum curvature, (2001). Submitted.