Planar cubic Hermite $G^1$ splines with small strain energy

Gašper Jaklič$^a$, Emil Žagar$^{b,*}$

$^a$IMFM, Jadranska 19, Ljubljana, Slovenia
$^b$FMF and IMFM, Jadranska 19, Ljubljana, Slovenia

Abstract

In this paper, a classical problem of the construction of a cubic Hermite $G^1$ continuous spline curve is considered. The only data given are interpolation points, while tangent directions are unknown. They are constructed in such a way that a particular minimization of the strain energy of the spline curve is applied. The resulting spline curve is regular, cusp-, loop- and fold-free. Even more, it is independent of a particular parameterization. Numerical examples demonstrate that it is satisfactory as far as the shape of the curve is concerned.

Key words: Hermite interpolation, spline curve, minimization

1 Introduction

The construction of planar parametric polynomial curves (or splines) based on the interpolation of data points is one of the fundamental problems in computer aided geometric design (CAGD). In the spline case, it is usually required that the resulting parametric curve is at least $G^1$ continuous, i.e., geometrically continuous of order 1. One of the basic problems is how to choose interpolating parameters for a particular parameterization of the curve. If they are given in advance, interpolating schemes become linear (see [1], [2] and [3], e.g.). On the other hand, break points might be left as unknowns, which leads to so called geometric (or parametric) interpolation. This approach is nonlinear and thus much more difficult to handle. Neither approach gives satisfactory results in various cases encountered in practice. While geometric interpolation

* Corresponding author.

Email address: emil.zagar@fmf.uni-lj.si (Emil Žagar).

Preprint submitted to Elsevier Science 30 May 2008
is usually superior when asymptotic analysis is concerned, linear schemes can give better results when interpolation of separated data is needed.

When geometric continuity is required, the choice of tangent directions at data points becomes important. Usually there are three possibilities:

(a) Tangent directions are given in advance.
(b) The choice of appropriate directions is left to the designer.
(c) An interpolating spline is required to be $G^1$ continuous, but the tangent directions are not specified.

While the first two approaches usually lead to local interpolating schemes, the last one implies a global set of conditions, given as a large (fortunately banded) system of linear equations. Note that the second approach should be applied just for local changes of the spline, otherwise quickly geometrically nonfeasible curves can be obtained, that are not pleasing to the human eye.

Since tangent directions are rarely available in practice, they should be determined by some simple procedure, preferably by an easy and geometrically evident construction.

In this paper we consider the following problem. Suppose that data points

$$T_j \in \mathbb{R}^2, \quad T_j \neq T_{j+1}, \quad j = 0, 1, \ldots, n,$$

and associated interpolation parameters

$$t_j \in \mathbb{R}^2, \quad j = 0, 1, \ldots, n, \quad t_0 < t_1 < \cdots < t_n,$$

are given. Interpolating parameters can of course be derived from data points, e.g., by the centripetal or the chord length parameterization, but generally we can assume that they are prescribed in advance. Our goal is to find a $G^1$ continuous parametric spline curve $s: [t_0, t_n] \to \mathbb{R}^2$ such that

$$s_i := s|_{[t_{i-1}, t_i]} \in \mathbb{P}^3, \quad i = 1, 2, \ldots, n,$$

$$s_i(t_k) = T_k, \quad k = i - 1, i, \quad i = 1, 2, \ldots, n,$$

$$s'_i(t_k) = \alpha_{i,k-i+1} d_k, \quad k = i - 1, i, \quad i = 1, 2, \ldots, n,$$

where $\alpha_{i,k-i+1} > 0$ are unknown positive scalars, $d_k$ normalized tangent direction vectors, and $\mathbb{P}^3$ is the space of parametric polynomials of degree $\leq 3$. There is of course infinitely many solutions of the above problem, since it is very well known that any set of $\alpha_{i,k-i+1} > 0$ and $d_k$ gives a unique spline curve $s$. Thus we have a large set of free parameters which can be used to design an appropriate shape of the spline curve.

One of the natural approaches how to deal with the free parameters is to define a suitable functional and minimize it. Usually, the shape of the curve
depends mostly on its curvature and therefore it seems reasonable to minimize the functional (the so called strain energy)

\[ \varphi(\alpha) := \int_{t_0}^{t_n} \| s''(t) \|^2 \, dt, \quad (3) \]

where \( \alpha := (\alpha_{i,k-i+1})_{i=1,k=i-1}^{n,i} \in \mathbb{R}^{2n} \) and \( \| \cdot \| \) is the Euclidean norm in \( \mathbb{R}^2 \). Note that \( s'' \) might not be continuous, but it has only a finite number of finite jumps thus the integral (3) clearly exists. This approach has been used in [4], but tangent directions were supposed to be known there. In this paper we consider another reasonable approach. Instead of minimizing (3), one would like to minimize its discrete approximation. Of course, the positivity of scalar parameters \( \alpha_{i,k-i+1} \) must be taken into account, which obviously leads to a constrained optimization problem. On the contrast to [4] where tangent directions were given in advance, we shall consider here that they have to be determined in a feasible way.

The paper is organized as follows. In the second section the minimization approach is outlined. Necessary and sufficient conditions for the existence of the optimal spline curve are given and the regularity of the spline is proved. In Section 3 a detailed construction of the spline curve based on the results from Section 2 is described. An efficient algorithm for the construction of a cubic Hermite \( G^1 \) spline by minimization is presented. The last section gives a number of numerical examples.

2 Minimization technique

First, some basic notation used through the paper will be introduced. Let \( \langle \cdot, \cdot \rangle \) be the standard inner product in \( \mathbb{R}^2 \) and \( \angle(\mathbf{a}, \mathbf{b}) \) the angle formed by vectors \( \mathbf{a} \) and \( \mathbf{b} \). Recall that

\[ \langle \mathbf{a}, \mathbf{b} \rangle = \| \mathbf{a} \| \| \mathbf{b} \| \cos \angle(\mathbf{a}, \mathbf{b}), \quad \mathbf{a} \times \mathbf{b} = \| \mathbf{a} \| \| \mathbf{b} \| \sin \angle(\mathbf{a}, \mathbf{b}) \],

where \( \mathbf{a} \times \mathbf{b} := a_1 b_2 - a_2 b_1 \) is the planar vector product. We will also need the standard forward difference notation, i.e., \( \Delta(\bullet)_i = (\bullet)_{i+1} - (\bullet)_i \).

Consider now the functional \( \varphi \) in (3). If

\[ \varphi_i(\alpha) := \int_{t_{i-1}}^{t_i} \| s''_i(t) \|^2 \, dt, \quad i = 1, 2, \ldots, n, \]

then \( \varphi \) can be written as

\[ \varphi(\alpha) = \sum_{i=1}^{n} \varphi_i(\alpha). \]
We want to minimize it on the open set \( \mathcal{D} = \{ \alpha \in \mathbb{R}^{2n} | \alpha > 0 \} \). Here the inequality \( \alpha > 0 \) is considered componentwise. But \( \varphi_i \) depends only on the local parameters \( \alpha_i := (\alpha_{i,0}, \alpha_{i,1}) \), thus \( \varphi_i(\alpha) = \varphi_i(\alpha_i) \) and obviously

\[
\min_{\alpha \in \mathcal{D}} \varphi(\alpha) = \sum_{i=1}^{n} \min_{\alpha_i \in \mathcal{D}_i} \varphi_i(\alpha_i),
\]

where \( \mathcal{D}_i = \{ \alpha_i \in \mathbb{R}^2 | \alpha_i > 0 \} \). Instead of explicitly deriving \( s''_i \) and applying minimization as in [4], a discrete approximation of \( \varphi_i \) will be used. It is clear that \( \varphi_i \equiv 0 \) if \( s_i \) is a line segment joining \( T_{i-1} \) and \( T_i \). In general, \( s_i \) will be a cubic polynomial of course, but a choice of an appropriate \( \alpha_i \) might minimize its first two leading coefficients and thus make it closer to its control polygon (the line \( T_{i-1}T_i \)). To find a suitable discrete approximation of \( \varphi_i \), let us first write \( s_i \) in the Newton form as

\[
s_i(t) = T_{i-1} + (t - t_{i-1}) [t_{i-1}, t_{i-1} - 1] s_i + (t - t_{i-1})^2 [t_{i-1} - 1, t_{i-1}, t_i] s_i + (t - t_{i-1})^3 [t_{i-1} - 1, t_{i-1}, t_i, t_i] s_i,
\]

where \( [t_{i-1}, t_{i-1}] s_i = \alpha_{i,0} d_{i-1} \) and \( [t_i, t_i] s_i = \alpha_{i,1} d_i \). It is now clear that one way to minimize its two leading coefficients is to minimize

\[
||[t_{i-1}, t_{i-1}, t_i] s_i||^2 + ||[t_{i-1}, t_{i-1}, t_i, t_i] s_i||^2.
\]

But it turns out that this leads to a complicated analysis. In order to keep things as simple as possible, the observation that

\[
||[t_{i-1}, t_{i-1}, t_i] s_i|| \leq \frac{1}{\Delta t_{i-1}} (||[t_{i-1}, t_{i-1}, t_i] s_i|| + ||[t_{i-1}, t_i] s_i||)
\]

suggests to minimize

\[
\psi_i(\alpha_i) := ||[t_{i-1}, t_{i-1}, t_i] s_i||^2 + ||[t_{i-1}, t_i] s_i||^2,
\]

instead. Indeed, if \( \psi_i(\alpha_i) = 0 \), then

\[
[t_{i-1}, t_{i-1}, t_i] s_i = [t_{i-1}, t_{i-1}, t_i, t_i] s_i = 0,
\]

and \( s_i \) in (4) reduces to a straight line. On the other hand, \( \psi_i \) can be viewed as a discrete approximation of \( \varphi_i \). Namely,

\[
[t_{i-1}, t_{i-1}, t_i] s_i = \frac{1}{2} s''_i(\xi_1) \quad [t_{i-1}, t_i, t_i] s_i = \frac{1}{2} s''_i(\xi_2), \quad \xi_1, \xi_2 \in [t_{i-1}, t_i],
\]

which implies

\[
\psi_i(\alpha_i) = \frac{1}{2} ||s''_i(\xi_3)||^2, \quad \xi_3 \in [t_{i-1}, t_i],
\]

i.e., a zeroth order approximation of \( 2\varphi_i/\Delta t_{i-1} \). Thus, instead of minimizing \( \varphi_i \), the minimization of \( \psi_i \) will be done. An optimal choice of \( \alpha_i \) is stated in the following theorem.
Theorem 1 The nonlinear functional $\psi_i$, $i = 1, 2, \ldots, n$, has a unique global minimum in the interior of $D_i$ iff

$$\alpha_i^* := \frac{1}{\Delta t_{i-1}} \langle (d_{i-1}, \Delta T_{i-1}), (d_i, \Delta T_{i-1}) \rangle^T > 0.$$ 

Furthermore,

$$\min_{\alpha_i \in D_i} \psi_i(\alpha_i) = \frac{2 - c_{i,0}^2 - c_{i,1}^2}{(\Delta t_{i-1})^4} \| \Delta T_{i-1} \|^2,$$

where

$$c_{i,k} = \cos \angle (d_{i+k-1}, \Delta T_{i-1}), \quad k = 0, 1.$$ 

PROOF. Some basic properties of divided differences together with (2) simplifies (5) to

$$\psi_i(\alpha_i) = \frac{1}{(\Delta t_{i-1})^2} \left( (\Delta t_{i-1})^2 \left( \alpha_{i,0}^2 + \alpha_{i,1}^2 \right) - 2 \Delta t_{i-1} \langle \alpha_{i,0} d_{i-1} + \alpha_{i,1} d_i, \Delta T_{i-1} \rangle 
+ 2 \| \Delta T_{i-1} \|^2 \right).$$

Minima of $\psi_i$ are either on the boundary of $D_i$ or they are obtained by partial derivations of $\psi_i$. In the later case, a local minimum appears at $\alpha_i^* := (\alpha_{i,0}^*, \alpha_{i,1}^*)^T$, where

$$\alpha_{i,k}^* = \frac{\langle d_{i+k-1}, \Delta T_{i-1} \rangle}{\Delta t_{i-1}}, \quad k = 0, 1,$$

which leads to

$$m := \psi_i(\alpha_i^*) = \frac{2 - c_{i,0}^2 - c_{i,1}^2}{(\Delta t_{i-1})^4} \| \Delta T_{i-1} \|^2.$$ 

It remains to prove that this is a global minimum as well. Take any $\alpha_i = (\alpha_{i,0}, \alpha_{i,1})^T$ on the boundary of $D_i$. Thus $\alpha_{i,k} = 0$ for at least one $k \in \{0, 1\}$. If $\alpha_{i,0} = \alpha_{i,1} = 0$ then $\psi_i(\alpha_i) = 2 \| \Delta T_{i-1} \|^2 / (\Delta t_{i-1})^4 \geq m$, and it remains to consider the case when only one of $\alpha_{i,k}$ is positive. Due to the symmetry, it is enough to study $\alpha_{i,0} > 0$ only. But in this case it is easy to check that the functional $\psi_i |_{\alpha_{i,1} = 0}$ attains its global minimum at $\alpha_{i,0} = \langle d_{i-1}, \Delta T_{i-1} \rangle / \Delta t_{i-1}$ implying $\psi_i(\alpha_{i,0}, 0) = \left( 2 - c_{i,0}^2 \right) \| \Delta T_{i-1} \|^2 / (\Delta t_{i-1})^4 \geq m$. This concludes the proof of the lemma.

Notice that the minimum can also be zero. In this case $c_{i,0} = c_{i,1} = 0$ and the cubic parametric spline segment $s_i$ reduces to a straight line $s_i(t) = T_{i-1} + (t - t_{i-1})[t_{i-1}, t_i] s_i$.

Corollary 2 The conditions $\alpha_i^* > 0$, $i = 1, 2, \ldots, n$, have a simple geometric interpretation, i.e., $\angle (d_{i+k-1}, \Delta T_{i-1}) \in [0, \frac{\pi}{2})$, $k = 0, 1$. 

5
Now suppose that the assumptions of Theorem 1 are satisfied. Then an important question arises whether the resulting cubic spline segment $s_i$ is regular on $[t_{i-1}, t_i]$, $i = 1, 2, \ldots, n$. The answer is confirmative, even more, $s_i$ has no cusps, loops or folds and is independent of the parameterization (1) as stated in the following lemma.

**Lemma 3** Let the assumptions of Theorem 1 be satisfied and let $s_i$ be the resulting Hermite geometric interpolant defined by (2). Then the spline segment $s_i$ is regular, loop-, cusp-, fold-free and parameterization independent on $[t_{i-1}, t_i]$, $i = 1, 2, \ldots, n$.

**Proof.** Let us reparameterize $s_i$ by a local parameter $u = (t - t_{i-1})/\Delta t_{i-1}$, i.e., let $p_i(u) := s_i(t)$, $t \in [t_{i-1}, t_i]$. It is enough to show that $p_i$ is regular, loop-, cusp-, fold-free and parameterization independent on $[0, 1]$. By (2) and Theorem 1

\[
p_i(k) = T_{i+k-1}, \quad k = 0, 1,
\]

\[
p_i'(k) = \Delta t_{i-1} \alpha_{i,k} d_{i+k-1} = \langle d_{i+k-1}, \Delta T_{i-1} \rangle d_{i+k-1}, \quad k = 0, 1.
\]

It is clear that $p_i$ is independent of the parameterization (1). To show that it is regular and has no cusps, loops and folds, let us first rewrite it in the Bézier form

\[
p_i(u) = T_{i-1} B^3_0(u) + \left( T_{i-1} + \frac{1}{3} \langle d_{i-1}, \Delta T_{i-1} \rangle d_{i-1} \right) B^3_1(u) + \left( T_i - \frac{1}{3} \langle d_i, \Delta T_{i-1} \rangle d_i \right) B^3_2(u) + T_i B^3_3(u),
\]

where $B^3_i$, $i = 0, 1, 2, 3$, are cubic Bernstein polynomials. By using a translation and a rotation we can further assume that $T_{i-1} = (0, 0)^T$ and $T_i = (x_1, 0)^T$, $x_1 > 0$. Since by Theorem 1 $\alpha_i^* > 0$, the conclusion that $d_{i+k-1,1} > 0$, $k = 0, 1$, where $d_{i+k-1,1}$ is the first component of $d_{i+k-1}$, follows immediately. A differentiation of $p_i$ yields

\[
p_i'(u) = \langle d_{i-1}, \Delta T_{i-1} \rangle d_{i-1} B^2_0(u) + \left( 3 \Delta T_{i-1} - \langle d_{i-1}, \Delta T_{i-1} \rangle \right) d_{i-1} - \langle d_i, \Delta T_{i-1} \rangle d_i B^2_1(u) + \langle d_i, \Delta T_{i-1} \rangle d_i B^2_2(u).
\]

Let $p'_{i,1}$ be the first component of $p_i'$. To see that $p_i$ is regular and loop-, cusp- and fold-free on $[0, 1]$, it is enough to verify that $p'_{i,1}(u) > 0$ for $u \in [0, 1]$. Quite clearly

\[
p'_{i,1}(u) = x_1 \left( d_{i-1,1}^2 B^2_0(u) + \left( 3 - d_{i-1,1}^2 - d_{i,1}^2 \right) B^2_1(u) + d_{i,1}^2 B^2_2(u) \right).
\]

Since $d_i$ are normalized and $d_{i+k-1,1} > 0$, the conclusion $0 < d_{i+k-1,1} < 1$, $k = 0, 1$, follows. Since also $x_1 > 0$, all the Bézier coefficients of $p'_{i,1}$ are
positive and by the convex hull property of Bézier curves $p'_{i,1}$ is positive on $[0, 1]$.

3 Construction of tangent directions

In the previous section a construction of the cubic Hermite $G^1$ polynomial spline curve has been explained. We have assumed that the unit tangent directions $d_i$ are known and the parameters $\alpha^*_i$ are obtained by a particular minimization technique. As it is clear from Theorem 1, tangent directions must be chosen properly, otherwise the conclusion of the theorem can not be applied.

In this section the problem of the construction of tangent directions $d_i$ will be considered. They should be chosen in such a way that they provide a minimization of the functional (5) as stated in Theorem 1. Thus it is enough to require
\[
\langle d_{i+k-1}, \Delta T_{i-1} \rangle > 0, \quad i = 1, 2, \ldots, n, \quad k = 0, 1.
\]

In order to fulfill these conditions, consider the $(i-1)$-th and $i$-th segment of the spline curve $s$ (see Fig. 1). Let us define a rotation
\[
R := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}
\]
and $z_i := \text{sign} (\Delta T_{i-1} \times \Delta T_i)$. Further, let
\[
\begin{align*}
  u_i &:= z_i R \Delta T_{i-1}, \\
  v_i &:= -z_i R \Delta T_i.
\end{align*}
\]
Define $w_i := u_i + \beta_i v_i$, $\beta_i \in \mathbb{R}$. It is now easy to prove the following lemma.

Lemma 4 If $z_i \neq 0$ and $\beta_i > 0$, then $\langle w_i, \Delta T_j \rangle > 0$, $j = i - 1, i$.

PROOF. We will prove that $\langle w_i, \Delta T_j \rangle > 0$, $j = i - 1$. The proof for $j = i$ is very similar and will be omitted. Take any $w_i = u_i + \beta_i v_i$ with $\beta_i > 0$. By (6)-(8)
\[
\begin{align*}
  \langle w_i, \Delta T_{i-1} \rangle &= \langle u_i, \Delta T_{i-1} \rangle + \beta_i \langle v_i, \Delta T_{i-1} \rangle \\
  &= -\beta_i z_i \langle R \Delta T_i, \Delta T_{i-1} \rangle,
\end{align*}
\]
since $u_i$ and $\Delta T_{i-1}$ are perpendicular. Obviously, it is enough to show that
\[
z_i = -\text{sign}(R \Delta T_i, \Delta T_{i-1}).
\]
Since $z_i \neq 0$, $\angle (\Delta T_{i-1}, \Delta T_i) \neq 0, \pi$, and there are two possibilities.

(a) If $z_i > 0$, then $\angle (R \Delta T_i, \Delta T_{i-1}) > \pi/2$ and sign $\langle R \Delta T_i, \Delta T_{i-1} \rangle < 0$.
(b) If $z_i < 0$, then $\angle (R \Delta T_i, \Delta T_{i-1}) < \pi/2$ and sign $\langle R \Delta T_i, \Delta T_{i-1} \rangle > 0$.

From Lemma 4 it follows that any $\beta_i > 0$ gives $w_i$ that leads to a minimization of the functional (5). Namely, one just takes $d_i = w_i/\|w_i\|$. Thus $\beta_i, i = 1, 2, \ldots, n$, can be considered as free shape parameters. One of the choices would, e.g., be $\beta_i = \|u_i\|/\|v_i\|$, which implies that $d_i$ points from $T_i$ in the direction of the bisector of the angle $\angle (u_i, v_i)$ (see Fig. 1). This is a natural choice since this $d_i$ stays away from the unwanted directions implying $\alpha_{i,k} = 0$, for $k = 0$ or $k = 1$, as much as possible. Lemma 4 also excludes two possibilities, namely $\angle (\Delta T_{i-1}, \Delta T_i) = 0, \pi$. But, if the considered angle is equal to 0 then $w_i$ can be taken as $w_i = \Delta T_i$, and again the conclusions of the lemma follow. On the other hand, the case when the angle equals $\pi$ would imply that any parameterization of such data has a fold. Thus, this case should be excluded from possible data sets by using some kind of preprocessing of data points (by inserting an additional point, e.g.).

For the first and the last tangent direction the above procedure can not be applied, but in this case $d_0$ and $d_n$ can be, e.g., chosen as $d_0 = \Delta T_0/\|\Delta T_0\|$ and $d_n = \Delta T_{n-1}/\|\Delta T_{n-1}\|$. Thus for given shape parameters $\beta_i > 0$, Algorithm 1 can be used for the construction of tangent directions and the resulting Hermite $G^1$ cubic spline.
Algorithm 1. Returns tangent directions $d_i$ and cubic $G^1$ Hermite spline pieces $s_i$.

\begin{algorithm}
for $i = 1$ to $n - 1$
  if $\angle(\Delta T_{i-1}, \Delta T_i) = \pi$
    Exit.
  end if
end for

$d_0 = \Delta T_0 / \| \Delta T_0 \|$
for $i = 1$ to $n - 1$
  if $\angle(\Delta T_{i-1}, \Delta T_i) = 0$
    $d_i = \Delta T_i / \| \Delta T_i \|$
  else
    $u_i = z_i \cdot R \cdot \Delta T_{i-1}$
    $v_i = -z_i \cdot R \cdot \Delta T_i$
    $w_i = u_i + \beta_i \cdot v_i$
    // $\beta_i$ are given in advance or calculated as $\beta_i = \| u_i \| / \| v_i \|$
    $d_i = w_i / \| w_i \|$
  end if
end for

$d_n = \Delta T_{n-1} / \| \Delta T_{n-1} \|$  
for $i = 1$ to $n$
  Construct spline piece $s_i$ using $T_j, d_j, j = i - 1, i$, and Theorem 1.
end for
\end{algorithm}

4 Numerical examples

Let us conclude the paper by numerical examples. Our cubic $G^1$ Hermite interpolant can closely resemble the $C^2$ cubic interpolating spline, but for nonuniformly distributed data points (exchange of short and long segments) larger differences between the curve shapes can occur (Fig. 2). In Fig. 3 the influence of the shape parameters $\beta_i$ on the curve is presented.

![Fig. 2. The cubic $G^1$ Hermite interpolant (solid) closely resembles the $C^2$ cubic interpolating spline (left). For nonuniformly distributed data points larger differences between the curves can occur (right).](image-url)
Fig. 3. The interpolants for very small ($\beta = 1/100$, dashed line) or large ($\beta = 1000$, dash-dot line) shape parameters can give a good approximation of the control polygon at near right angles, but can have unwanted behavior elsewhere.

References


