Planar cubic G^1 interpolatory splines with small strain energy

Gašper Jaklič [a], Emil Žagar [b,*]

[a] *FMF and IMFM, University of Ljubljana and PINT, University of Primorska, Jadranska 19, Ljubljana, Slovenia*

[b] *FMF and IMFM, University of Ljubljana, Jadranska 19, Ljubljana, Slovenia*

Abstract

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

Key words: Hermite interpolation, spline curve, minimization, strain energy

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 in CAGD is how to choose parameters of interpolation (break points). If they are given in advance, interpolation schemes become linear (see [4], [6] and [15], e.g.). On the other hand, they might be left as unknowns, which leads to so called geometric interpolation (see [5], [18], [9], [16], [14], [11], [13], [10], e.g.). Neither approach gives satisfactory results in various cases encountered in practical applications. While geometric interpolation is usually superior when asymptotic

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

analysis is concerned, linear schemes can give better results when interpolation of separated data is needed. In the later case, some other criteria become important, such as the shape of the curve, its strain energy, curvature, convexity,... This leads to so called shape preserving techniques, nowadays a very well understood topic (see [8], [17], [12], [1], [3], [2], e.g.).

In general, there are two classes of interpolation problems, Lagrange and Hermite. While the first one requires only data points, the second one needs also some information on derivatives. Although Lagrange interpolation problems seem easier to solve, Hermite ones are much more simple to handle. But they have a serious drawback, derivatives are rarely available in practice. This usually requires a generation of artificial data.

When dealing with splines, two kinds of smoothness conditions at the breakpoints are encountered, parametric and geometric continuity. While the first one requires continuity of derivatives up to some order, the second one relaxes these conditions to continuity of geometric quantities, such as tangent directions, curvatures,..., only. When G^1 (or C^1) smoothness is required, the choice of tangent directions at data points becomes vital, since it significantly influences the shape of the spline. Usually there are three possibilities:

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

While the first two approaches 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 non-feasible 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 a new method for the construction of a cubic G^1 continuous interpolatory spline curve is proposed, where the only data prescribed are interpolation points, while tangent directions are unknown. They are constructed automatically in such a way that a particular minimization of the strain energy of the spline curve is done. The resulting spline curve is constructed locally and is regular, cusp-, loop- and fold-free. Furthermore, it is satisfactory as far as the shape of the curve is concerned.

The paper is organized as follows. In the next section the problem considered is outlined. In Section 3 the minimization approach is described. Necessary and sufficient conditions for the existence of the optimal spline curve are given and the regularity of the spline is proved. In Section 4, a detailed construction of the spline curve, based on the results from Section 3, is presented. Furthermore, an optimization process for the choice of the tangent directions is proposed. An efficient algorithm for the construction of a cubic Hermite G^1 spline is outlined. The last section gives a number of numerical examples that illustrate the results of the paper.

2 Interpolation problem

We will start with the description of some basic notation needed. For $\boldsymbol{a} = (a_1, a_2)^T$, $\boldsymbol{b} = (b_1, b_2)^T \in \mathbb{R}^2$, let $\boldsymbol{a} \cdot \boldsymbol{b} := a_1 b_1 + a_2 b_2$ be the standard inner product in $\mathbb{R}^2$ and $\angle(\boldsymbol{a}, \boldsymbol{b})$ the angle formed by vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Note that all considered angles will be unsigned, and the term angle will be used also for the magnitude of an angle. Recall that

$$\boldsymbol{a} \cdot \boldsymbol{b} = \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \, \cos \angle(\boldsymbol{a}, \boldsymbol{b}), \quad |\boldsymbol{a} \times \boldsymbol{b}| = \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \, |\sin \angle(\boldsymbol{a}, \boldsymbol{b})|,$$

where $\boldsymbol{a} \times \boldsymbol{b} := a_1 b_2 - a_2 b_1$ is the planar vector product and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^2$. We will use the standard difference notation, i.e., $\Delta(\bullet)_i = (\bullet)_{i+1} - (\bullet)_i$ and the standard divided differences, defined by

$$\begin{aligned} [\underbrace{x_i, x_i, \ldots, x_i}_{k}]f &:= \frac{1}{(k-1)!} f^{(k-1)}(x_i), \\ [x_i, x_{i+1}, \ldots, x_j]f &:= \frac{[x_{i+1}, \ldots, x_j]f - [x_i, \ldots, x_{j-1}]f}{x_j - x_i}, \quad x_i \neq x_j. \end{aligned}$$

The problem considered is as follows. Suppose that data points

$$\boldsymbol{T}_j \in \mathbb{R}^2, \quad j = 0, 1, \ldots, n,$$

with $\boldsymbol{T}_j \neq \boldsymbol{T}_{j+1}$ and associated interpolation parameters

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

are given. We will assume that the interpolation parameters are prescribed (usually they are derived from data points, e.g., by the centripetal, chord length or α-parameterization, see [7]). Our goal is to find a G^1 continuous parametric spline curve $\boldsymbol{s} : [t_0, t_n] \to \mathbb{R}^2$ such that

$$\begin{aligned} &\boldsymbol{s}_i := \boldsymbol{s}|_{[t_{i-1}, t_i]} \in \mathbb{P}_3, \quad i = 1, 2, \ldots, n, \\ &\boldsymbol{s}_i(t_k) = \boldsymbol{T}_k, \quad k = i-1, i, \quad i = 1, 2, \ldots, n, \\ &\dot{\boldsymbol{s}}_i(t_k) = \alpha_{i,k-i+1} \, \boldsymbol{d}_k, \quad k = i-1, i, \quad i = 1, 2, \ldots, n, \end{aligned} \tag{1}$$

where $\alpha_{i,k-i+1} > 0$ are unknown scalars, $\boldsymbol{d}_k$ are normalized tangent direction vectors, and $\mathbb{P}_3$ is the space of planar parametric polynomials of degree ≤ 3. Of course, there are infinitely many solutions for the above problem, since it is well known that any set of $\alpha_{i,k-i+1} > 0$ and $\boldsymbol{d}_k$ gives a unique spline curve $\boldsymbol{s}$. Thus we have a large set of free parameters which can be used as shape parameters of the spline.

One of the approaches for choosing the parameters $\boldsymbol{\alpha} := (\alpha_{i,k-i+1})_{i=1,k=i-1}^{n,i} \in \mathbb{R}^{2n}$ is to define a suitable functional and minimize it. Usually, the shape of the curve depends mostly on its curvature κ and therefore it seems reasonable to minimize the functional

$$\varphi_s(\boldsymbol{\alpha}) := \int_{t_0}^{t_n} \|\kappa(t)\|^2 \, dt = \int_{t_0}^{t_n} \frac{\|\dot{\boldsymbol{s}}(t) \times \ddot{\boldsymbol{s}}(t)\|^2}{\|\dot{\boldsymbol{s}}(t)\|^6} \, dt. \tag{2}$$

The expression (2) is called **the strain energy** of the curve. In practice ([4], [19]), the approximate strain energy (called also **linearized bending energy**)

$$\varphi(\boldsymbol{\alpha}) := \int_{t_0}^{t_n} \|\ddot{\boldsymbol{s}}(t)\|^2 \, dt, \tag{3}$$

is used instead of (2). Note that the approximate strain energy is close to a real one if $\|\dot{\boldsymbol{s}}(t)\| \approx 1$. If this is not the case, it can be far away from the real strain energy. But the beauty of the approximant lies in the fact that the minimization problem for the coefficients becomes linear. Note that $\ddot{\boldsymbol{s}}$ might not be continuous, but it has only a finite number of finite jumps, thus the integral (3) clearly exists. It is obvious that the minimization can be done locally. Namely,

$$\min_{\boldsymbol{\alpha}} \varphi(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \min_{\boldsymbol{\alpha}_i} \varphi_i(\boldsymbol{\alpha}_i),$$

where

$$\varphi_i(\boldsymbol{\alpha}_i) := \int_{t_{i-1}}^{t_i} \|\ddot{\boldsymbol{s}}_i(t)\|^2 \, dt, \tag{4}$$

and $\boldsymbol{\alpha}_i := (\alpha_{i,0}, \alpha_{i,1})$. It is clear from the geometry that the components of $\boldsymbol{\alpha}_i$ should be positive, since otherwise the tangent vectors of $\boldsymbol{s}_i$ at t_{i-1} and t_i would not have the same directions as given tangent directions $\boldsymbol{d}_{i-1}$ and $\boldsymbol{d}_i$. So one should have in mind that actually a constrained minimization of (4) has to be done, i.e.,

$$\min_{\boldsymbol{\alpha}_i \in \mathcal{D}_i} \varphi_i(\boldsymbol{\alpha}_i), \quad \mathcal{D}_i := \{\boldsymbol{\alpha}_i \in \mathbb{R}^2 \mid \boldsymbol{\alpha}_i > \boldsymbol{0}\}.$$

Note that the inequality $\boldsymbol{\alpha}_i > \boldsymbol{0}$ is considered componentwise. Unfortunately, as already observed in [20], for given tangent directions $\boldsymbol{d}_{i-1}$, $\boldsymbol{d}_i$, the global minimum is not always in $\mathcal{D}_i$, since the following theorem holds true.

Theorem 1 ([20]) *Let $\boldsymbol{s}_i$ be the local Hermite interpolant on $[t_{i-1}, t_i]$ satisfying* (1) *and let $\beta_k = \angle(\Delta \boldsymbol{T}_{i-1}, \boldsymbol{d}_k)$, $k = i-1, i$ (see Fig. 1) and $\beta_{i-1,i} :=$*

$\angle(\boldsymbol{d}_{i-1}, \boldsymbol{d}_i)$. *Then the global minimum* $\boldsymbol{\alpha}_i$ *of* φ_i *is given by*

$$\alpha_{i,0} = \frac{6\,\Delta \boldsymbol{T}_{i-1} \cdot \boldsymbol{d}_{i-1} - 3\,(\Delta \boldsymbol{T}_{i-1} \cdot \boldsymbol{d}_i)\,(\boldsymbol{d}_{i-1} \cdot \boldsymbol{d}_i)}{\Delta t_{i-1}\left(4 - (\boldsymbol{d}_{i-1} \cdot \boldsymbol{d}_i)^2\right)},$$
$$\alpha_{i,1} = \frac{6\,\Delta \boldsymbol{T}_{i-1} \cdot \boldsymbol{d}_i - 3\,(\Delta \boldsymbol{T}_{i-1} \cdot \boldsymbol{d}_{i-1})\,(\boldsymbol{d}_{i-1} \cdot \boldsymbol{d}_i)}{\Delta t_{i-1}\left(4 - (\boldsymbol{d}_{i-1} \cdot \boldsymbol{d}_i)^2\right)}.$$

Furthermore, $\boldsymbol{\alpha}_i > \mathbf{0}$ *if and only if*

$$2\cos\beta_{i-1} - \cos\beta_i \cos\beta_{i-1,i} > 0 \quad \textit{and} \quad 2\cos\beta_i - \cos\beta_{i-1} \cos\beta_{i-1,i} > 0.$$

The resulting Hermite interpolant $\boldsymbol{s}_i$ *is regular on* $[t_{i-1}, t_i]$ *if additionally*

$$\cos\beta_{i-1} > \frac{1}{3} \quad \textit{and} \quad \cos\beta_i > \frac{1}{3}. \tag{5}$$

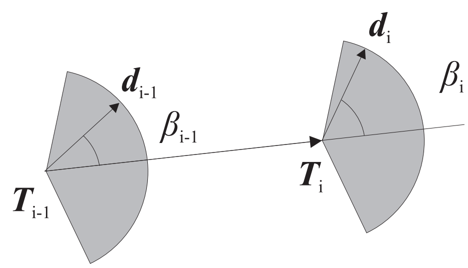

Fig. 1. Single segment and critical betas.

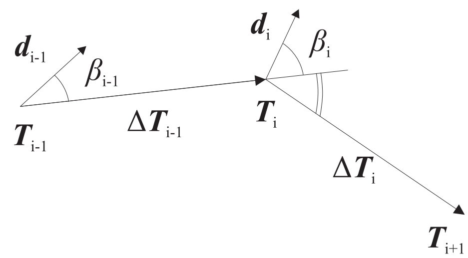

Fig. 2. Two segments.

It is clear from Theorem 1 that tangent directions $\boldsymbol{d}_{i-1}$ and $\boldsymbol{d}_i$ should satisfy some geometric constraints to assure that the minimum of φ_i is in the desirable domain $\mathcal{D}_i$ and the resulting Hermite interpolant is regular. For the regularity, e.g., admissible positions of tangent directions are shown in Fig. 1. This becomes very important when local Hermite interpolants are put together to form a spline curve. Consider namely two segments (see Fig. 2). If the angle $\angle(\Delta\boldsymbol{T}_{i-1}, \Delta\boldsymbol{T}_i) > \beta_{i-1} + \beta_i$, then no tangent direction $\boldsymbol{d}_i$ will guarantee the regularity of either $\boldsymbol{s}_i$ or $\boldsymbol{s}_{i+1}$. Since from (5) critical values for β_{i-1} and β_i are $\approx 70.53°$, the angle $\angle(\Delta\boldsymbol{T}_{i-1}, \Delta\boldsymbol{T}_i)$ should be less than $141.06°$. In [20] this problem was solved by a preprocessing of data. The authors suggested inserting additional two or three segments to achieve a desirable bound on the above angle. Although the methods proposed in [20] are relatively simple, they require an amount of additional work which becomes a bottleneck in applications dealing with huge amounts of data. Furthermore, the assumption that tangent directions are given in advance is usually not realistic at all.

In this paper we consider another approach that overcomes the mentioned problems. Instead of inserting additional segments and calculating new interpolation points and tangent directions if the minimum of φ_i is not guaranteed to be in an admissible domain $\mathcal{D}_i$, we assume that tangent directions are unknown. Of course, restrictions on tangent directions given by Theorem 1, might imply that no tangent direction at a given interpolation point would be admissible. So it is clear that if we persist in a fixed number of spline segments, the functional φ_i might not have a minimum in an admissible region $\mathcal{D}_i$ for some tangent directions $\boldsymbol{d}_{i-1}$ and $\boldsymbol{d}_i$. Thus a reasonable approximation of φ_i, leading to more relaxed conditions on admissible regions for tangent directions, will be proposed.

The developed method will not require prescribed tangent directions and no additional (artificial) data will be needed. An algorithm for constructing optimal tangent directions, such that the approximating strain energy functional is minimized, will be provided.

3 Minimization technique

The main goal of this section is to provide such an approximation for the functional (4), that the magnitudes of admissible angles β_{i-1} and β_i for tangent directions $\boldsymbol{d}_{i-1}$ and $\boldsymbol{d}_i$ are as large as possible. Intuitively, a tangent direction at a given interpolation point $\boldsymbol{T}_{i-1}$ should always point into the same halfplane as the vector $\Delta\boldsymbol{T}_{i-1}$. Thus the maximal expected interval for the magnitudes of (unsigned angles) β_{i-1} and β_i to lie in is $[0, \pi/2)$. If additionally the resulting Hermite interpolant is regular, a pleasant shape of the spline is expected. One of the natural ways to approximate (4) is to use a particular quadrature. To keep things as simple as possible, the trapezoidal rule will be chosen here. Thus

$$\varphi_i(\boldsymbol{\alpha}) = \int_{t_{i-1}}^{t_i} \|\ddot{\boldsymbol{s}}_i(t)\|^2 \, dt \approx \frac{\Delta t_{i-1}}{2} \left(\|\ddot{\boldsymbol{s}}_i(t_{i-1})\|^2 + \|\ddot{\boldsymbol{s}}_i(t_i)\|^2 \right). \tag{6}$$

However $\|\ddot{\boldsymbol{s}}_i(t_{i-1})\|$ and $\|\ddot{\boldsymbol{s}}_i(t_i)\|$ both depend on tangent directions $\boldsymbol{d}_{i-1}$ and $\boldsymbol{d}_i$, thus similar conditions on admissible regions for tangent directions as stated in Theorem 1 are expected. In order to avoid this, a further approximation of (6) will be done. The main idea is to find the best approximation of $\ddot{\boldsymbol{s}}_i(t_{i-1})$, given as a linear combination of $\boldsymbol{s}_i(t_{i-1})$, $\dot{\boldsymbol{s}}_i(t_{i-1})$, and $\boldsymbol{s}_i(t_i)$, and similarly, the best approximation of $\ddot{\boldsymbol{s}}_i(t_i)$ by $\boldsymbol{s}_i(t_{i-1})$, $\dot{\boldsymbol{s}}_i(t_i)$ and $\boldsymbol{s}_i(t_i)$. The best approximation should be considered here as an approximation which is exact on the polynomial space of order $\le k$, where k is as large as possible. The method of undetermined coefficients or the Newton's form of the interpolation polynomial lead to

$$\ddot{\boldsymbol{s}}_i(t_{i-1}) \approx 2\,[t_{i-1}, t_{i-1}, t_i]\boldsymbol{s}_i, \quad \ddot{\boldsymbol{s}}_i(t_i) \approx 2\,[t_{i-1}, t_i, t_i]\boldsymbol{s}_i,$$

and by (6), φ_i can be approximated by

$$\varphi_i(\boldsymbol{\alpha}) \approx \frac{2}{\Delta t_{i-1}}\psi_i(\boldsymbol{\alpha}), \tag{7}$$

where

$$\psi_i(\boldsymbol{\alpha}) := \left\| \frac{1}{\Delta t_{i-1}} \Delta \boldsymbol{T}_{i-1} - \alpha_{i,0}\,\boldsymbol{d}_{i-1} \right\|^2 + \left\| \alpha_{i,1}\,\boldsymbol{d}_i - \frac{1}{\Delta t_{i-1}} \Delta \boldsymbol{T}_{i-1} \right\|^2. \tag{8}$$

Theorem 2 *The nonlinear functional ψ_i, $i = 1, 2, \ldots, n$, has a unique global minimum in the interior of $\mathcal{D}_i$ iff*

$$\boldsymbol{\alpha}_i^* := \frac{1}{\Delta t_{i-1}} \left(\boldsymbol{d}_{i-1} \cdot \Delta \boldsymbol{T}_{i-1}, \boldsymbol{d}_i \cdot \Delta \boldsymbol{T}_{i-1}\right)^T > \boldsymbol{0}.$$

Furthermore,

$$\min_{\boldsymbol{\alpha}_i \in \mathcal{D}_i} \psi_i(\boldsymbol{\alpha}_i) = \frac{2 - c_{i,0}^2 - c_{i,1}^2}{(\Delta t_{i-1})^2} \, \|\Delta \boldsymbol{T}_{i-1}\|^2,$$

where

$$c_{i,k} = \cos \angle \left(\boldsymbol{d}_{i+k-1}, \Delta \boldsymbol{T}_{i-1}\right), \quad k = 0, 1.$$

PROOF. The functional (8) can be simplified to

$$\begin{aligned} \psi_i(\boldsymbol{\alpha}_i) = \frac{1}{(\Delta t_{i-1})^2} \Big(&(\Delta t_{i-1})^2 \left(\alpha_{i,0}^2 + \alpha_{i,1}^2\right) \\ &- 2\,\Delta t_{i-1} (\alpha_{i,0}\,\boldsymbol{d}_{i-1} + \alpha_{i,1}\,\boldsymbol{d}_i) \cdot \Delta \boldsymbol{T}_{i-1} + 2\,\|\Delta \boldsymbol{T}_{i-1}\|^2 \Big). \end{aligned}$$

Note that ψ_i is convex. Its gradient vanishes at $\boldsymbol{\alpha}_i^* := \left(\alpha_{i,0}^*, \alpha_{i,1}^*\right)^T$, where

$$\alpha_{i,k}^* = \frac{\boldsymbol{d}_{i+k-1} \cdot \Delta \boldsymbol{T}_{i-1}}{\Delta t_{i-1}}, \quad k = 0, 1,$$

which leads to a unique minimum

$$\psi_i(\boldsymbol{\alpha}_i^*) = \frac{2 - c_{i,0}^2 - c_{i,1}^2}{(\Delta t_{i-1})^2} \, \|\Delta \boldsymbol{T}_{i-1}\|^2.$$

The point $\boldsymbol{\alpha}_i^*$ is in $\mathcal{D}_i$ iff its components are positive. This concludes the proof of the theorem.

Notice that the minimal value of the functional ψ_i can also be zero. In this case $c_{i,0} = c_{i,1} = 1$ (or -1, but in this case the curve has an undesired fold), and the cubic parametric spline segment $\boldsymbol{s}_i$ reduces to a straight line $\boldsymbol{s}_i(t) = \boldsymbol{T}_{i-1} + (t - t_{i-1})[t_{i-1}, t_i]\boldsymbol{s}_i$.

Corollary 3 *The conditions $\boldsymbol{\alpha}_i^* > \boldsymbol{0}$, $i = 1, 2, \ldots, n$, have a simple geometric interpretation, i.e., $\angle(\boldsymbol{d}_{i+k-1}, \Delta\boldsymbol{T}_{i-1}) \in [0, \frac{\pi}{2})$, $k = 0, 1$.*

Now suppose that the assumptions of Theorem 2 are satisfied. Then an important question arises whether the resulting cubic spline segment $\boldsymbol{s}_i$ is regular on $[t_{i-1}, t_i]$, $i = 1, 2, \ldots, n$. The answer is affirmative, even more, $\boldsymbol{s}_i$ has no cusps, loops or folds.

Theorem 4 *Let the assumptions of Theorem 2 be satisfied and let $\boldsymbol{s}_i$, $i = 1, 2, \ldots, n$, be the resulting Hermite geometric interpolant defined by* (1)*. Then the spline segment $\boldsymbol{s}_i$ is regular, loop-, cusp-, and fold-free.*

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

$$\begin{aligned} &\boldsymbol{p}_i(k) = \boldsymbol{T}_{i+k-1}, \quad k = 0, 1,\\ &\boldsymbol{p}_i'(k) = \Delta t_{i-1}\, \alpha_{i,k}\, \boldsymbol{d}_{i+k-1} = (\boldsymbol{d}_{i+k-1} \cdot \Delta\boldsymbol{T}_{i-1})\, \boldsymbol{d}_{i+k-1}, \quad k = 0, 1. \end{aligned}$$

Clearly, $\boldsymbol{p}_i$ can be written in the Bézier form as

$$\begin{aligned} \boldsymbol{p}_i(u) = \boldsymbol{T}_{i-1}\, B_0^3(u) &+ \left(\boldsymbol{T}_{i-1} + \frac{1}{3}(\boldsymbol{d}_{i-1} \cdot \Delta\boldsymbol{T}_{i-1})\, \boldsymbol{d}_{i-1}\right) B_1^3(u)\\ &+ \left(\boldsymbol{T}_i - \frac{1}{3}(\boldsymbol{d}_i \cdot \Delta\boldsymbol{T}_{i-1})\, \boldsymbol{d}_i\right) B_2^3(u) + \boldsymbol{T}_i\, B_3^3(u), \end{aligned}$$

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

$$\begin{aligned} \boldsymbol{p}_i'(u) = (\boldsymbol{d}_{i-1} \cdot \Delta\boldsymbol{T}_{i-1})\, \boldsymbol{d}_{i-1}\, B_0^2(u) &+ (3\,\Delta\boldsymbol{T}_{i-1} - (\boldsymbol{d}_{i-1} \cdot \Delta\boldsymbol{T}_{i-1})\, \boldsymbol{d}_{i-1}\\ &-(\boldsymbol{d}_i \cdot \Delta\boldsymbol{T}_{i-1})\, \boldsymbol{d}_i)\; B_1^2(u) + (\boldsymbol{d}_i \cdot \Delta\boldsymbol{T}_{i-1})\boldsymbol{d}_i\, B_2^2(u). \end{aligned}$$

Let $p_{i,1}'$ be the first component of $\boldsymbol{p}_i'$. To see that $\boldsymbol{p}_i$ is regular and loop-, cusp- and fold-free on $[0, 1]$, it is enough to verify that $p_{i,1}'(u) > 0$ for $u \in [0, 1]$ (see [6], e.g.). Quite clearly

$$p_{i,1}'(u) = x_1 \left(d_{i-1,1}^2\, B_0^2(u) + \left(3 - d_{i-1,1}^2 - d_{i,1}^2\right) B_1^2(u) + d_{i,1}^2\, B_2^2(u)\right).$$

Since $\boldsymbol{d}_i$ 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 $[0, 1]$. This concludes the proof of the theorem.

4 Construction of tangent directions

In the previous section, a construction of the cubic Hermite G^1 polynomial spline curve, based on a minimization of energy, has been derived. We have assumed that the unit tangent directions $\boldsymbol{d}_i$ are known in advance and the parameters $\boldsymbol{\alpha}_i^*$ are obtained by a particular minimization technique. As it is clear from Theorem 2, tangent directions must be chosen carefully.

In this section the problem of the construction of tangent directions $\boldsymbol{d}_i$ will be addressed. From Theorem 2 it is enough to require

$$\boldsymbol{d}_{i+k-1} \cdot \Delta \boldsymbol{T}_{i-1} > 0, \quad i = 1, 2, \ldots, n, \quad k = 0, 1.$$

In order to fulfill these conditions, consider the i-th and $(i+1)$-th segment of the spline curve $\boldsymbol{s}$ (see Fig. 3). Let us define a rotation

$$R := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{9}$$

and $z_i := \operatorname{sign}(\Delta \boldsymbol{T}_{i-1} \times \Delta \boldsymbol{T}_i)$. Further, let

$$\boldsymbol{u}_i := z_i\, R\, \Delta \boldsymbol{T}_{i-1}, \quad \boldsymbol{v}_i := -z_i\, R\, \Delta \boldsymbol{T}_i, \quad \boldsymbol{w}_i := \lambda_i\, \boldsymbol{u}_i + (1 - \lambda_i)\, \boldsymbol{v}_i, \ \lambda_i \in \mathbb{R}. \tag{10}$$

It is now easy to prove the following lemma.

Lemma 5 *If $z_i \neq 0$ and $\lambda_i \in (0, 1)$, then $\boldsymbol{w}_i \cdot \Delta \boldsymbol{T}_j > 0$, $j = i-1, i$.*

PROOF. We will prove that $\boldsymbol{w}_i \cdot \Delta \boldsymbol{T}_j > 0$, $j = i - 1$. The proof for $j = i$ is very similar and will be omitted. Take any $\boldsymbol{w}_i = \lambda_i\, \boldsymbol{u}_i + (1 - \lambda_i)\, \boldsymbol{v}_i$ with $\lambda_i \in (0, 1)$. By (9) and (10)

$$\begin{aligned} \boldsymbol{w}_i \cdot \Delta \boldsymbol{T}_{i-1} &= \lambda_i\, \boldsymbol{u}_i \cdot \Delta \boldsymbol{T}_{i-1} + (1 - \lambda_i)\, \boldsymbol{v}_i \cdot \Delta \boldsymbol{T}_{i-1} \\ &= -(1 - \lambda_i)\, z_i\, (R\, \Delta \boldsymbol{T}_i) \cdot \Delta \boldsymbol{T}_{i-1}, \end{aligned}$$

since $\boldsymbol{u}_i$ and $\Delta \boldsymbol{T}_{i-1}$ are perpendicular to each other. Obviously, it is enough to show that

$$z_i = -\operatorname{sign}\left((R\, \Delta \boldsymbol{T}_i) \cdot \Delta \boldsymbol{T}_{i-1}\right).$$

Since $z_i \neq 0$, $\angle(\Delta \boldsymbol{T}_{i-1}, \Delta \boldsymbol{T}_i) \neq 0, \pi$, and there are two possibilities.

(a) If $z_i > 0$, then $\angle(R\Delta \boldsymbol{T}_i, \Delta \boldsymbol{T}_{i-1}) > \pi/2$ and $\operatorname{sign}((R\Delta \boldsymbol{T}_i) \cdot \Delta \boldsymbol{T}_{i-1}) < 0$.
(b) If $z_i < 0$, then $\angle(R\Delta \boldsymbol{T}_i, \Delta \boldsymbol{T}_{i-1}) < \pi/2$ and $\operatorname{sign}((R\Delta \boldsymbol{T}_i) \cdot \Delta \boldsymbol{T}_{i-1}) > 0$.

This concludes the proof of the lemma.

From Lemma 5 it follows that any $\lambda_i \in (0,1)$ gives $\boldsymbol{w}_i$ that leads to the minimization of the functional (8). Namely, one just takes $\boldsymbol{d}_i = \boldsymbol{w}_i/\|\boldsymbol{w}_i\|$. Thus λ_i, $i = 1, 2, \ldots, n$, can be considered as free shape parameters. One of the choices would, e.g., be $\lambda_i = \|\Delta \boldsymbol{T}_i\|/(\|\Delta \boldsymbol{T}_{i-1}\| + \|\Delta \boldsymbol{T}_i\|)$, which implies that $\boldsymbol{d}_i$ points from $\boldsymbol{T}_i$ in the direction of the bisector of the angle $\angle(\Delta \boldsymbol{T}_{i-1}, \Delta \boldsymbol{T}_i)$ (see Fig. 3 and Corollary 7). This is a natural heuristic choice since this $\boldsymbol{d}_i$ stays away

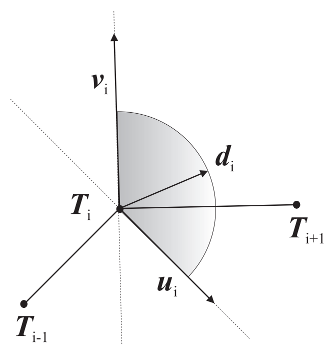

Fig. 3. Admissible positions of $\boldsymbol{d}_i$.

from the unwanted directions implying $\alpha_{i,k} = 0$, for $k = 0$ or $k = 1$, as much as possible. Lemma 5 excludes two possibilities, namely $\angle(\Delta \boldsymbol{T}_{i-1}, \Delta \boldsymbol{T}_i) = 0, \pi$. But, if the considered angle is equal to 0, then $\boldsymbol{w}_i$ can be taken as $\boldsymbol{w}_i = \Delta \boldsymbol{T}_i$, and again the conclusions of the lemma follow. On the other hand, the case when the angle equals π 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 $\boldsymbol{d}_0$ and $\boldsymbol{d}_n$ can be, e.g., chosen as $\boldsymbol{d}_0 = \Delta \boldsymbol{T}_0/\|\Delta \boldsymbol{T}_0\|$ and $\boldsymbol{d}_n = \Delta \boldsymbol{T}_{n-1}/\|\Delta \boldsymbol{T}_{n-1}\|$. Thus for given shape parameters $\lambda_i \in (0,1)$, the tangent directions and the resulting Hermite G^1 cubic spline can be constructed. Theorem 2 allows a broad range of admissible tangent directions. Then the spline is constructed locally. A natural question arises, which is the optimal admissible tangent direction. The answer is given by the following theorem.

Theorem 6 *The optimal tangent directions $\boldsymbol{d}_i$, such that the approximate* *strain energy* (7) *is minimized, are obtained by* $\boldsymbol{d}_i := \boldsymbol{w}_i/\|\boldsymbol{w}_i\|$, (10), *and*

$$\lambda_i = \frac{2\,(\Delta t_i)^3\ \boldsymbol{u}_i \cdot \boldsymbol{v}_i}{A_1 \pm \sqrt{A_2}}, \tag{11}$$

with

$$\begin{aligned} A_1 &:= (\Delta t_{i-1})^3\,\|\Delta \boldsymbol{T}_i\|^2 + 2\,(\Delta t_i)^3\ \boldsymbol{u}_i \cdot \boldsymbol{v}_i - (\Delta t_i)^3\,\|\Delta \boldsymbol{T}_{i-1}\|^2, \\ A_2 &:= \left((\Delta t_i)^3\,\|\Delta \boldsymbol{T}_{i-1}\|^2 - (\Delta t_{i-1})^3\,\|\Delta \boldsymbol{T}_i\|^2\right)^2 \\ &\quad + 4\,(\Delta t_{i-1})^3\,(\Delta t_i)^3\,(\boldsymbol{u}_i \cdot \boldsymbol{v}_i)^2. \end{aligned}$$

If $\boldsymbol{u}_i \cdot \boldsymbol{v}_i = 0$, *then*

a) $\lambda_i = 0$ *if* $z_i = -1$, *or*
b) $\lambda_i = 1$ *if* $z_i = 1$.

PROOF. Our goal is to find optimal tangent directions $\boldsymbol{d}_i$, such that the approximate strain energy (7) is minimized. Recall that if the tangent directions are already known, by Theorem 2 optimal $\boldsymbol{\alpha}_i$ are obtained. Since the direction $\boldsymbol{d}_i$ appears just in two neighbouring segments in (8), it is enough to minimize the expression

$$\frac{2}{\Delta t_{i-1}}\left\|\alpha_{i,1}\boldsymbol{d}_i - \frac{1}{\Delta t_{i-1}}\Delta \boldsymbol{T}_{i-1}\right\|^2 + \frac{2}{\Delta t_i}\left\|\frac{1}{\Delta t_i}\Delta \boldsymbol{T}_i - \alpha_{i+1,0}\boldsymbol{d}_i\right\|^2.$$

By using Theorem 2 and (10), computing partial derivatives on λ_i, and a somewhat tedious computation, the following quadratic equation is obtained

$$\begin{aligned} \rho(\lambda_i) := \lambda_i^2 &\left(((\Delta t_{i-1})^3 - (\Delta t_i)^3)\boldsymbol{u}_i \cdot \boldsymbol{v}_i + (\Delta t_i)^3\,\|\Delta \boldsymbol{T}_{i-1}\|^2 - (\Delta t_{i-1})^3\,\|\Delta \boldsymbol{T}_i\|^2\right) \\ &+ \lambda_i\left((\Delta t_{i-1})^3\,\|\Delta \boldsymbol{T}_i\|^2 + 2\,(\Delta t_i)^3\ \boldsymbol{u}_i \cdot \boldsymbol{v}_i - (\Delta t_i)^3\,\|\Delta \boldsymbol{T}_{i-1}\|^2\right) \\ &- (\Delta t_i)^3\ \boldsymbol{u}_i \cdot \boldsymbol{v}_i = 0. \end{aligned}$$

Note that $\rho(0)\,\rho(1) < 0$, which implies that there is always a unique solution in $(0,1)$. This determines the choice of the sign in (11). The case $\boldsymbol{u}_i \cdot \boldsymbol{v}_i = 0$ has to be analyzed separately. This concludes the proof.

For a particular parameterization, the expression (11) considerably simplifies. In this case a part of A_2 vanishes together with the square root. The following result is obtained.

Corollary 7 *For the* 2/3*-parameterization (see [7]),*

$$\frac{\Delta t_{i-1}}{\Delta t_i} = \left(\frac{\|\Delta \boldsymbol{T}_{i-1}\|}{\|\Delta \boldsymbol{T}_i\|}\right)^{\frac{2}{3}},$$

the optimal tangent directions $\boldsymbol{d}_i$ *lie on bisectors of angles* $\angle(\boldsymbol{u}_i, \boldsymbol{v}_i)$,

$$\lambda_i := \frac{\|\Delta \boldsymbol{T}_i\|}{\|\Delta \boldsymbol{T}_{i-1}\| + \|\Delta \boldsymbol{T}_i\|}.$$

5 Numerical examples

Let us conclude the paper by some numerical examples. The cubic G^1 Hermite interpolant can closely resemble the C^2 cubic interpolating spline (Fig. 4), but for non-uniformly distributed data points (exchange of short and long segments) larger differences between the curve shapes can occur.

The influence of the shape parameters λ_i on the curve can be clearly seen in (Fig. 4, right). For a comparison, approximate strain energies (3) for all the curves are computed. Clearly the energy of the C^2 spline is the smallest, but in some cases a simple ad-hoc procedure of selecting the bisector direction as an admissible tangent direction, gives a curve with suitably small approximate strain energy.

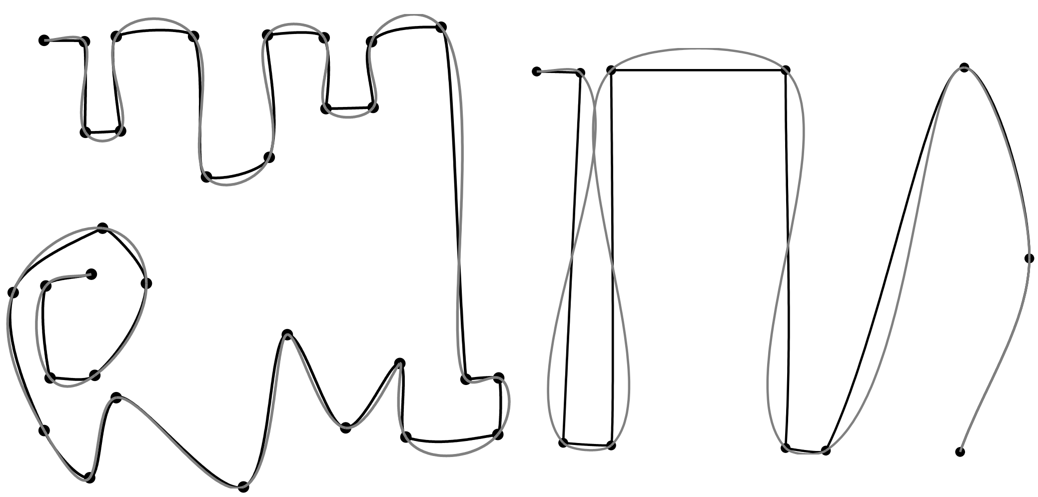

Fig. 4. The cubic G^1 Hermite interpolant (black) closely resembles the C^2 cubic interpolating spline (gray) (left figure). Approximate strain energies are 218.8 and 144.3, respectively. The optimal cubic G^1 Hermite interpolant (black) together with the one, obtained by an ad-hoc bisector choice of admissible tangent directions are shown in the right figure. Approximate strain energies are 599.2 and 740.1, respectively.

In Fig. 5, a comparison of curves, obtained by our method and [20], is presented. The data points are sampled from a unit circle. Since the algorithm in [20] requires tangent directions, two versions are considered. In the left figure, the gray curve is obtained by using tangents sampled from the circle, and in the right figure it is obtained by using tangent directions, computed by our method.

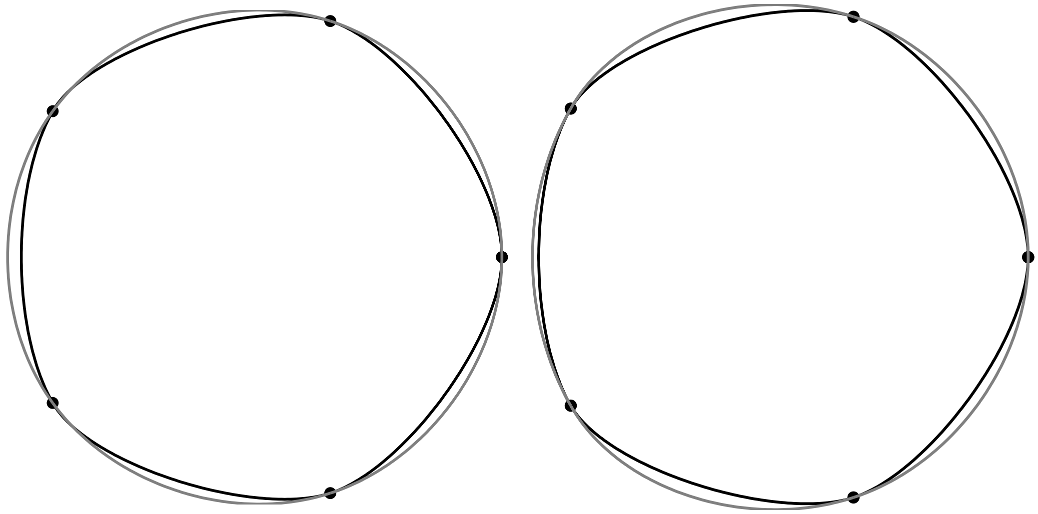

Fig. 5. Comparison of the cubic G^1 Hermite interpolant (black) with [20] using the data points and tangent directions sampled from a unit circle (left), and using the tangent directions, computed by our algorithm (right).

In general, the method from [20] needs preprocessing and inserting new, artificial points. There are 11 particular cases, but not all of them are explicitly derived. This makes an objective comparison of methods hard to do. Thus in Fig. 5 the data were chosen in such a way that no additional points were required. The results show that the shapes of both curves are comparable. The time complexities of both methods are similar. Our method requires some additional operations for tangent direction computation, however, [20] needs tangent directions as data, and additional operations are needed for choosing a particular of 11 subroutines and new data construction.

References

[1] C. Conti and R. Morandi. Piecewise C^1-shape-preserving Hermite interpolation. *Computing*, 56(4):323–341, 1996.

[2] P. Costantini and M. L. Sampoli. A general scheme for shape preserving planar interpolating curves. *BIT*, 43(2):297–317, 2003.

[3] I. Cravero and C. Manni. Shape-preserving interpolants with high smoothness. *J. Comput. Appl. Math.*, 157(2):383–405, 2003.

[4] C. de Boor. *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York, revised edition, 2001.

[5] C. de Boor, K. Höllig, and M. Sabin. High accuracy geometric Hermite interpolation. *Comput. Aided Geom. Design*, 4(4):269–278, 1987.

[6] G. Farin. *Curves and surfaces for computer-aided geometric design*. Computer Science and Scientific Computing. Academic Press Inc., San Diego, CA, fourth edition, 1997.

[7] M. S. Floater. On the deviation of a parametric cubic spline interpolant from its data polygon. *Comput. Aided Geom. Design*, 25(3):148–156, 2008.

[8] T. N. T. Goodman and K. Unsworth. Shape-preserving interpolation by parametrically defined curves. *SIAM J. Numer. Anal.*, 25(6):1453–1465, 1988.

[9] K. Höllig and J. Koch. Geometric Hermite interpolation with maximal order and smoothness. *Comput. Aided Geom. Design*, 13(8):681–695, 1996.

[10] G. Jaklič, J. Kozak, M. Krajnc, and E. Žagar. On geometric interpolation of circle-like curves. *Comput. Aided Geom. Design*, 24(5):241–251, 2007.

[11] G. Jaklič, J. Kozak, M. Krajnc, and E. Žagar. On geometric interpolation by planar parametric polynomial curves. *Math. Comp.*, 76(260):1981–1993, 2007.

[12] P. D. Kaklis and N. S. Sapidis. Convexity-preserving interpolatory parametric splines of non-uniform polynomial degree. *Comput. Aided Geom. Design*, 12(1):1–26, 1995.

[13] J. Kozak and M. Krajnc. Geometric interpolation by planar cubic G^1 splines. *BIT*, 47(3):547–563, 2007.

[14] J. Kozak and E. Žagar. On geometric interpolation by polynomial curves. *SIAM J. Numer. Anal.*, 42(3):953–967, 2004.

[15] E. T. Y. Lee. Choosing nodes in parametric curve interpolation. *Comput. Aided Design*, 21(6):363–370, 1989.

[16] K. Mørken and K. Scherer. A general framework for high-accuracy parametric interpolation. *Math. Comp.*, 66(217):237–260, 1997.

[17] S. Pruess. Shape preserving C^2 cubic spline interpolation. *IMA J. Numer. Anal.*, 13(4):493–507, 1993.

[18] A. Rababah. High order approximation method for curves. *Comput. Aided Geom. Design*, 12(1):89–102, 1995.

[19] J. Wallner. Note on curve and surface energies. *Comput. Aided Geom. Design*, 24(8-9):494–498, 2007.

[20] J.-H. Yong and F. Cheng. Geometric Hermite curves with minimum strain energy. *Comput. Aided Geom. Design*, 21(3):281–301, 2004.