Applied Mathematics and Computation 276 (2016) 172–181

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

A constructive framework for minimal energy planar curves

Michael J. Johnson [a,*], Hakim S. Johnson [b]

[a] Department of Mathematics, Faculty of Science, Kuwait University, Kuwait
[b] Kuwait English School, Kuwait

ARTICLE INFO

MSC:
65D05
65D17
65D07

Keywords:
Geometric interpolation
Hermite interpolation
Parametric cubic spline
Nonlinear spline

ABSTRACT

Given points $P_1, P_2, \ldots, P_n$ in the plane, we are concerned with the problem of finding a fair curve which interpolates the points. We assume that we have a method in hand, called a basic curve method, for solving the geometric Hermite interpolation problem of fitting a regular C^∞ curve between two points with prescribed tangent directions at the endpoints. We also assume that we have an energy functional which defines the energy of any basic curve. Using this basic curve method repeatedly, we can then construct G^1 curves which interpolate the given points $P_1, P_2, \ldots, P_n$. The tangent directions at the interpolation points are variable and the idea is to choose them so that the energy of the resulting curve (i.e., the sum of the energies of its pieces) is minimal. We give sufficient conditions on the basic curve method, the energy functional, and the interpolation points for (a) existence, (b) G^2 regularity, and (c) uniqueness of minimal energy interpolating curves. We also identify a one-parameter family of basic curve methods, based on parametric cubics, whose minimal energy interpolating curves are unique and G^2 under suitable conditions. One member of this family looks very promising and we suggest its use in place of conventional C^2 parametric cubic splines.

1. Introduction

Let $P_1, P_2, \ldots, P_n$ be a sequence of points in the complex plane $\mathbb{C}$ satisfying $P_j \neq P_{j+1}$, and consider the problem of finding a 'fair' curve which passes sequentially (i.e., interpolates) the points. Whereas there do not exist interpolating curves with minimal bending energy (see [3] and also [10]), except when the points lie sequentially along a line, it was shown recently [2] that they do exist if one imposes the additional constraint that each piece of the interpolating curve be an s-curve (a curve which turns monotonically at most 180° in one direction and then turns monotonically at most 180° in the opposite direction). Such interpolating curves with minimal bending energy are called *elastic splines*. While work on [2] was in progress, the authors of the present article took up the numerical challenge of computing elastic splines. As in [7], the problem was formulated as an optimization problem where the interpolation points $P_1, P_2, \ldots, P_n$ are given but corresponding tangent directions $d_1, d_2, \ldots, d_n$ are variable. An important sub-problem, which was extensively addressed in [2], is that of finding an s-curve with minimal bending energy which solves the first order geometric Hermite interpolation problem of constructing a curve which begins at P_j with direction d_j and ends at P_{j+1} with direction d_{j+1}. The s-curve condition places feasibility restrictions on the directions d_j and d_{j+1}. In case $P_j = 0$ and P_{j+1} lies on the positive real axis (which can be obtained by a translation and rotation) and writing $d_j = e^{i\alpha}$ and $d_{j+1} = e^{i\beta}$, these feasibility restrictions reduce to the inequalities $|\alpha|, |\beta| < \pi$ and $|\alpha - \beta| \leq \pi$. The presence of these coupled restrictions on the directions $\{d_j\}$ gives rise to a rather complicated feasible region in $\mathbb{C}^n$ and this in turn complicates

* Corresponding author. Tel.: +96524985329.
E-mail address: yohnson1963@hotmail.com (M.J. Johnson).

http://dx.doi.org/10.1016/j.amc.2015.11.047
0096-3003/

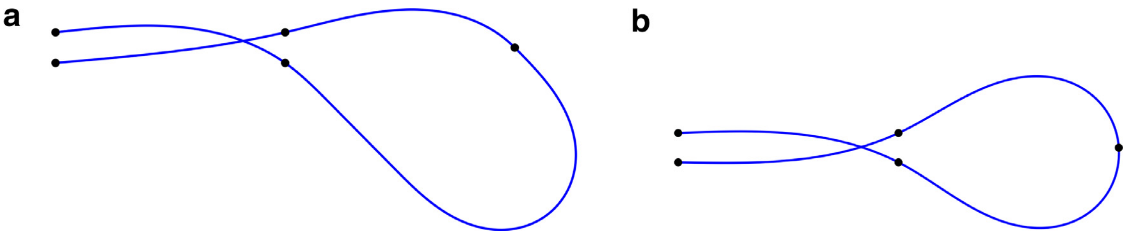

Fig. 1. An elastic spline (a) before and (b) after the restriction $|\alpha|, |\beta| \le \pi/2$.

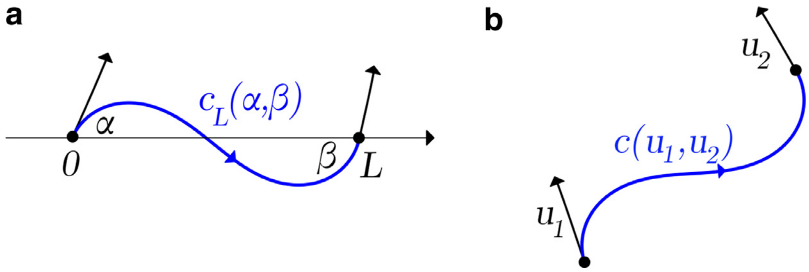

Fig. 2. Basic curves in (a) canonical and (b) general position.

the optimization algorithm (see [9] for an alternative feasible region). After completing the numerics, it was observed that the elastic splines were often fair, but could also be unsightly, particularly when the interpolation points force abrupt changes in direction (see Fig. 1a). After much experimentation, we decided that the best way to eliminate the unsightly elastic splines is to replace the s-curve feasibility inequalities $|\alpha|, |\beta| < \pi$ and $|\alpha - \beta| \le \pi$ with the simple restriction $|\alpha|, |\beta| \le \pi/2$ (see Fig. 1b). With these simple uncoupled restrictions on $\{d_j\}$, the feasible region in $\mathbb{C}^n$ reduces to a Cartesian product and the optimization algorithm simplifies to that of optimizing each direction individually while cycling through the points. In addition to eliminating the unsightly elastic splines, the simplified restriction $|\alpha|, |\beta| \le \pi/2$ also makes the theoretic study of elastic splines much more tractable. Moreover, the basic theory can be developed in a general context where potentially any method for solving the above-mentioned first order geometric Hermite interpolation problem can be used in place of that for elastic splines.

We now describe the basic setup. A **unit tangent vector** $u = (P, d)$ is an ordered pair of complex numbers with $|d| = 1$ and can be visualized as a directed line segment with **base-point** P and **direction** d. A C^∞ regular **curve** is a C^∞ function $f : [a, b] \to \mathbb{C}$ whose first derivative f' is non-vanishing. We say that f **connects** $u_1 = (P_1, d_1)$ to $u_2 = (P_2, d_2)$ if $f(a) = P_1,\ f'(a) = |f'(a)|d_1,\ f(b) = P_2$ and $f'(b) = |f'(b)|d_2$. We use the term basic curve method to refer to a method for solving the first order geometric Hermite interpolation problem mentioned above. Precisely, a **basic curve method** is a mapping $(\alpha, \beta, L) \mapsto c_L(\alpha, \beta)$, which is defined for angles $\alpha, \beta \in [-\Omega, \Omega]$ and lengths $L > 0$ ($\Omega \in (0, \pi)$ is a given constant), whose image $c_L(\alpha, \beta)$ is a C^∞ regular curve which connects $u = (0, e^{i\alpha})$ to $v = (L, e^{i\beta})$ (see Fig. 2a). Associated with the basic curve method is a functional E_L, whereby the 'energy' of the curve $c_L(\alpha, \beta)$ equals $E_L(\alpha, \beta)$. In practice, the energy of $c_L(\alpha, \beta)$ is often its bending energy, defined by $\frac{1}{2}\int_a^b [\kappa(s)]^2 \frac{ds}{dt}\, dt$ with κ denoting signed curvature and s arclength, or an approximation of bending energy, such as $\frac{1}{2}\int_a^b |f''(t)|^2\, dt$, but there is no theoretical requirement that energy has a physical interpretation. It could just as well be the *cosmic energy* of the curve. The basic curve method is extended to other pairs of unit tangent vectors by the use of translation and rotation (see Fig. 2b). Specifically, let $u_1 = (P_1, d_1)$ and $u_2 = (P_2, d_2)$ be two unit tangent vectors with distinct base points, and set $\alpha = \arg \frac{d_1}{P_2 - P_1}$, $\beta = \arg \frac{d_2}{P_2 - P_1}$ and $L = |P_2 - P_1|$. Here arg is defined, as usual, by $\arg re^{i\theta} = \theta$ when $r > 0$ and $\theta \in (-\pi, \pi]$. If the angles α and β belong to $[-\Omega, \Omega]$, then the basic curve connecting u_1 to u_2 is defined by $c(u_1, u_2) := T \circ c_L(\alpha, \beta)$, where the transformation $T(z) = a_1 z + a_2$ is determined by the requirements $T(0) = P_1$ and $T(L) = P_2$ (i.e., $a_1 = (P_2 - P_1)/L$ and $a_2 = P_1$). The energy of $c(u_1, u_2)$ is defined by $\mathrm{Energy}(c(u_1, u_2)) := E_L(\alpha, \beta)$. As a consequence of these definitions, the extended basic curve method and its energy functional are invariant under translations and rotations.

Examples of basic curve methods pertaining to parametric cubics are given in [7,12]; we will have more to say about these in Section 4. A basic curve method employing A-splines is given in [1], and we mention that second order basic curve methods (where curvature data is also interpolated) can be found in [5,11] and the references therein.

With a basic curve method in hand, one can construct G^1 curves which interpolate the points $P_1, P_2, \ldots, P_n$ (following [7]) by assigning 'feasible' directions $d_1, d_2, \ldots, d_n$, and then use the resultant unit tangent vectors $u_j := (P_j, d_j)$ to obtain an interpolating curve $c(u_1, u_2) \sqcup c(u_2, u_3) \sqcup \cdots \sqcup c(u_{n-1}, u_n)$, called an **admissible curve**, whose energy is defined to be the sum of energies of its constituent pieces. Here, the directions $d_1, d_2, \ldots, d_n$ are deemed **feasible** if all of the basic curves $c(u_j, u_{j+1})$ are defined. The goal is then to choose the feasible directions $\{d_j\}$ so that the energy of the corresponding admissible curve is minimized.

Our primary purpose is to prove sufficient conditions (and occasionally necessary conditions) for (a) existence, (b) G^2 regularity, and (c) uniqueness of minimal energy admissible curves. These sufficient conditions depend on the constant Ω, which