CURVE FITTING
WITH
NONLINEAR SPIRAL SPLINES

by

Ian D. Coope

Department of Mathematics, University of Canterbury, Christchurch, New Zealand.

No. 63 July, 1991.

Abstract. Collocating spiral splines are derived as an approximation to the curve of least energy. The defining equations, although nonlinear, are easily solved because the Jacobian matrix has banded structure. A simple but effective iterative scheme for the solution of these equations is described together with a useful scheme for determining initial approximations for nonlinear splines. The resulting curve is invariant with respect to translation and rotation of axes and is usually much smoother than is possible with polynomial splines because the curvature of the spiral spline varies linearly with respect to arc length.

Key words: curve fitting, spiral spline, nonlinear spline, least energy, interpolation.

AMS(MOS) subject classifications: 65D07, 65D10, 41A15.

CURVE FITTING WITH NONLINEAR SPIRAL SPLINES

Ian D. Coope

Department of Mathematics
University of Canterbury

14 January 1991

Abstract. Collocating spiral splines are derived as an approximation to the curve of least energy. The defining equations, although nonlinear, are easily solved because the Jacobian matrix has banded structure. A simple but effective iterative scheme for the solution of these equations is described together with a useful scheme for determining initial approximations for nonlinear splines. The resulting curve is invariant with respect to translation and rotation of axes and is usually much smoother than is possible with polynomial splines because the curvature of the spiral spline varies linearly with respect to arc length.

Key words: curve fitting, spiral spline, nonlinear spline, least energy, interpolation.

AMS(MOS) subject classifications: 65D07, 65D10, 41A15.

1 Introduction

The problem of fitting a smooth, twice continuously differentiable plane curve through a given set of data points, (x_j, y_j), $j = 1, 2, \ldots, n$, is considered. The curve to be calculated is required to pass through the given points in the order prescribed and be invariant with respect to rotation and translation of axes. The least energy spline, also called 'true nonlinear spline', is one such curve which has received renewed attention recently [4, 7, 8] and this is the starting point for the investigation in this paper. After deriving the differential equation describing the least energy spline a simple approximation leads to the development of the 'nonlinear spiral spline' which shares the required desirable properties of 'invariance' and 'smoothness' and which gives visually pleasing curves. Although nonlinear, the defining equations for the spiral spline are quite easily solved and an algorithm for calculating the required curve, together with a highly useful technique for automatically generating initial approximations to nonlinear splines (applicable also to the least energy spline) are presented.

[1] *Dedicated to my parents.*

Stoer[18] has also proposed the use of spiral splines (see also Mehlum[13, 15] and Pal and Nutbourne[16]) and this paper has much in common with his approach. An important difference is that Stoer determines a spiral spline by minimizing its energy whereas the approach taken here is to show first that spiral splines are good approximations to the true nonlinear spline (c.f. Mehlum[15]) and then to fix the particular spiral spline by imposing appropriate end conditions.

2 The variational problem

In this section the differential equation describing a least energy curve passing through two points with Cartesian coordinates $(x_1, y_1), (x_2, y_2)$, and having prescribed inclinations ψ_1 and ψ_2 at these points is derived. Extensive use is made of intrinsic coordinates (ψ, s), where $\psi(s)$ denotes the inclination, (measured from the positive x-axis as angular origin), at the point at distance s measured along the curve. The variational problem is to determine that curve, with continuous curvature, which minimizes the functional:

$$I[\psi] = \int_0^L \left(\frac{d\psi}{ds}\right)^2 ds \tag{1}$$

and which satisfies the collocation conditions:

$$x_2 - x_1 = \int_0^L \cos\psi \, ds \tag{2}$$

$$y_2 - y_1 = \int_0^L \sin\psi \, ds \tag{3}$$

where L is the (unknown) curve length (setting $s=0$ at (x_1, y_1)). Introducing the Lagrangian function,

$$F(s, \psi, \psi') = (\psi')^2 + \lambda\cos\psi + \mu\sin\psi \tag{4}$$

where λ and μ are Lagrange multipliers, the Euler-Lagrange equation for the variational problem can be written

$$\frac{\partial F}{\partial\psi} - \frac{d}{ds}\left(\frac{\partial F}{\partial\psi'}\right) = 0 \tag{5}$$

with boundary conditions, $\psi(0)=\psi_1$, $\psi(L)=\psi_2$. Because L is unknown this latter boundary condition leads to the transversality condition (see, for example, [5])

$$\left[F - \psi'\frac{\partial F}{\partial\psi'} = 0\right]_{\psi=\psi_2}$$

which on substituting for F from equation (4) provides the condition

$$\left[(\psi')^2 = \lambda\cos\psi + \mu\sin\psi\right]_{\psi=\psi_2} \tag{6}$$

Moreover, with F defined by (4) equation (5) takes the simple form

$$-\lambda\sin\psi + \mu\cos\psi - \frac{d}{ds}(2\psi') = 0$$

which is easily integrated to give

$$\left(\frac{d\psi}{ds}\right)^2 = \lambda \cos\psi + \mu \sin\psi \quad (+\text{constant}) \tag{7}$$

and the constant of integration is clearly zero from the transversality condition (6). Equation (7) is well known as an intrinsic differential equation of an elastica and has been derived by several different (and sometimes lengthy) methods (see, for example, [2, 7, 8, 9, 11, 13, 17]) and it may be integrated to give $x(s)$ and $y(s)$ in terms of elliptic functions [9]. Thus in principle it is possible to express ψ as a function of s involving three as yet unknown parameters, λ, μ, and the constant of integration arising from the solution of the nonlinear differential equation (7). These three parameters together with the unknown curve length L must then be determined by satisfying the two collocation conditions (2), (3) and the two prescribed end inclinations ψ_1, ψ_2 in order to complete the specification of the required curve in parametric form.

To fit a curve with continuous slope and curvature through a set of points $\{(x_j, y_j)\}_1^n$ the unknown end inclinations at interior knots are replaced by continuity conditions for slope and curvature at each interior knot. This was essentially the approach taken in 1966 by Larkin[9] in which the resulting system of nonlinear algebraic equations was solved by a simple, although somewhat slow, relaxation method. This appears to be the first successful attempt to calculate accurate solutions to the curve of least energy. His method is reviewed together with other early approaches [6, 14, 19] in the survey paper by Malcolm[12]. More recently, a new approach to solving the exact equations of the nonlinear least energy spline is presented by Edwards[4]. He obtains simplifications in the spline representation by using a local coordinate system on each subinterval and solves the resulting nonlinear system of algebraic equations by an efficient modified Newton iteration. Thus there are methods available for constructing least energy splines. However, there are still some fundamental difficulties. Existence of a finite, stable equilibrium solution cannot be guaranteed for an arbitrary configuration. A necessary condition for existence is established by Larkin[9] and is most easily seen by rewriting equation (7) (which must hold on each subinterval) in the form

$$\left(\frac{d\psi}{ds}\right)^2 = \rho_j^2 \cos(\psi - \alpha_j), \quad \psi \in [\psi_j, \psi_{j+1}]$$

Then the need to keep the right hand side non-negative on $[\psi_j, \psi_{j+1}]$ requires, in particular, that

$$|\psi_{j+1} - \psi_j| \le \pi \tag{8}$$

Unfortunately, the inclinations $\psi_j, j = 2, \ldots, n-1$, are not known a priori. Even when existence is not at issue it may be that many stable least energy splines exist for a given configuration since any solution necessarily corresponds to a local and not global minimum of the energy curve. This is a difficulty that is likely to arise with all nonlinear splines including the spiral splines introduced in the next section. For the present author, however, the main disadvantage with least energy curves is that in some cases the approach does not lead to very pleasing curves. For example, many people would agree that a circle is a very smooth curve and if the data points and end conditions are consistent with a circle then that is what should be fitted. The examples given by Lee and Forsythe[11], however, show that circles are not least energy curves. Therefore, the problem of approximating the least energy curve in a way that preserves the highly desirable property of invariance with respect to translation and rotation of axes and which fits circles when appropriate is considered.

3 Approximating the Curve of Least Energy

For configurations which result in small inclinations $\frac{d\psi}{ds} \approx \frac{d^2y}{dx^2}$ and a linearization of the variational problem (1–3) in Cartesian coordinates leads to the familiar cubic polynomial spline which is not rotationally invariant. However, linearizing the right hand side of equation (7) gives

$$\left(\frac{d\psi}{ds}\right)^2 = \lambda + \mu\psi$$

which has general solution

$$\psi = \frac{\mu}{4}(s-\gamma)^2 - \lambda/\mu \tag{9}$$

where γ is a constant of integration, and this does lead to a rotationally invariant spline as is shown later. Equation (9) is the intrinsic form of an Euler spiral, also called Cornu spiral or clothoid [10] and consequently a nonlinear spline comprising piecewise curves of the form (9) satisfying continuity of slope and curvature at the knots is subsequently referred to as a **spiral spline** - it includes a circle as a special case.

The spiral curve (9) expresses the inclination, ψ, as a quadratic function of the arc length s. In practice, it is more convenient to write the quadratic, $Q_j(s)$, on the jth sub-interval, $[s_j, s_{j+1}]$, using the Newton form of the interpolating quadratic through the points (s_j, ψ_j), (s_{j+1}, ψ_{j+1}) and $(\frac{s_j+s_{j+1}}{2}, \phi_j)$, where ϕ_j denotes the inclination half-way along the curve on this sub-interval (i.e. $\phi_j = Q_j(\frac{s_j+s_{j+1}}{2})$). Then the spiral spline takes the simple form

$$\psi = Q_j(s) = \psi_j + b_j(s-s_j) + c_j(s-s_j)(s-s_{j+1}), \quad s \in [s_j, s_{j+1}] \tag{10}$$

where

$$b_j = \frac{\psi_{j+1} - \psi_j}{L_j}, \qquad c_j = \frac{2(\psi_j - 2\phi_j + \psi_{j+1})}{L_j^2}$$

and L_j is the arc length, $L_j = s_{j+1} - s_j$.

Introducing $t = (s - s_j)/L_j$ and writing $P_j(t) = Q_j(s_j + tL_j)$ reduces the quadratic (10) to

$$P_j(t) = \psi_j + (\psi_{j+1} - \psi_j)t + 2(\psi_j - 2\phi_j + \psi_{j+1})t(t-1), \qquad t \in [0,1] \tag{11}$$

which is readily seen to satisfy $P_j(0) = \psi_j$, $P_j(\frac{1}{2}) = \phi_j$, $P_j(1) = \psi_{j+1}$, as required. The continuity of curvature conditions, $Q'_{j-1}(s_j) = Q'_j(s_j)$, $j = 2, 3, \ldots, n-1$, then become

$$\frac{1}{L_{j-1}}P'_{j-1}(1) = \frac{1}{L_j}P'_j(0), \quad j = 2, 3, \ldots, n-1,$$

which, on substituting the form for P from (11) and rearranging, lead to the equations:

$$L_j\psi_{j-1} + 3(L_{j-1} + L_j)\psi_j + L_{j-1}\psi_{j+1} = 4(L_j\phi_{j-1} + L_{j-1}\phi_j), \tag{12}$$
$$j = 2, 3, \ldots, n-1.$$

The collocation conditions (2), (3) also have a simple form when expressed in terms of $P(t)$,

$$\Delta x_j = L_j \int_0^1 \cos P_j(t)\, dt$$
$$\Delta y_j = L_j \int_0^1 \sin P_j(t)\, dt$$

which may be rewritten in the computationally more convenient form

$$\int_0^1 \sin(P_j(t) - \theta_j)\, dt = 0 \tag{13}$$

$$L_j = \hat{L}_j / \int_0^1 \cos(P_j(t) - \theta_j)\, dt \tag{14}$$

where $\hat{L}_j$ denotes the chord length,

$$\hat{L}_j = \sqrt{(\Delta x_j)^2 + (\Delta y_j)^2}\,, \tag{15}$$

and θ_j denotes the chord angle,

$$\theta_j = \arctan(\Delta y_j / \Delta x_j)\,, \tag{16}$$

and where $P_j(t)$ depends on the unknown parameters ψ_j, ϕ_j, and ψ_{j+1} through the definition (11). The $3n-4$ nonlinear equations (12), (13), (14) in the $3n-4$ unknowns $\{\psi_j\}_2^{n-1}$, $\{\phi_j\}_1^{n-1}$ and $\{L_j\}_1^{n-1}$ can then in principle be solved to give the intrinsic parametric form of the spiral spline. Of course, it is possible to eliminate the variables L_j, $j = 1, \ldots, n-1$, from the equations (12) by substituting from (14), and this reduces the size of the nonlinear system to $2n-3$ equations in the $2n-3$ unknowns $\{\psi_j\}_2^{n-1}, \{\phi_j\}_1^{n-1}$. However, the arc lengths are still needed to complete the description of the spiral spline in parametric form.

3.1 End Conditions

If the end inclinations ψ_1, ψ_n, are not known then other end conditions can be incorporated; for example, the 'curvature' end condition $\frac{d\psi}{ds} = \kappa_1$ at $s = 0$ leads to the extra equation $P_1'(0) = L_1\kappa_1$ or

$$-3\psi_1 + 4\phi_1 - \psi_2 = L_1\kappa_1\,,$$

and at the other end $\frac{d\psi}{ds} = \kappa_n$ at $s = L$ $(= \sum_1^{n-1} L_j)$ becomes $P_{n-1}'(1) = L_{n-1}\kappa_n$ or

$$-\psi_{n-1} + 4\phi_{n-1} - 3\psi_n = L_{n-1}\kappa_n\,.$$

The special case of 'natural' end conditions, $\kappa_1 = 0, \kappa_n = 0$, then gives the equations:

$$\begin{array}{rcrcl} 3\psi_1 & + & \psi_2 & = & 4\phi_1 \\ \psi_{n-1} & + & 3\psi_n & = & 4\phi_{n-1} \end{array} \tag{17}$$

Different end conditions such as 'closed loop' or 'not-a-knot' or others could also be applied in a similar fashion.

3.2 Energy

In terms of $P_j(t)$, the energy, $\mathcal{E}_j$, of the spiral spline on the jth span is

$$\mathcal{E}_j = \frac{1}{L_j}\int_0^1 \left(P_j'(t)\right)^2\, dt$$

and using the linearity of $P_j'(t)$ this evaluates to

$$\mathcal{E}_j = \frac{1}{3P_j''L_j}\left[\left(P_j'(t)\right)^3\right]_0^1$$

Using the linearity of $P_j'(t)$ again, gives $P_j'' = P_j'(1) - P_j'(0)$ and hence

$$\mathcal{E}_j = \frac{\left(P_j'(1)\right)^2 + P_j'(1)\,P_j'(0) + \left(P_j'(0)\right)^2}{3L_j}$$

which is the ratio of a quadratic form in the inclinations $\psi_j, \phi_j, \psi_{j+1}$, to the arc length, L_j. Substituting the appropriate values from (11) and simplifying then gives

$$\mathcal{E}_j = \frac{1}{3L_j}\mathbf{p}_j^T\mathbf{A}\mathbf{p}_j \tag{18}$$

where the vector, $\mathbf{p}_j$, and matrix, $\mathbf{A}$, are defined as

$$\mathbf{p}_j = \begin{bmatrix} \psi_j \\ \phi_j \\ \psi_{j+1} \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix} \tag{19}$$

The matrix, $\mathbf{A}$, in (19) is, as expected, positive semi-definite and has eigenvalues and corresponding (un-normalised) eigenvectors

$$\lambda_1 = 0,\ \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \quad \lambda_2 = 6,\ \mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}; \quad \lambda_3 = 24,\ \mathbf{u}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}. \tag{20}$$

Thus an alternative to expression (18) for the energy is

$$\mathcal{E}_j = \frac{(\psi_j - \psi_{j+1})^2 + \frac{4}{3}(\psi_j - 2\phi_j + \psi_{j+1})^2}{L_j} \tag{21}$$

and this form is preferable for computational purposes because it is faster to evaluate, because it is usually more accurate, and because it is guaranteed to be non-negative even in the presence of rounding errors.

3.3 Rotational Invariance

The structure of the eigensystem (20) also shows, again as expected, that the energy, $\mathcal{E}_j$, is zero if and only if $\psi_j = \phi_j = \psi_{j+1}$, which corresponds to a straight line segment. Moreover, it also *illustrates* the rotational invariance of the spiral spline since the addition of an angle α to each of $\psi_j, \phi_j, \psi_{j+1}$ clearly leaves the energy unchanged because $\mathbf{A}\mathbf{u}_1 = \mathbf{0}$. To establish this invariance property formally, it is sufficient to show that if the axes are rotated through an angle α then the solutions for the inclinations $\psi_j, \phi_j, \psi_{j+1}$ are incremented by α. Now rotation of the axes by α causes the chord angles to be incremented by α and it is easy to see that if $\{\psi_j^*\}_1^n, \{\phi_j^*\}_1^{n-1}, \{L_j^*\}_1^{n-1}$, solves (12), (13), (14), then $\{\psi_j^* + \alpha\}_1^n, \{\phi_j^* + \alpha\}_1^{n-1}, \{L_j^*\}_1^{n-1}$, solves (12) as well as equations (13) and (14) if the chord angles $\{\theta_j\}_1^{n-1}$ are replaced by $\{\theta_j^* = \theta_j + \alpha\}_1^{n-1}$.

4 Initial Approximations

Existence of a solution to the nonlinear equations (12), (13), (14), for an arbitrary configuration of data points is still an open question. Stoer[18] has shown that countably many spiral curves can be found if there are only two data points and the inclination and curvature are prescribed at one end. Therefore, it seems likely that the equations (12), (13), (14), have at least one solution, and the present author has never failed to calculate at least one spiral spline for any data set to date, which includes some awkward cases for which it is known that no finite length, stable equilibrium, least energy spline exists (some of these examples are included in §5). In practice the real difficulty to overcome is that of choosing the most appropriate spline from the many possibilities.

Clearly, the problem of choosing initial approximations is important, irrespective of the particular method used to solve the system of nonlinear equations, since it is this choice which will influence most strongly the particular solution (if any) obtained. Therefore, this aspect is considered next and it will be seen that this problem, in addition to being interesting in its own right, also leads very naturally to a simple and effective iterative scheme for solving equations (12), (13), (14).

Initial approximations are required for the arc lengths L_j, $j = 1, 2, \ldots, n-1$ and for the angles $\{\phi_j\}_{j=1}^{n-1}$, and $\{\psi_j\}_{j=2}^{n-1}$. If, in addition, the spline is not clamped at each end then one or both of ψ_1 and ψ_n must also be assigned initial values. Because equations (12) are linear in the variables $\{\psi_j\}_1^n$, it is sufficient to determine initial approximations for L_j and ϕ_j, $j = 1, 2, \ldots, n-1$, and then solve (12).

The chord length, $\hat{L}_j$, is used to estimate the arc length, L_j, since it gives a rotationally invariant estimate. The chord length also has the property of *consistently* under-estimating the arc length, which is advantageous because it is only the *relative* sizes of the arc lengths that are important to equation (12). It is also possible to use the chord angle, θ_j, as an initial approximation to ϕ_j but this can give very poor approximations in some cases since no account is taken of the end conditions. As a simple illustration, consider the two points (0,0), (1,1) for which it is required to fit a spiral spline with end inclinations $\psi_1 = 0°, \psi_2 = 0°$. For this case, the chord angle is $\theta_1 = 45°$, and this would be the correct value for ϕ_1 if natural end conditions were in force. However, with the given inclinations the true value should be close to 67.1°.

In order to derive rotationally invariant initial approximations to $\{\phi_j\}_1^{n-1}$ in a way which does take account of end conditions the equation (13) is approximated by

$$\int_0^1 (P_j(t) - \theta_j)\, dt \approx 0,$$

on the assumption that $|P_j(t) - \theta_j|$ is small for $0 \le t \le 1$. This is easily integrated after substituting for $P_j(t)$ from (11) to give the approximation:

$$\phi_j \approx \hat{\phi}_j = \frac{6\theta_j - \psi_j - \psi_{j+1}}{4}. \tag{22}$$

Substituting the approximations (15), (22) for L_j and ϕ_j in the continuity of curvature equations (12) and then rearranging gives the tridiagonal system of linear equations:

$$\begin{aligned} \hat{L}_j\psi_{j-1} + 2(\hat{L}_{j-1} + \hat{L}_j)\psi_j + \hat{L}_{j-1}\psi_{j+1} \approx\ & 3(\hat{L}_j\theta_{j-1} + \hat{L}_{j-1}\theta_j), \\ & j = 2, 3, \ldots, n-1. \end{aligned} \tag{23}$$

which, together with the known initial inclinations ψ_1, ψ_n, is easily solved in $O(n)$ arithmetic operations to give initial approximations $\{\hat{\psi}_j\}$ for the unknowns $\{\psi_j\}$. Initial approximations for $\{\phi_j\}$ can then be recovered from (22).

This procedure requires slight modification for different end conditions. For example, the natural end condition equations (17) give, in addition to equations (23), after substituting from (22) for the relevant value of ϕ_j:

$$\begin{array}{rcrcl} 2\psi_1 & + & \psi_2 & \approx & 3\theta_1 \\ \psi_{n-1} & + & 2\psi_n & \approx & 3\theta_{n-1} \end{array}$$

and this is still a tridiagonal system of linear equations when combined with (23). Clearly, other types of end condition can be accommodated in a similar way. Note also that the resulting linear system (23) is strictly diagonally dominant and a unique solution is therefore guaranteed.

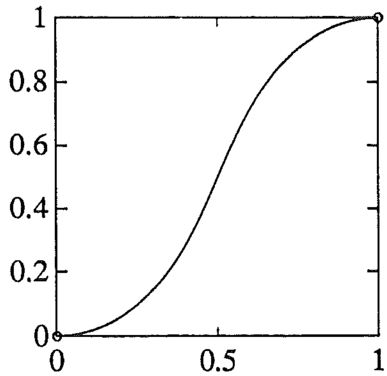

Figure 1
Clamped spline on 2 points
$\mathcal{E} = 4.863459$; $L = 1.503891$

This technique produced very good initial approximations for nonlinear spiral splines and was used for determining the initial approximations in all the numerical examples presented in §5. It is also recommended for finding initial approximations to the true least energy nonlinear spline in the method of Edwards[4]. For the simple example above, with only 2 data points, it produces the initial approximation $\hat{\phi}_1 = \frac{1}{4}(6 \times 45° - 0° - 0°) = 67.5°$, which is much closer to the true value of 67.097 (to 5 significant figures) computed by the method of §5. The resulting curve is displayed in Figure 1.

Finally it is noted that the recommended procedure for determining initial approximations is rotationally invariant. If the original axes are rotated through an angle α then the new chord angles become $\{\theta_j + \alpha\}$. Thus the new values computed by solving equations (23) will add α to each original $\hat{\psi}_j$, which in turn will cause α to be added to the original estimates $\hat{\phi}_j$ in equation (22).

5 Constructing the Spiral Spline

In principle any iterative method for solving nonlinear equations could be applied to the nonlinear system (12), (13), (14), using the initial approximations obtained by the technique described in the previous section. Usually a modified Newton or quasi-Newton method which takes account of the banded structure of the Jacobian matrix would be recommended. However, there are still some difficulties to overcome. The chord angles , $\{\theta_j\}_1^{n-1}$ are used in (23) to determine the initial approximations $\hat{\psi}_j$, $j = 2, 3, \ldots, n-1$, but what are the chord angles? The following example illustrates well the difficulty concerned. Suppose there are three data points $(0,0)$, $(1, 1.35)$, $(2,0)$, and it is required to fit a smooth curve passing through these points, in the order given, but with the end points clamped to have zero slope. Two possible curves satisfying these requirements are displayed in Figures 2 and 3.

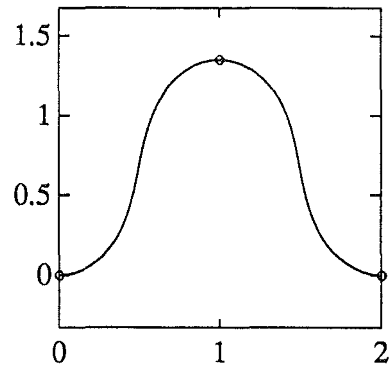

Figure 2
Clamped spline on 3 points
$\mathcal{E} = 11.21629$; $L = 3.664278$

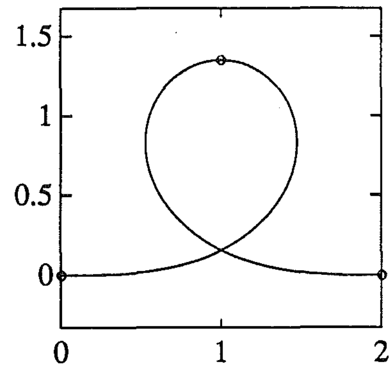

Figure 3
Clamped spline on 3 points
$\mathcal{E} = 9.842400$; $L = 5.347986$

In each case, the nonlinear equations (12), (13), (14) are satisfied to high accuracy by the displayed curves. Although neither curve is a true least energy curve, each curve can be seen to satisfy condition (8) which is necessary for the existence of a finite length least energy curve. Both curves also have positive curvature at each end point although, in the case of Figure 3, the curvature is very slight. The curve in Figure 2 was computed using end conditions: $\psi_1 = 0°, \psi_3 = 0°$ with chord angles $\theta_1 = 53.47°, \theta_2 = -53.47°$, leading to initial approximations $\hat{\phi}_1 = 80.21°, \hat{\phi}_2 = -80.21°$, and final values $\phi_1 = 79.53°, \phi_2 = -79.53°$. The curve in Figure 3 was computed using end conditions: $\psi_1 = 0°, \psi_3 = 360°$ with chord angles $\theta_1 = 53.47°$, $\theta_2 = 360° - 53.47 = 306.53°$, leading to initial approximations $\hat{\phi}_1 = 35.21°, \hat{\phi}_2 = 324.79°$, and final values $\phi_1 = 45.00°, \phi_2 = 315.00°$. It is difficult to make a choice, out of context, as to which is the better of these two curves because the topology is different in each case. Both curves are spiral splines but the curve in Figure 3 has less energy but longer length than that in Figure 2. The curve in Figure 3 may not be acceptable, however, because it crosses itself. If the orientation is also specified at each knot then the choice is easily resolved in the case above but for large data sets this may not be a practical proposition.

The difficulty is compounded if the configuration of points depicted in Figures 2 and 3 lies well away from the endpoints in a much larger set of data because then the interior inclinations would not be known a priori. Changing these inclinations by a few degrees may well affect the decision on which fitted curve is preferred. In practice this difficulty needs to be handled by the nonlinear equation solver which must be prevented from flipping from one configuration to another between iterations. This could be done by adding linear inequality constraints in the variables ψ_j, ϕ_j, to the nonlinear system of equations but a different approach is preferred.

The chord lengths $\hat{L}_j$, $j = 1, 2, \ldots, n-1$, are calculated first and used as initial approximations for the arc lengths, $L_j^{(0)} = \hat{L}_j$. The chord angles θ_j, $j = 1, 2, \ldots, n-1$, are also calculated and adjusted modulo 360° to fix the topology. Then initial values, $\psi_j^{(0)} = \hat{\psi}_j$, are calculated using the calculated chord lengths and chord angles as described in §4. The following simple iterative algorithm can then be applied.

Algorithm 5.1

For $k = 0, 1, 2, \ldots$

1. For $j = 1, 2, \ldots, n-1$:

   (a) Keeping the values of $\{\psi_j\}$ fixed at $\{\psi_j^{(k)}\}$ solve the single nonlinear equation in the single variable ϕ_j:

   $$\int_0^1 \sin\left(P_j^{(k)}(t) - \theta_j\right) dt = 0 \tag{24}$$

   where $P_j^{(k)}(t) = \psi_j^{(k)} + (\psi_{j+1}^{(k)} - \psi_j^{(k)})t + 2(\psi_j^{(k)} - 2\phi_j + \psi_{j+1}^{(k)})t(t-1)$.
   Denote the solution by $\phi_j^{(k+1)}$.

   (b) Re-estimate the arc length, L_j, using the latest estimates $\{\psi_j^{(k)}\}$, $\{\phi_j^{(k+1)}\}$ in the equation:

   $$L_j^{(k+1)} = \hat{L}_j / \int_0^1 \cos\left(\overline{P}_j^{(k)}(t) - \theta_j\right) dt \tag{25}$$

   where $\overline{P}_j^{(k)}(t) = \psi_j^{(k)} + (\psi_{j+1}^{(k)} - \psi_j^{(k)})t + 2(\psi_j^{(k)} - 2\phi_j^{(k+1)} + \psi_{j+1}^{(k)})t(t-1)$,

2. Obtain new estimates $\{\psi_j^{(k+1)}\}$ by solving the tridiagonal system of linear equations (12) using the most recent values $\{\phi_j = \phi_j^{(k+1)}\}$, $\{L_j = L_j^{(k+1)}\}$.

3. Terminate if $\max_j |\psi_j^{(k+1)} - \psi_j^{(k)}| \le \varepsilon$, (a preset tolerance) or if the iteration appears to be diverging. (The results in this paper were obtained using $\varepsilon = 10^{-6}$.)

The only real difficulty lies in step 1. The integrals (24), (25) can be expressed in terms of the Fresnel integrals:

$$\begin{aligned} C(\alpha) &= \int_0^{\alpha} \cos\left(\frac{\pi}{2}t^2\right) dt \\ S(\alpha) &= \int_0^{\alpha} \sin\left(\frac{\pi}{2}t^2\right) dt \end{aligned} \tag{26}$$

for which there are standard approximations (see for example [1, 3]). For the results in this paper the Harwell subroutine library routine FC10AD was used for evaluation of the integrals in (24), (25) and subroutine NB02AD to solve the single nonlinear equation (24).

An advantage of this approach is that it enables control over the topology of the spline to be maintained easily. If neighbouring chords have inclinations satisfying

$$|\theta_{j+1} - \theta_j| < 90°, \ j = 1, 2, \ldots, n-2, \tag{27}$$

then no difficulties have been experienced and the algorithm converges reliably and quickly in 2–10 iterations for data sets varying in size from $n = 2$ to $n = 1000$. When condition (27) is not satisfied there is usually some difficulty in deciding the appropriate topology and this seems to manifest itself in the algorithm by a slower convergence rate. As an illustration, consider the data of Reinsch[17], which comprises the 8 ordered points (-184, 44), (-76, 100), (-45, 121), (0, 120), (29, 112), (55, 14), (104, 100), (135, 120). Here the chord angles θ_5, θ_6, violate condition (27) and some difficulty can be anticipated. Two possible curves were, however, successfully computed using Algorithm 5.1 by making appropriate choices for the initial chord angles. These curves are displayed in Figures 4 and 5.

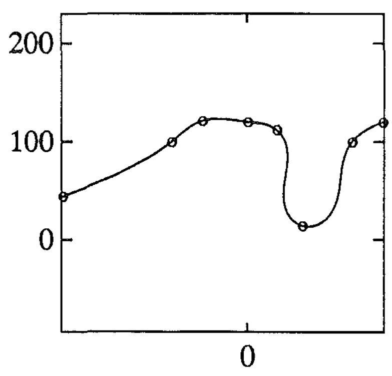

Figure 4. Reinsch[17] data
Natural spline on 8 points
$\mathcal{E} = 0.186045$; $L = 492.9196$

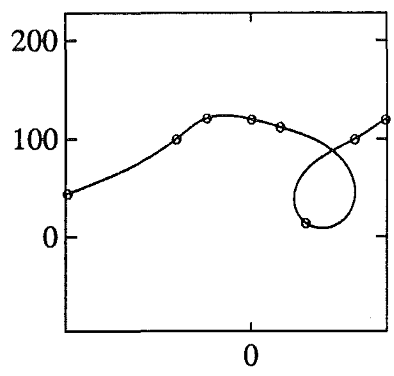

Figure 5. Reinsch[17] data
Natural spline on 8 points
$\mathcal{E} = 0.154888$; $L = 572.7310$

The algorithm was also applied to the data of Birkoff, Burchard and Thomas[2], which comprises the 4 ordered points (1, 0), (2, 0), (0, 2), (0, 1). They claim that no stable equilibrium, finite length, *least* energy spline exists for this configuration but the algorithm of this section was able to calculate a finite length *low* energy natural spiral spline that has the required smoothness properties. The resulting curve is displayed in Figure 6. This data also violates (27) and the fitted curve violates condition (8), showing that it cannot be a true least energy curve. The curve has an arc of a circle on the second span.

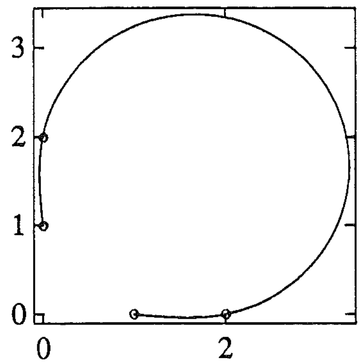

Figure 6. BBT[2] data
Natural spline on 4 points
$\mathcal{E}=2.771497$; $L=9.350514$

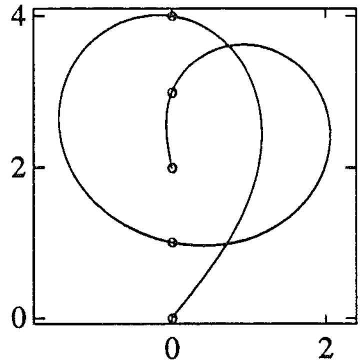

Figure 7. Collinear data
Natural spline on 5 points
$\mathcal{E}=7.397542$; $L=16.41264$

As a final (and pathological) example the collinear set of points $(0,0)$, $(0,4)$, $(0,1)$, $(0,3)$, $(0,2)$, in the order given was fitted by a natural spiral spline. Setting the chord angles to be $\{\theta\}_1^4 = \{90°, 270°, 450°, 630°\}$ the curve in Figure 7 was obtained. Its reflection in the y-axis could have been obtained if chord angles $\{\theta\}_1^4 = \{90°, -90°, -270°, -450°\}$ had been used instead. Of course, it is unlikely that such a configuration of points would arise in any practical application but the example does indicate the robustness of the given approach.

6 Discussion

The convergence properties of Algorithm 5.1 have not yet been studied extensively but numerical experience so far indicates that for sensible end conditions and appropriately calculated chord angles convergence always occurs. It would, however, be a simple matter to modify the algorithm to guarantee convergence to a 'solution' which at least minimizes some norm of the residuals of the nonlinear equations (12), (13), (14), but there are advantages in studying the convergence properties of the unmodified algorithm. The following result, although rather obvious, is stated as a theorem to emphasise its importance.

Theorem 6.1 *If the sequences* $\{\psi_i^{(k)}\}_{i=1}^n$, $\{\phi_j^{(k)}\}_{j=1}^{n-1}$, $\{L_j^{(k)}\}_{j=1}^{n-1}$, $k=1,2,3,\ldots$, *generated by Algorithm 5.1 converge as k tends to infinity, with limit points denoted (in vector notation) by* $\boldsymbol{\psi}^\infty$, $\boldsymbol{\phi}^\infty$, $\boldsymbol{L}^\infty$, *respectively, then the nonlinear equations (12), (13), (14), are satisfied when* $\psi_i=\psi_i^\infty$, $i=1,2,\ldots,n$, $\phi_j=\phi_j^\infty$, $L_j=L_j^\infty$, $j=1,2,\ldots,n-1$.

The importance of this result is that it can be used to establish conditions which guarantee the existence of a spiral spline if conditions can be found for which Algorithm 5.1 converges. The results presented in this paper indicate that convergence is usually achieved and that

Algorithm 5.1 is particularly effective when condition (27) holds for all the data points. These topics of existence and convergence are currently under further investigation.

A particularly attractive feature of Algorithm 5.1 is its simplicity; the original pilot code was written in an evening. However, some improvements to efficiency could still be made. In particular the evaluation of the Fresnel integrals (26) via the Harwell subroutine FC10AD is somewhat slow and also gives poor relative accuracy for small values of α. Improved procedures could easily be incorporated.

Acknowledgement

It is a pleasure to acknowledge the very many people who have at various stages influenced the work described here. In particular I am most grateful for the advice, and helpful comments given by my colleague Dr D.J.N. Wall. I am also indebted to John Edwards, for providing the initial impetus which started this project and without whose encouragement this paper would never have been written.

References

[1] M. Abromowitz, I.A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Na. Bur. Stand., Washington D.C., 1972.

[2] G. Birkhoff, H. Burchard, D. Thomas, *Nonlinear interpolation by splines, pseudosplines, and elastica*, Research Publication GMR-468, General Motors Research Laboratories, Warren, Michigan, 1965.

[3] J. Boersma, *Computation of Fresnel integrals*, Math. Comp. **14**, 1960, p380.

[4] J. Edwards, *Exact equations of the spline*, A.C.M. Trans. Math. Softw. (to appear)

[5] M.J. Forray, *Variational Calculus in Science and Engineering*, McGraw-Hill, New York, 1968.

[6] J.M. Glass, *Smooth-curve interpolation: A generalised spline-fit procedure*, BIT **6**, 1966, pp277-293.

[7] B.K.P. Horn, *The curve of least energy*, A.C.M. Trans. Math. Softw. **9**, 1983, pp441-460.

[8] M. Kallay, *Plane curves of minimal energy*, A.C.M. Trans. Math. Softw. **12**, 1986, pp219-222.

[9] F.M. Larkin, *An interpolation procedure based on fitting elasticas to given data points*, Culham Operating System Note 5/66, Theory Division, Culham Laboratory, Abingdon, U.K., 1966, Re-issued as CLM-R284, 1988.

[10] J.D. Lawrence, *A catalog of plane curves*, Dover Publications, New York, 1972.

[11] E.H. Lee, G.E. Forsythe, *Variational study of nonlinear spline curves*, SIAM Review **15**, 1973, pp120-133.

[12] M.A. Malcolm, *On the computation of nonlinear spline functions*, SIAM J. Numer. Anal. **14**, 1977, pp254-293.

[13] E. Mehlum, *Nonlinear splines*, in "Computer aided geometric design", R.E. Barnhill, R.F. Riesenfield (Eds), Academic Press (New York), 1974, pp173-207.

[14] E. Mehlum, *Curve and surface fitting based on variational criteria for smoothness*, Technical Report, Central institute for industrial research, Oslo, 1969 (83 pages).

[15] E. Mehlum, *Appell and the apple (nonlinear splines in space)*, Technical Report No. 1676, Central institute for industrial research, Oslo, 1981 (28 pages).

[16] T.K. Pal, A.W. Nutbourne, *Two-dimensional curve synthesis using linear curvature elements*, Computer Aided Design **9**, 1977, pp121-134.

[17] K-D Reinsch, *Numerische Berechnung von Biegelinien in der Ebene*, Technical Report No. M8108, University of Munich, 1981.

[18] J. Stoer, *Curve fitting with clothoidal splines*, Journal of Research of the National Bureau of Standards **87**, 1982, pp317-346.

[19] C.H. Woodford, *Smooth curve interpolation*, BIT **9**, 1969, pp69-77.

INTERNET: idc@math.canterbury.ac.nz

Department of Mathematics
University of Canterbury
Christchurch 1
NEW ZEALAND