January 1985

Report No. STAN-CS-85-1047

Smooth, Easy to Compute Interpolating Splines

by

John D. Hobby

Department of Computer Science

Stanford University
Stanford, CA 94305

Smooth, Easy to Compute Interpolating Splines

John D. Hobby
Computer Science Dept.
Stanford University
Stanford, California 94305

Abstract: We present a system of interpolating splines with first and approximate second order geometric continuity. The curves are easily computed in linear time by solving a system of linear equations without the need to resort to any kind of successive approximation scheme. Emphasis is placed on the need to find aesthetically pleasing curves in a wide range of circumstances; favorable results are obtained even when the knots are very unequally spaced or widely separated. The curves are invariant under scaling, rotation, and reflection, and the effects of a local change fall off exponentially as one moves away from the disturbed knot.

Approximate second order continuity is achieved by using a linear "mock curvature" function in place of the actual endpoint curvature for each spline segment and choosing tangent directions at knots so as to equalize these. This avoids extraneous solutions and other forms of undesirable behavior without seriously compromising the quality of the results.

The actual spline segments can come from any family of curves whose endpoint curvatures can be suitably approximated, but we propose a specific family of parametric cubics. There is freedom to allow tangent directions and "tension" parameters to be specified at knots, and special "curl" parameters may be given for additional control near the endpoints of open curves.

This research was supported in part by the National Science Foundation under grants IST-820-1926 and MCS-83-00984 and by the Systems Development Foundation.

1. Introduction

The problem of fitting a smooth curve through a set of points on the plane has many important applications in computer graphics, computer aided design, and typesetting. Often there is no pre-existing curve to approximate except possibly a freehand drawing, and the only requirement is to find an aesthetically pleasing curve that the computer can easily manipulate. For some interactive applications the curves can be controlled by manipulating points that do not lie on the curve, but many applications require the control points to lie on the curve. For example, the control points may be obtained by digitizing key points on a drawing, or there may be a priori knowledge that the curve must pass through certain points.

Suppose the curve must pass through points $z_0, z_1, \ldots, z_n$; either z_0 and z_n are to be the endpoints of the curve, or $z_0 = z_n$ and the curve is to be a closed loop. Optionally, there may be direction vectors w_i specifying the curve slope at some z_i. For example, some of the z_i may have been selected as vertical extrema so that the curve must pass through them horizontally. It is desirable for the curve to be invariant under scaling, rotation, and reflection in the sense that if T is such a transformation then applying T to the computed curve should yield the same result as computing a new curve through Tz_i for $0 \le i \le n$ with direction vectors Tw_i.

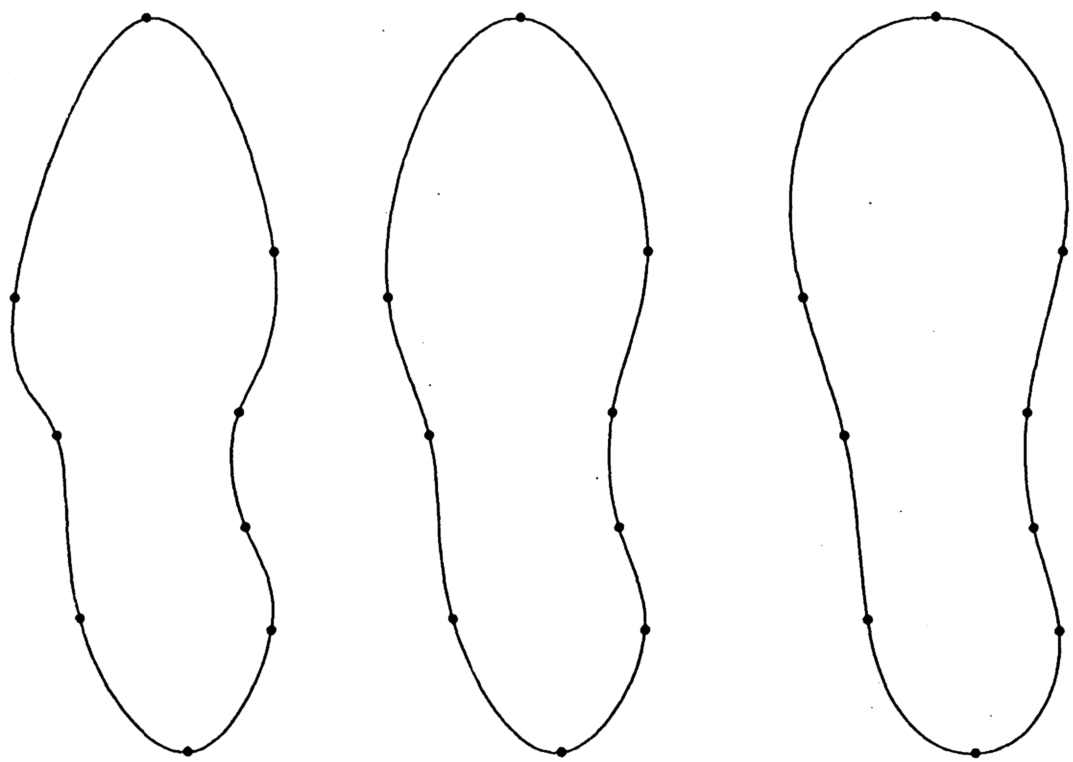
The curve should have at least approximate continuity of slope and curvature where no directions are given, and it would also be desirable to have some notion of extensibility and locality. A system of splines is extensible if the curve generated from knots $z_0, z_1, \ldots, z_n$ is identical to that generated from knots $z'_0, z'_1, \ldots, z'_{n+1}$ where $z'_i = z_i$ for $i < k$, $z'_i = z_{i-1}$ for $i > k$, and z'_k is on the curve segment joining z_{k-1} and z_k. In other words, adding a new knot already on the curve must not change it. In practice it is extremely difficult to achieve exact extensibility. The only well-known extensible spline family is the "curve of least energy" that minimizes the integral of squared curvature with respect to arc length [3, 6], but this curve is difficult to work with. It is interesting to note that when the knots are nearly collinear, the curve of least energy approaches the simple non-parametric cubic spline passing through the given knots with continuous second derivative. The splines that we deal with here will share this property.

1. The effect of changing w_0 while preserving exact locality

The concept of locality is that if one of the knots or direction vectors is perturbed, the changes should be confined to a few surrounding spline segments. Here we will settle for a kind of exponential decline in influence rather than a strict limitation to a few surrounding knots. As the example of Figure 1 shows, it is difficult to have both exact locality and continuity of curvature even for nearly straight curves. If w_0 is in the direction of $z_1 - z_0$ then the desired curve is obviously a straight line, yet there is no way cubic curve can join a straight line with continuous curvature.

B-splines have locality and continuous curvature, but of course they do not interpolate. The interpolating splines analogous to cubic B-splines, sometimes called "natural cubic splines," do not have locality but can easily be computed by solving linear equations. If no directions are given, there is a unique piecewise parametric cubic, closed curve that is C^2 continuous with respect to the parameter and passes through n given points in order. Such a curve can be uniquely represented as a cubic B-spline, and its control points are linear combinations of $z_0, z_1, \ldots z_n$.

As shown in [2], natural cubic splines do not perform well for unequally spaced knots because the spacing of parameter values at knots does not reflect the spacing of the knots. Better results can be obtained by setting the parameter at each knot z_i to a value t_i where $t_j - t_{j-1} = \|z_j - z_{j-1}\|$ for $1 \le j \le n$, and requiring second order continuity with respect to the parameter as shown in Figure 2b. This chordal paramcterization improves on the uniform parameterization of Figure 2a, but. the splines that we shall develop still have more gentle curvature in this case as shown in Figure 2c.



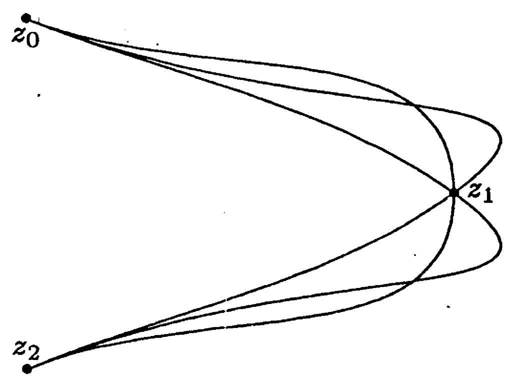
2a. Natural cubics 2b. Cubics with chordal parameterization 2c. Cubics with mock curvature constraints

Figure 2 points out the difference between geometric and parametric continuity. Requiring first and second order continuity with respect to the parameter uses up four degrees of freedom per knot, enough co completely determine a parametric cubic spline. One of these dcgrccs of freedom can be reclaimed and put to better use by altering the parameter spacing as shown in Figure 2b, but another dcgreo of freedom can be made available by requiring only continuity of slope and curvature.

In [1], Barsky and Beatty show how two extra degrees of freedom can be obtained for B-splines by requiring only geometric continuity. We need to obtain similar degrees of freedom for interpolating splines, but rather than trying to adapt the bias and tension parameters of [1], we shall first concentrate on finding good defaults to work from. The new parameters will be of little value if it is difficult to set them so as to obtain reasonable results. Any system of cubic interpolating splines must implicitly provide some mechanism for fixing the two parameters and it'is not at all clear that this best done by requiring any form of parametric C^2 continuity.

In [5], J. R. Manning takes an interesting approach to this problem. He defines a specific family of curves so that so that there is a unique one for each pair of initial and final points and directions. He then selects spline directions at each knot so as to achieve geometric continuity. Although Manning does not does not deal with the possibility of some directions being specified in advance, his approach provides a certain degree of locality in that effects of local perturbations do not propagate past knots where a direction is given.

With Manning's approach, both degrees of freedom are available to control the shape of the curve, and defaults can be selected so as to obtain the most pleasing curves. Section 2 explains how to select the defaults by choosing two functions and using them to determine the magnitudes of the velocities at each knot in such a way as to guarantee that the curves generated will be independent of scaling, rotation, and reflection. We can then provide two "tension" parameters for each knot by simply dividing them into these functions. Essentially the same approach would work for other kinds of curves, although there may be more parameters to choose. We select parametric cubics here because they are essentially the simplest curves that can pass through two arbitrary points in two arbitrary directions. Conic sections do not suffice because of their inability to handle points of inflection.



3. Three splines of the type proposed in [5].

One apparent disadvantage to this approach is the difficulty in solving for the directions that provide continuity of curvature. Manning proposes an iterative approximation scheme that seems to work well in practice, but he admits that there is not always a unique solution and there is no guarantee that the iteration always converges to the desired solution. Cubic splines often have very low curvature at their endpoints when they have very sharp bends internally, and this can introduce extraneous solutions as shown in Figure 3. The three curves shown are all curvature continuous open curves that have given directions at z_0 and z_2, but regardless of the initial conditions, Manning's iteration always converges to one of the asymmetrical ones with sharp bends. If z_0 is raised and z_2 lowered until the angle $z_0 z_1 z_2$ is about 122°, the asymmetrical solutions merge with the symmetrical ones and the rate of convergence for Manning's iteration approaches zero.

While these kinds of problems do not seem to occur when the angles involved arc not so large, much additional testing would be necessary in order to verify this. In section 3, we show how all such problems can be avoided by setting up a system of linear equations that are easy to solve and guarantee approximate continuity of curvature. We derive the specific equations appropriate for the family of curves discussed in Section 2, but similar equations could be derived for many different classes of curves.

2. The Magnitudes of the Velocity Vectors

The subproblem to be solved in this section can be stated as follows: Given two points z_i and z_{i+1}, and two unit vectors w_i and w_{i+1}, find an aesthetically pleasing parametric cubic $z(t)$ so that $z(0) = z_i$, $z(1) = z_{i+1}$, $z'(0) = \alpha w_i$, and $z'(1) = \beta w_{i+1}$, where α and β are positive real numbers and $z'(t)$ is the componentwise derivative $(x'(t), y'(t))$ of $z(t) = (x(t), y(t))$. We also wish to introduce two "tension" parameters τ_i and $\bar{\tau}_{i+1}$ such that pleasing curves will be obtained when $\tau_i = \bar{\tau}_{i+1} = 1$, and as the tensions approach ∞, the curves will approach the line segment joining z_i to z_{i+1}.

In order to guarantee that the results are independent of translation, rotation, and scaling, we shall begin by finding a function $\hat{z}(t)$ such that

$$\hat{z}(0) = (0,0), \quad \hat{z}(1) = (1,0),$$
$$\hat{z}'(0) = \frac{1}{\tau_i}\rho(\theta, \phi) \cdot (\cos\theta, \sin\theta), \quad \text{and} \quad \hat{z}'(1) = \frac{1}{\bar{\tau}_{i+1}}\sigma(\theta, \phi) \cdot (\cos\phi, -\sin\phi) \tag{1}$$

where ρ and σ are positive real functions to be determined later, $\theta = \arg w_i - \arg(z_{i+1} - z_i)$, and $\phi = \arg(z_{i+1} - z_i) - \arg w_{i+1}$. (Here $\arg(x, y)$ is the angle ω such that (x, y) is a positive multiple of $(\cos\omega, \sin\omega)$.) We then set

$$z(t) = z_i + \begin{pmatrix} x_1 - x_0 & y_0 - y_1 \\ y_1 - y_0 & x_1 - x_0 \end{pmatrix} \hat{z}(t). \tag{2}$$

It is not hard to see that the parametric cubic satisfying (1) has Bézier control points $(0,0)$, $(\rho/3\tau_i) \cdot (\cos\theta, \sin\theta)$, $(1 - (\sigma/3\bar{\tau}_{i+1})\cos\phi, (\sigma/3\bar{\tau}_{i+1})\sin\phi)$, and $(1,0)$, so that

$$\hat{z}(t) = \frac{\rho}{\tau_i}t(1-t)^2(\cos\theta, \sin\theta) + t^2(1-t)\left(3 - \frac{\sigma\cos\phi}{\bar{\tau}_{i+1}}, \frac{\sigma\sin\phi}{\bar{\tau}_{i+1}}\right) + t^3(1,0). \tag{3}$$

It only remains to choose positive functions $\rho(\theta, \phi)$ and $\sigma(\theta, \phi)$ so that $\rho(\theta, \phi) = \sigma(\phi, \theta) = \rho(-\theta, -\phi)$.

In [5] Manning chooses

$$\rho(\theta, \phi) = \frac{2}{1 + (1-c)\cos\theta + c\cos\phi} \quad \text{and} \quad \sigma(\theta, \phi) = \frac{2}{1 + c\cos\phi + (1-c)\cos\theta} \tag{4}$$

and then empirically selects $c = 2/3$ to obtain the most pleasing family of curves. Here we shall attempt to do a systematic analysis of the vast range of possible functions to determine whether slightly more complicated functions will yield better results. These functions will have to be evaluated only once for each segment of the spline curve, and they have a strong influence on the final shape.

2.1. Mathematical measures of smoothness

One common way of evaluating the smoothness of curves is to integrate the square of curvature with respect to arc length. For $\hat{z} = (\hat{x}, \hat{y})$

$$\int k^2\, ds = \int_0^1 \frac{(\hat{y}''\hat{x}' - \hat{x}''\hat{y}')^2}{(\hat{x}'^2 + \hat{y}'^2)^{5/2}} dt. \tag{5}$$

This can be simplified somewhat but it still proves to be intractable analytically. This is not surprising considering the complex behavior of numerical solutions.

Equation (5) is exactly the energy function that the curve of **least** energy minimizes, but if we restrict $\hat{z}$ to be the cubic spline (3), we can investigate the functions ρ and σ that minimize (5). Actually we should consider the smallest local minimum since (5) approaches 0 as ρ and σ approach ∞ for fixed θ and ϕ.

Unfortunately, numerical integration of $k^2\, ds$ proves to be slow and imprecise, and it would have to be repeated a large number of times in order to get a good idea what the functions $\rho(\theta, \phi)$ and $\sigma(\theta, \phi)$ should be. Instead we shall introduce two other measures of smoothness that behave similarly: .

$$\max_{0 \le t \le 1} |k(t)| \tag{6}$$

and

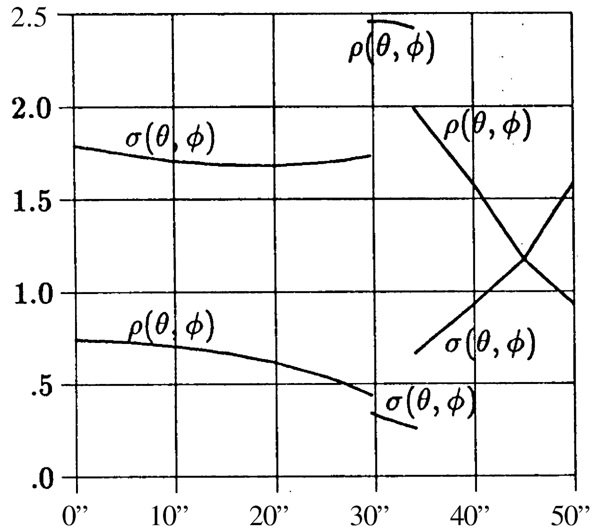
$$\max_{0 \le t \le 1} \left| \frac{dk}{ds} \right|. \tag{7}$$

Using (6) to measure smoothness corresponds to taking the oo-norm of curvature instead of the 2-norm; using (7) gives roughly similar results but applies a greater penalty to short periods of relatively high curvature. The functions $\rho(\theta, \phi)$ and $\sigma(\theta, \phi)$ that minimize (7) turn out to be somewhat better behaved than those that minimize (6), but the overall character is the same for both measures of smoothness.

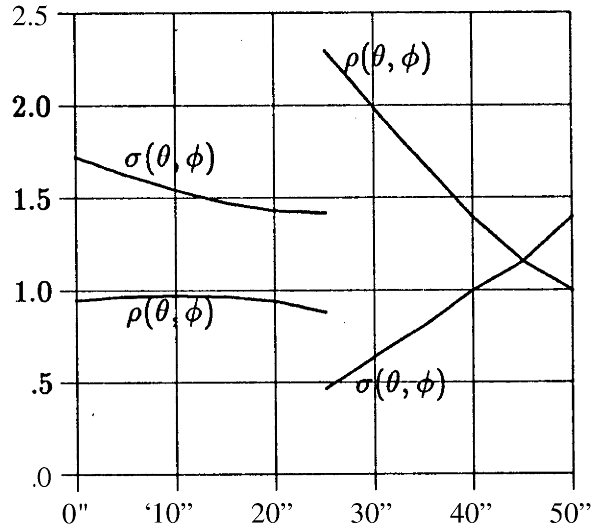
For fixed θ and ϕ, the smoothness measures can have multiple local minima and the relative smoothness at the local minima can change as θ and ϕ change. Therefore it should not be surprising that the "optimum" ρ and σ functions have large discontinuitics where they catastrophically change from one local minimum to another. When this happens there tend to be (p, a) values between the two minima that also generate relatively "smooth" curves, so it is not really necessary to use discontinuous functions for ρ and σ.

Figure 4 illustrates the most basic catastrophe. Ncsr $\theta = \phi = 45°$, the "optimum" ρ increases and the σ decreases as 0 decreases. This action tends to reduce the curvature where it is maximum near $t = 1$ without introducing other points of high curvature. When $0 \approx \theta \ll \phi$, the situation is entirely different. The high curvature near $t = 1$ is best controlled by making σ large, and increasing ρ beyond what is needed to control the curvature at $t = 0$ just makes the problem worse.

When ρ and σ arc chosen to minimize (6) as shown in Figure 4a, there arc actually two catastrophes, and the short segment between them is particularly interesting. Ordinarily, extremely small σ values lead to high curvature near $t = 1$, but at $\theta = 34.1°$, $\phi = 55.9°$, the curvature is actually minimized by choosing $\sigma = .261$. The choice of $\rho = 2.423$ here

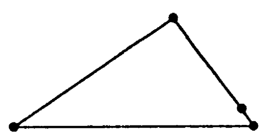


4a. The functions $\rho(\theta, \phi)$ and $\sigma(\theta, \phi)$ versus θ for ϕ = 90” $-\ \theta$, minimizing the magnitude of curvature.



4b. The functions $\rho(\theta,\phi)$ and $\sigma(\theta, \phi)$ versus θ for ϕ = 90” $-\ \theta$, minimizing the magnitude of curvature change.

is extremely critical. As shown in Figure 5, this has the effect of making the last three **Bézier** control points almost collinear, so that the endpoint curvature is not too large in spite of the low velocity.



5. The Bdzicr control polygon for (3) with θ = **34.1°**, ϕ = **55.9°**, where ρ and σ are chosen to minimize curvature.

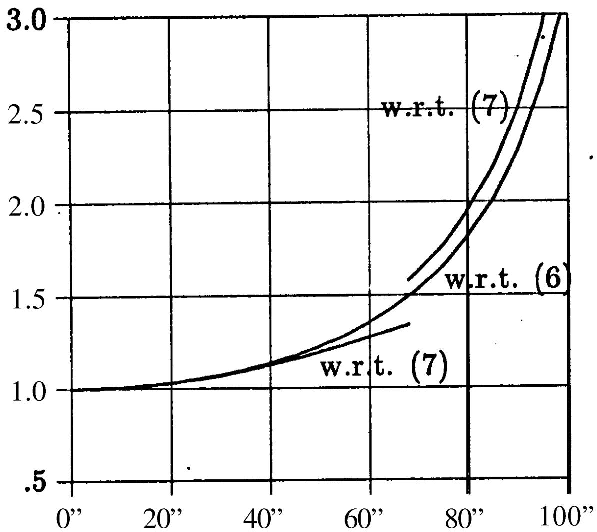
The bizarre situation shown in Figure 5 does not occur when minimizing the derivative of curvature, but there is still a catastrophe near 25”. When $\theta + \phi = 90°$ and $\theta <$ 25” the “optimum” cubic has a point of inflection, but when $\theta >$ 25” it has none.

Figure 6 shows how the “optimal” velocity parameters grow as θ and ϕ **increase.** Minimizing (7) produces a catastrophe at 68” where ρ and σ increase to about 1.58 and the cubic acquires a single point of maximum curvature at $t = .5$. For $\rho = \sigma \approx 1.34$, the maximum curvature occurs at $t \approx .08$ and $t \approx .92$. Intermediate values for ρ and σ produce rclntivcly high change in curvature near $t = 0$ and $t = 1$.

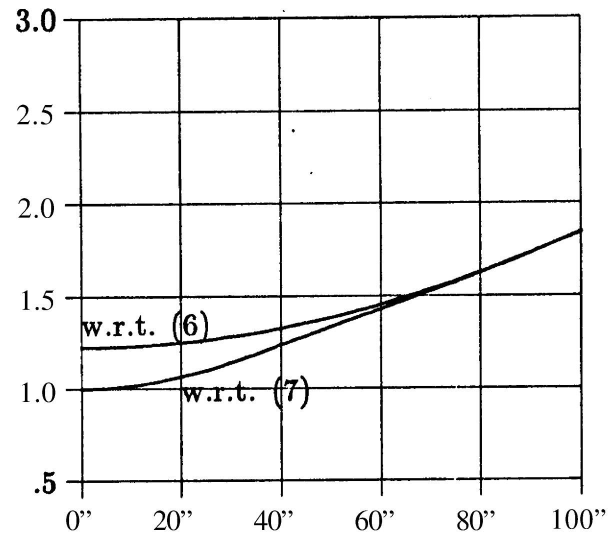
Minimizing (G) instead of (7) avoids the catastrophe at $\theta = \phi = 68°$, but then ρ and σ do not approach unique limits as (0, $\phi) \rightarrow (0,0)$. Along $\theta = -\phi$ the limit is $\sqrt{6}/2$, while ρ and σ approach 1 as $\theta \rightarrow 0$ when $\theta = \phi$. Under the approximation $k \approx \frac{d^2y}{dx^2}$, when either (5) or (7) is used as the measure of smoothness, it can be shown that the optimum curves are cubics where $\rho \cos\theta = \sigma \cos\phi = 1$. Thus, it seems reasonable to let *(p, a)* $\rightarrow (1,1)$ as $(\theta,\phi) \rightarrow (0,0)$.

2.2. Practical equations for the velocity parameters

Practical equations for ρ and σ must bc continuous and fairly easy to **evaluate.** The



6a. "Optimum" ρ and σ versus θ when $\theta = \phi$ for both smoothness measures.



6b. "Optimum" ρ and σ versus θ when $\theta = -\phi$ for both smoothness measures.

"optimum" functions illustrated in Figures 4 and 6 fail on both counts, but they still provide useful guidelines. The actual choice of functions is necessarily somewhat arbitrary, and there is a trade-off between "smoothness" and simplicity. Other properties such as approximate extensibility and predictable response to changes are also important, but empirical studies indicate that these goals also tend to be served by maximizing "smoothness".

We have already decided that $\rho(0,0) = \sigma(0,0) = 1$. Thus for small angles we can approximate the behavior of the curve of least energy and achieve very good approximate extensibility. For $\theta = \phi$, we should approximate the behavior of the functions shown in Figure 6a, but these functions increase too rapidly for large angles: they'seem to approach ∞ well before θ and ϕ reach 180". It is convenient to let

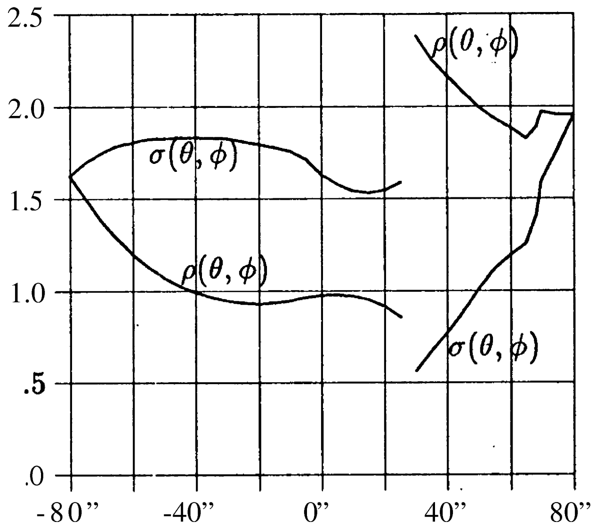
$$\rho = \sigma = \frac{2}{1 + \cos\theta} \quad \text{for } e = 4 \tag{8}$$

so as to obtain good approximations to circles.

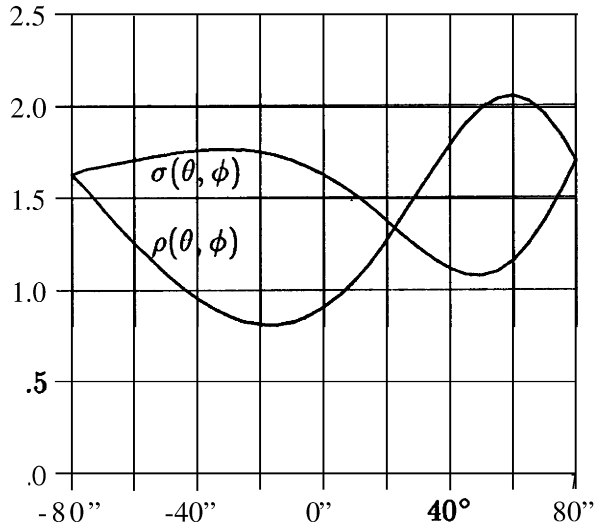
Because of the symmetry requirements $\rho(\theta, \phi) = \sigma(\phi, \theta) = \rho(-0, -\phi)$, it suffices to choose ρ and σ for $0 \leq |\theta| \leq \phi \leq 180$". Figure 7a shows the $\rho(\theta, \phi)$ and $\sigma(\theta, 4)$ that minimize (7) for $4 = 80$". Figure 7b shows practical functions ρ and σ that smooth out the catastrophes and are consistent with (8). Similar plots for smaller ϕ would have ρ and σ closer to each other and closer to' 1. (The slope discontinuities at 64.7" and 69.6° arc due to changes in the relative sizes of extrema in $\frac{dk}{ds}$ at different parts of the cubic curve.)

If complexity is of no concern, wc might want' to choose $\rho(\theta, \phi)$ and $\sigma(\theta, \phi)$ as follows for $0 < |\theta| \leq 4 < \pi$ with angles measured in radians:

$$P = f(\theta, 4) + \gamma(\phi) \cdot \sin(\psi_\phi(\theta/\phi)),$$
$$\sigma = f(\theta, 4) - \gamma(\phi) \cdot \sin(\psi_\phi(\theta/\phi)),$$
$$f(\theta, 4) = \frac{a\alpha^2 + \alpha + 2c}{\alpha + c\cos\beta + c},$$



7a. "Optimum" $\rho(\theta,\phi)$ and $\sigma(\theta,\phi)$ versus θ for ϕ = 80°.



7b. Practical functions $\rho(\theta,\phi)$ and $\sigma(\theta,\phi)$ versus θ for ϕ = 80°.

$$\alpha = (\phi-\theta)\cdot\left(\frac{\phi-\theta}{2\phi}\right)^{d}, \quad \beta = \frac{\theta+\phi}{2}\cdot\left(\frac{2\phi}{\theta+\phi}\right)^{e}$$

$$\gamma(\phi) = \frac{1.17}{\pi}\phi - .15\sin(2\phi),$$

$$\psi_\phi(x) = \pi\cdot\left(x + (x^2-1)((.32-\phi/2\pi)x + .5 - \phi/2\pi)\right). \qquad (9)$$

A least squares fit of f to $(\rho$ + a)/ 2 with ρ and σ chosen to minimize (7) yielded a = 0.2678306, c = 0.2638750, d = 1.402539, and e = 0.7539063. A possible refinement is to require $\rho \le 1.5\sin\phi/\sin(\theta+\phi)$ when $\phi > \pi$, $\theta < 0$, and $\theta \ne -\phi$ SO as to avoid any possibility of generating a curve with a cusp in it. (This only effects the above functions when $\phi > 145°$.)

It is desirable to have a simpler approximation that does not use transcendental functions other than sines and cosines of θ and ϕ. One such approximation is the following functions which were developed for the new METRFONT system [4]:

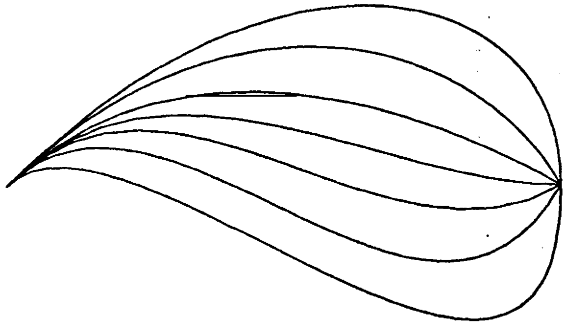
$$\rho = \frac{2+\alpha}{1+(1-c)\cos\theta + c\cos\phi},$$

$$\sigma = \frac{2-\alpha}{1+(1-c)\cos\phi + c\cos\theta}, \quad \text{where}$$

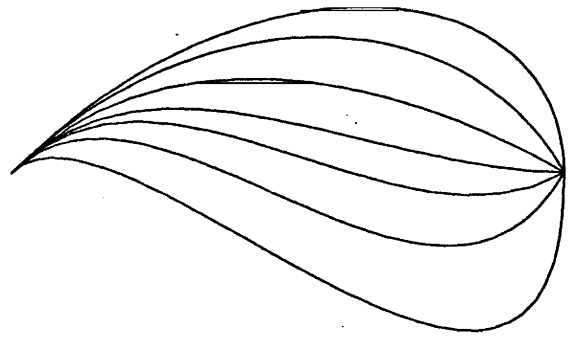
$$\alpha = a(\sin\theta - 6\sin\phi)(\sin\phi - 6\sin 0)(\cos\theta - \cos\phi). \qquad (10)$$

The constants a, 6, and c were chosen to minimize an error function based on the value of (10) for 116 different (0, ϕ) pairs. This suggested a = 1.597, 6 = .0700, and c = .370, but since empirical evidence indicated that large values for $|\rho-\sigma|$ were causing problems, METAFONT uses the slightly perturbed values a = $\sqrt{2}$, 6 = $\frac{1}{16}$, and c = $(3-\sqrt{5})/2$.

Figure 8 shows some of the curves generated by (9) and (10). They are similar for moderate angles, but the simpler equations set ρ too small and σ too large when $\phi = -90°$. Equation (10) does not perform well in such extreme cases because it does not allow $\rho - \sigma$



8a. Curves from (9) with $\theta = 45°$.



8b. Curves from (10) with $\theta = 45°$.

to be large enough when $0 \ll \theta/\phi < 1$ without making $\sigma - \rho$ too large when $-1 < \theta/\phi < 0$ or moving the cross-over point where $\rho = \sigma$ too close to $\theta/\phi = 0$.

3. Mock Curvature Constraints

Here we need to extend the notation of Section 2 so that $\theta_j = \arg w_j - \arg(z_{j+1} - z_j)$ for $0 \le j < n$, $\phi_j = \arg(z_j - z_{j-1}) - \arg w_j$ for $0 < j \le n$, and d_j is the Euclidean length of the vector $z_{j+1} - z_j$ for $0 \le j < n$. If the problem is to find a closed curve with no directions given, it will be convenient to sometimes use alternative names $z_{n+1}, \theta_n, \phi_{n+1}, \tau_n$, and $\bar\tau_{n+1}$ for $z_1, \theta_0, \phi_1, \tau_0$, and $\bar\tau_1$ respectively. We can then define $\psi_j = \arg(z_{j+1} - z_j) - \arg(z_j - z_{j-1})$ for $1 \le j \le n'$, where $n' = n$ for closed curve problems with no directions given and $n' = n - 1$ otherwise. Unless stated otherwise, all ψ_j are at most 180" in absolute value.

Where w_i has been given in advance, it simply determines ϕ_i and θ_i; other w_i need to be determined by solving for θ_i and ϕ_i. Since the problem of finding direction angles can be broken into independent subproblems separatatcd by knots where directions are given, we can assume that no directions other than w_0 and w_n are given. For closed curve problems we can assume that no directions at all are given, otherwise the problem could be reduced to one or more open curve problems.

The requirement that the curvature be continuous at some knot z_i, $0 < i < n$, is

$$k_1(z_{i-1}, z_i, w_{i-1}, w_i, \tau_{i-1}, \bar\tau_i) = k_0(z_i, z_{i+1}, w_i, w_{i+1}, \tau_i, \bar\tau_{i+1})$$

where k_0 and k_1 are functions that give the curvature at $t = 0$ and $t = 1$ in terms of the endpoints, terminal directions, and tension parameters for the family of curves being used. Because of the requirement for invariance under translation, rotation, and scaling, there exists a function k such that

$$\begin{aligned} k_0(z_j, z_{j+1}, w_j, w_{j+1}, \tau_j, \bar\tau_{j+1}) &= k(\theta_j, \phi_{j+1}, \tau_j, \bar\tau_{j+1})/d_j \text{ and} \\ k_1(z_j, z_{j+1}, w_j, w_{j+1}, \tau_j, \bar\tau_{j+1}) &= k(\phi_{j+1}, \theta_j, \bar\tau_{j+1}, \tau_j)/d_j. \end{aligned} \tag{11}$$

Any particular family of curves determines a specific function k that satisfies (11). The corresponding *mock curvature* function $\hat{k}$ consists of the linear terms in the Taylor series for $k(\theta, \phi, \tau, \bar\tau)$, expanded about $(\theta, \phi) = (0, 0)$. For the curves determined by (2) and (3) with ρ and σ determined by (9) or (10),

$$k(\theta, \phi, \tau, \bar\tau) = \frac{2\sigma(\theta, \phi)\sin(\theta + \phi)/\bar\tau - 6\sin\theta}{\bigl(\rho(\theta, \phi)/\tau\bigr)^2}$$

and

$$\hat{k}(\theta,\phi,\tau,\bar{\tau}) = \tau^2\left(\frac{2(\theta+\phi)}{\bar{\tau}} - 6\theta\right) \tag{12}$$

where the angles are measured in radians. Since the tension parameters are always known in advance, they are treated like constants in this expansion.

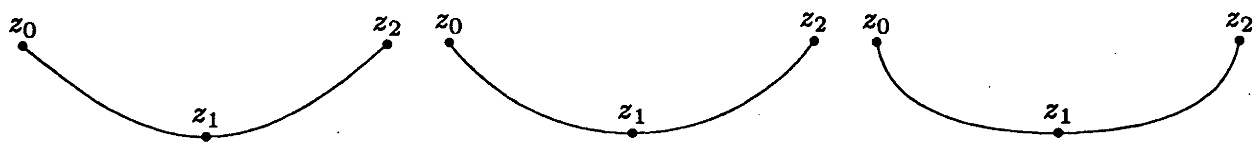
Continuity of mock curvature requires

$$\hat{k}(\phi_i,\theta_{i-1},\bar{\tau}_i,\tau_{i-1})/d_{i-1} - \hat{k}(\theta_i,\phi_{i+1},\tau_i,\bar{\tau}_{i+1})/d_i = 0 \quad \text{for } 1 \le i \le n'. \tag{13}$$

Combining this with the first order continuity equations

$$\theta_i + \phi_i = -\psi_i \quad \text{for } 1 \le i \le n' \tag{14}$$

gives enough equations to determine all θ_i and ϕ_i for closed curve problems. For open curve problems when directions w_0 and w_n are given in advance, these provide the necessary additional equations by fixing θ_0 and ϕ_n, but otherwise additional constraints are needed.



9a. $\chi_0 = \chi_2 = 0.$ 9b. $\chi_0 = \chi_2 = 1.$ 9c. $\chi_0 = \chi_2 = \infty.$

The additional constraints are controlled by special "curl" parameters χ_0 and χ_n. There should be one such constraint for each endpoint where no direction is specified. They have the form

$$\hat{k}(\theta_0,\phi_1,\tau_0,\bar{\tau}_1) = \chi_0\hat{k}(\phi_1,\theta_0,\bar{\tau}_1,\tau_0) \tag{15a}$$

and

$$\hat{k}(\phi_n,\theta_{n-1},\bar{\tau}_n,\tau_{n-1}) = \chi_n\hat{k}(\theta_{n-1},\phi_n,\tau_{n-1},\bar{\tau}_n). \tag{15b}$$

The curl parameters give the ratio of the mock curvature at the endpoints to that at the adjacent knots. They should probably have default values of 1 so that the first and last spline segments will usually be good approximations to circular arcs as in Figure 9b.

We now have a system of equations consisting of (13), (14), and possibly (15a) and/or (15b). If θ_0 or ϕ_n have been given in advance then they may be regarded as constants. The first step is to rewrite (15a) and (15b) as

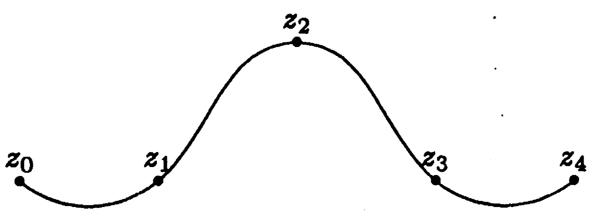
$$\theta_0 = \frac{\tau_0^3 + \chi_0\bar{\tau}_1^3(3\tau_0 - 1)}{\tau_0^3(3\bar{\tau}_1 - 1) + \chi_0\bar{\tau}_1^3}\phi_1 \quad \text{and} \quad \phi_n = \frac{\bar{\tau}_n^3 + \chi_n\tau_{n-1}^3(3\bar{\tau}_n - 1)}{\bar{\tau}_n^3(3\tau_{n-1} - 1) + \chi_n\tau_{n-1}^3}\theta_{n-1} \tag{16}$$

so that θ_0 and ϕ_n can be eliminated. Then (14) may be used to eliminate all ϕ_i so that the remaining variables are $\theta_1, \theta_2, \ldots, \theta_{n'}$, and the remaining equations are those given by (13) with appropriate substitutions. This system has some important properties that may be summarized as follows..

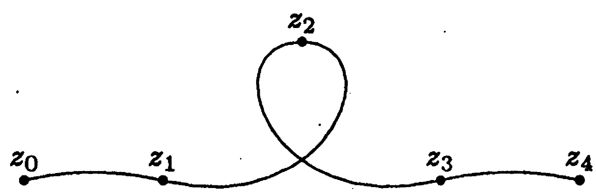
Theorem **1.** If $n \geq 2$, if all tension parameters satisfy *the* bound $\tau_i, \bar{\tau}_i \geq \tau_{\min} > \frac{2}{3}$, and if any curl parameters satisfy $\chi_0, \chi_n \geq 0$, *then* ***after*** *the* aforementioned substitutions, ***all coefficients of*** $\theta_1, \theta_2, \ldots, \theta_{n'}$ in ***(13)*** are nonnegative, and *for each* i, *the* coefficient of θ_i is at ***least*** $3\tau_{\min} - 1$ times *the sum of* all *the other coefficients in* that equation.

Proof. The bounds on the tension parameters guarantee that the coefficient of θ in (12) will be negative, the coefficient of ϕ will be positive, and the magnitude of the former will be at least $3\tau_{\min} - 1$ times the latter. When $1 < i < n'$ in (13), the only relevant substitutions are $\phi_j = -\psi_j - \phi_j$ for $|j - i| \leq 1$, so the coefficients of θ_{i-1}, θ_i and θ_{i+1} clearly have the required properties. For closed curve problems, the same holds for $i = 1$ and $i = n'$, **otherwise** additional substitutions eliminate θ_0 and ϕ_n so that $\hat{k}(\phi_1, \theta_0, \bar{\tau}_1, \tau_0)$ depends only on ϕ_1 and $\hat{k}(\theta_{n-1}, \phi_n, \tau_{n-1}, \bar{\tau}_n)$ depends only on θ_{n-1}. We need only show that both of these variables have non-positive **coefficients.** This is clearly true for given directions, and it also holds for curl constraints since the coefficients in (16) are at most $3\tau_0 - 1$ **and** $3\bar{\tau}_n - 1$ respectively. ▮

Theorem 1 shows that subject to certain reasonable limitations on the **tension** and curl parameters, the system of equations is diagonally dominant, and hence it has a unique solution. Actually, the solution is **unique** only up to the choice of the angles ψ_i. Ordinarily all ψ_i should be chosen so that they are at most 180" in absolute value, but it is possible to add multiples of 360" to **them.** The effect of such a change is usually to add a loop to the curve as in Figure 10.



10a. A spline computed with ψ_2 = -90"

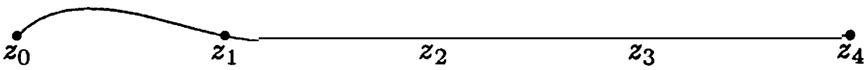


10b. A spline computed with $\psi_2 = \mathbf{270°}$

Theorem 1 also shows that **the** splines have approximate locality in **the** sense that changes in direction angles fall off exponentially as one **moves** away from a disturbance. **Specifically,** suppose a given direction θ_0 is displaced by an angle δ and let A bc the matrix **of coefficients of** $\theta_1, \theta_2, \ldots \theta_{n'}$ from (13) after **the** substitutions. The change in $\theta_1, \theta_2, \ldots \theta_{n'}$ due to this displacement is given by the solution vector Θ to $A\Theta = \delta e_1$ where $e_1 = (1, 0, 0, \ldots, 0)^T$.

We know that A is **tridiagonal** with **nonnegative entries,** and within each row **the diagonal element dominates the sum of the other** two **elements** by at least a factor of $3\tau_{\min} - 1$. It is not hard to **see** that for any two **adjacent components** of Θ, either $\Theta_k = 0$ or $\Theta_{k-1}/\Theta_k < 1 - 3\tau_{\min}$. This is trivial for $k = n'$, and it may be **extended** inductively to smaller k using the fact that $A_{kk} \geq (3\tau_{\min} - 1)(A_{k,k-1} + A_{k,k+1})$. Thus j knots away from **where** a given direction is **changed, the effect** of the change is reduced by at least a factor of $(3\tau_{\min} - 1)^j$. In practice **the** reduction is **often** by a **somewhat** greater amount as in Figure 11 **where** $\tau_{\min} = 1$ and $\theta_0/\theta_1 = -\frac{41}{11}$.

When a knot z_i is displaced, **three** mock **curvature** constraints are directly **affected** due to **changes in** d_{i-1}, d_i, ψ_{i-1}, ψ_i, and ψ_{j+1}. The **adjustment** will **cause** some change



11. Exponential decline in the effect of a 45” change in direction.

in ϕ_{i-1} and θ_{i+1}, and the effect on θ_{i+j} and θ_{i-j} for j > 1 is equivalent to what would happen if directions w_{i+1} and w_{i-1} were given in advance. The change in θ_{i+j} will be at most $1/(3\tau_{\min}-1)^{j-1}$ as great as the change in θ_{i+1}, and the change in θ_{i-j} will be at most $1/(3\tau_{\min}-1)^{j-1}$ as great as the change in ϕ_{i-1}. If the original problem was to find a closed curve with no directions given, then these two effects will add together so that the change in θ_{i+j} will be at most $1/(3\tau_{\min}-1)^{j-1}$ of the change in θ_{i+1} plus $1/(3\tau_{\min}-1)^{n-1-j}$ of the change in ϕ_{i-1}.

4. Conclusion

We have developed a tridiagonal system of linear equations that can be solved in linear time to determine the spline direction at each knot so as to match mock curvatures. It is necessary to use arctangents to set up the system of equations, and to use sines and cosines to recover the resulting spline directions; but this work can be reduced to one arctangent, one sine, and one cosine per knot on the spline.

We have shown that the splines have approximate locality in the sense that changes in direction angles fall off exponentially. The rate of decline depends on how small the tension parameters are allowed to be, but at least a factor of 2 per knot is guaranteed for the default tensions. It should be noted that an exponential decline in angular change does not guarantee that curve displacements decline similarly because it is technically feasible for d_j to be exponential in j.

The curve families discussed in Section 2 and defined by (9) and (10) are somewhat arbitrary, and the concept of mock curvature could bc applied to other families of curves. It would be desirable to find ρ and σ functions simpler than (9) that perform better than (10), although even the simplified functions of (10) produce very good results for problems such as that shown in Figure 2c.

Acknowledgements

The author is indebted to Prof. Donald Knuth of Stanford University for doing the first complete implementation and helping to refine some of the ideas. Prof. Kevin Karplus of Cornell University helped with the numerical investigation of the “optimal” ρ and σ functions.

References

1. Barsky, Brian A., Bentty, John C., “Varying the Betas in Beta-splints,” University of California, Bcrkolcy TR CSD 82/1 12, Dccembcr 1982.
2. Epstein, M. I'., “Parametric Intcrpolntion,” *SIAM Journal* of *Numerical* Analysis 13, 1976, 261–268.
3. Horn, B. K. I?, “The Curve of Least Energy,” Massachusetts Institute of Technology AI. Memo No. 612, January 1981.

4. Knuth, Donald E., Computers *and* Typesetting, Vol. 2: **METRFONT,** Addison Wesley, To appear.
5. Manning, J. R., "Continuity Conditions for Spline Curves," Computer Journal **17,** 1974, **181–186.**
6. Mehlum, Even, Curve ***and*** Surface ***Fitting*** Based on ***Variational Criteriæ for Smoothness,*** Oslo, 1969.