GMR 468

# NONLINEAR INTERPOLATION BY SPLINES, PSEUDOSPLINES, AND ELASTICA

G. Birkhoff
Harvard University
Cambridge, Massachusetts

H. Burchard and D. Thomas
Mathematics Department
General Motors Research Laboratories
Warren, Michigan

February 3, 1965

ABSTRACT

Many efficient linear methods are known for interpolation to plane curves which turn through an angle of less than $180^{\circ}$. Among these methods, linear "spline" interpolation is especially versatile. We describe below various methods for interpolating to general plane curves, including interpolation by (nonlinear) mechanical splines and "elastica", and discuss their relation to an adaptation of linear splines proposed by Fowler and Wilson.

"Linearized splines", defined by piecewise cubic polynomial functions $y = f(x)$ of class $C^2$, have recently been effectively used to represent smooth plane curves whose tangent direction $\theta$ changes by less than $180^\circ$. At least two proposals ([1], [3]) have been made for modifying this approach to define "nonlinear spline interpolation" between points on arbitrary smooth plane curves. The purpose of this note is to relate these proposals to true "nonlinear spline interpolation" by mechanical splines, and to interpolation by more general "elastica". In all cases, we suppose given $n + 1$ points $P_i = (x_i, y_i)$, $i = 0, 1, \cdots, n$, on a smooth plane curve $\Gamma$, and we interpolate a smooth approximating curve through these points.

The most natural way to do this is by trying to minimize the <u>strain energy</u> $(B/2)\int K^2\,ds$ of a mechanical spline of "stiffness" B passing through these points -- K denoting the curvature and s the arc-length. Local minima of this functional describe the positions of stable equilibrium of a mechanical spline constrained to pass through these points, but otherwise free to deform or slip. For such a spline, the Euler-Lagrange variational equations are, if dots refer to derivatives with respect to arc-length,

(1) $$2\ddot{K} + K^3 = 0 \quad \text{if and only if} \quad \delta \int K^2 \, ds = 0 ,$$

which reduces to $y^{IV} = 0$ in the linearized approximation.

True nonlinear spline interpolation consists in passing a curve of class $C^2$ through the $P_i$, which satisfies (1) (or, equivalently, (2)) at all points other than the "joints" $P_i$.

Except in the trivial case $K \equiv 0$ of a straight line, such minima can be only local. For, otherwise, the (positive) integral can be made arbitrarily small by interpolating large circular loops of length $O(r)$ and curvature $K = O(1/r)$, thus making $\int K^2 \, ds = O(1/r)$ less than any preassigned positive number.

We note that, in rectangular coordinates, the ordinary DE (1) assumes the rational form

(2) $$(1 + y'^2)\, y^{IV} - 8\, y'y''y''' - \frac{5}{2}\, y''^3 + 18\, \frac{y'^2\, y''^3}{(1 + y'^2)} = 0 .$$

This defines an interesting well-set two-endpoint boundary value problem, if one is given $y(a) = y(b) = 0$ and small endslopes $y'(a) = y'_o$, $y'(b) = y'_1$. We have not studied its effective computational solution, and have written down the preceding DE's (which are invariant under the four-parameter similarity group) mostly for purposes of comparison.

True nonlinear spline interpolation, defined mathematically in §1, should not be confused with analogous but different schemes. Among these, the most promising one seems to be that proposed by A. H. Fowler and C. H. Wilson. This scheme, which we will call <u>pseudo-spline interpolation</u>, replaces the nonlinear two-endpoint problem of §1 by linearized spline interpolation. It may be described in somewhat simplified form as follows.

Each curvilinear segment $\widehat{P_{i-1}\,P_i}$ is described, relative to Cartesian coordinates $\xi,\eta$ whose axes are respectively parallel to the chord $\overline{P_{i-1}\,P_i}$ and perpendicular to it, by a cubic polynomial $\eta = f_i(\xi)$, $i = 1, \cdots, n$. At each interior joint $P_i$, continuity of slope and curvature are required.

We now describe a procedure, which involves only <u>rational</u> operations, for effectively computing the curve defined by these conditions. Let

$$(3) \qquad a_i = x_i - x_{i-1}, \quad b_i = y_i - y_{i-1}, \quad c_i = \sqrt{a_i^2 + b_i^2},$$

and

$$(4) \qquad \bar{x}_i = (x_{i-1} + x_i)/2, \quad \bar{y}_i = (y_{i-1} + y_i)/2.$$

Then, on the i-th curvilinear segment $\widehat{P_{i-1}\,P_i}$, we make a similarity transformation to the new coordinates

(5a) $$\xi(x,y) = a_i(x - \overline{x}_i) + b_i(y - \overline{y}_i),$$

(5b) $$\eta(x,y) = b_i(x - \overline{x}_i) - a_i(y - \overline{y}_i).$$

The inverse transformation of coordinates is

(6) $$x = \overline{x}_i + (a_i\xi + b_i\eta)/c_i, \qquad y = \overline{y}_i + (b_i\xi - a_i\eta)/c_i.$$

Hence, writing $y' = dy/dx$, we have the formula

(7) $$\frac{d\eta}{d\xi} = \frac{b_i - a_i y'}{a_i + b_i y'} \quad \text{on} \quad \widehat{P_{i-1}\,P_i}.$$

The condition that $\eta = f_i(\xi)$ be a cubic function is satisfied by setting

(8) $$f_i(\xi) = (4\xi^2 - c_i^{\,2})(\alpha_i\xi + \beta_i), \qquad |\xi| \le c_i/2,$$

where the constants $\alpha_i$ and $\beta_i$ determine the endslopes

(9) $$f_i'(-\,c_i/2) = \lambda_i, \qquad f_i'(c_i/2) = \mu_i,$$

and are given by

(10) $$\alpha_i = \frac{\mu_i + \lambda_i}{4c_i^{\,2}}, \quad \beta_i = \frac{\mu_i - \lambda_i}{8c_i}.$$

Applying (7) to (9), the condition for continuity of slope is found to be

$$\lambda_{i+1} = \frac{E_i + D_i\mu_i}{D_i - E_i\mu_i}, \tag{11}$$

where

$$D_i = a_i a_{i+1} + b_i b_{i+1}, \qquad E_i = a_i b_{i+1} - b_i a_{i+1}. \tag{11'}$$

One considers these slopes $\mu_0, \mu_1, \cdots, \mu_{n-1}$ (or $\lambda_1, \cdots, \lambda_n$) as the unknowns to be computed by imposing the condition of continuity of curvature, or equivalently, of $y''$ which is given in the $(\xi_i, \eta_i)$ coordinate system by

$$y'' = -c_i^2\, \eta''/(a_i + b_i\eta')^3, \tag{12}$$

where

$$\eta' = d\eta/d\xi, \qquad \eta'' = d^2\eta/d\xi^2. \tag{12'}$$

The i-th cubic $f_i(\xi)$ connecting $P_{i-1}$ to $P_i$ and the $(i + 1)$-th cubic $f_{i+1}(\xi)$ connecting $P_i$ to $P_{i+1}$ are joined at $P_i$ by continuity of value, slope, and curvature. In particular at the joint $P_i$ one obtains from (8) - (10) the relations

$$\eta_i' = \mu_i, \qquad \eta_i'' = (4\mu_i + 2\lambda_i)/c_i \tag{13}$$

and

(13') $$\eta'_{i+1} = \lambda_{i+1}, \quad \eta''_{i+1} = -(4\lambda_{i+1} + 2\mu_{i+1})/c_{i+1}.$$

Substituting into (12) from (13) and (13') respectively, and equating, one obtains

(14) $$c_i(4\mu_i + 2\lambda_i)/(a_i + b_i\mu_i)^3 = -c_{i+1}(4\lambda_{i+1} + 2\mu_{i+1})/(a_{i+1} + b_{i+1}\lambda_{i+1})^3.$$

Equation (13) may be simplified using the bilinear transformation in (11) to express $(a_{i+1} + b_{i+1}\lambda_{i+1})$ in terms of $(a_i + b_i\mu_i)$, namely

(15) $$(a_{i+1} + b_{i+1}\lambda_{i+1}) = c_{i+1}(a_i + b_i\mu_i)/(D_i - E_i\mu_i)^3.$$

Substituting (15) into (14) the coupling equation can be written as

(16) $$(D_i - E_i\mu_i)^3 (4\lambda_{i+1} + 2\mu_{i+1}) + c_i c_{i+1}^2 (4\mu_i + 2\lambda_i) = 0.$$

Our experiments suggest that the $(n - 1)$ equations (16), although nonlinear, are well suited to fast iterative solution for any set of reasonably behaved (i.e., regularly spaced) points representing a plane curve.

Interpolation by mechanical splines is a special case of interpolation by "elastica" or "thin beams", defined by equations first derived by the Bernoullis and Euler [2, p. 3]. Given the sequence $P_0, P_1, \cdots, P_n$, we define interpolation by elastica with joints at the $P_i$, as a curve $P(s) = (x(s), y(s))$ whose position $P = (x,y)$ as a function of arc-length s satisfies the equilibrium conditions for a thin beam of constant stiffness B, acted on only by forces through the joints $P_i$. Denoting differentiation with respect to s by dots, these conditions [2, §§ 254, 255] are equivalent to:

(18) $\ddot{P}(s)$ is a continuous function of s,

(18') $P(s_i) = P_i$ for suitable $0 = s_0 < s_1 < \cdots < s_n = L$,

(18'') $\dot{x}(s)\,\ddot{y}(s) - \dot{y}(s)\,\ddot{x}(s) = \alpha_i x + \beta_i y + \gamma_i$ on $\widehat{P_{i-1} P_i}$,

where $\alpha_i$, $\beta_i$, $\gamma_i$ are 3n appropriate constants.

In (18''), the case $\alpha_i = \beta_i = 0$ gives a circular arc of curvature $\gamma_i$; it is the case that the stress in $\widehat{P_{i-1} P_i}$ is a pure couple with bending moment $B\gamma_i$. Otherwise, the stress in $\widehat{P_{i-1} P_i}$ is equipollent to the force $\vec{F}_i = B(-\beta_i, \alpha_i)$ acting along the straight line $\alpha_i x + \beta_i y + \gamma_i = 0$.

On each i-th segment $\widehat{P_{i-1} P_i}$, some other basic relations are worth noting. If $\theta$ is the angle between $\vec{F}_i$ and the tangent to the spline, then the curvature $K = \dot{\theta}$ and $\alpha_i \dot{x} + \beta_i \dot{y} = - R_i \sin\theta$ ($R_i = (\alpha_i^2 + \beta_i^2)^{1/2}$). Hence, differentiating (18''), we get

$$\ddot{\theta} = \alpha_i \dot{x} + \beta_i \dot{y} = - R_i \sin\theta, \tag{19}$$

just as in [2], p. 401, (8) (with B = 1). This "simple pendulum" equation can be integrated to give

$$\frac{1}{2} \dot{\theta}^2 = R \cos\theta + \lambda_i; \tag{20}$$

cf. [2], pp. 401-403, Eqs. (7), (9), (13).

By similarity transformations, one can obtain all non-circular "elastica" with given $\mu_i = \lambda_i / F_i$ from any one curve satisfying

$$\frac{1}{2} \dot{\theta}^2 = \cos\theta + \mu_i. \tag{20'}$$

Typical such curves are graphed in [2], pp. 404-405. For $\mu_i = -1$, the curve is a straight line; for $|\mu_i| \leq 1$, it has inflection points; for $\mu_i > 1$, it does not. The case of circular elastica corresponds to $\mu_i = \infty$.

Differentiating (20') and dividing through by $\dot{\theta}$, we get also $\ddot{\theta} = - \sin\theta$. Differentiating again, we get

$$\dddot{\theta} = - \cos\theta \cdot \dot{\theta} = (\mu_i - \frac{1}{2} \dot{\theta}^2)\, \dot{\theta}.$$

Setting $\dot{\theta} = K$ and simplifying, this gives

$$2\ddot{K} + K^3 = 2\mu_i K. \tag{21}$$

The preceding equation, which reduces to (1) when $\mu_i = 0$, is the Euler-Lagrange equation for the variational condition

$$\delta\left[\int (K^2 - 2\mu_i)\right] ds = 0. \tag{21'}$$

Formulas (1), (21), and (21') are not in [2], and may be new.

Comparing (21) with (1), we see clearly that interpolation by mechanical splines is the case $\mu_0 = \mu_1 = \cdots = \mu_n = 0$ of interpolation by elastica, in which tangential forces are permitted to act at the joints. In (20'), the case of mechanical splines (true nonlinear spline interpolation) gives a curve similar to

$$s = \int d\theta/\sqrt{2 \cos \theta}. \tag{22}$$

## CORRELATION

We now correlate the three interpretations of "nonlinear spline interpolation" described in §§ 1-3, respectively.

This correlation depends on the mechanical interpretation of equation (20), based on

$$B \dot{\theta} = B K = N, \quad B R_i \cos\theta = - T$$

(where N is the bending moment and T the tension acting on the spline at the point considered) from which we get by (20)

$$\frac{1}{2} \frac{1}{B^2} N^2 = - \frac{1}{B} T + \lambda_i . \tag{23}$$

The most direct correlation concerns "mechanical splines" and "elastica". Mechanical splines are always elastica with $\lambda_i \equiv 0$. More generally, suppose that we require

i) All external forces at internal joints are (finite) <u>shear</u> forces. As a consequence, the tension BT and bending moment BK will be continuous, the latter being stipulated already in (18); therefore, by (23), <u>$\lambda$ will be continuous</u>, which means that $\lambda$ will be <u>the same constant for all segments</u>. As R can be assumed to be non-negative, this restricts $\lambda/R$ to have the same sign for the entire beam, but only for $\lambda = 0$ will $\lambda/R$ be constant. In general, no further restrictions on the $\mu_i$ can be imposed.

A further natural condition is

ii) Zero bending moment, i.e., $K = 0$ at the endpoints $P_0, P_n$. Then, by (23), a tension $T = \lambda$ has to be applied at the endpoints. Therefore, $\lambda = 0$ is implied by

iii) At all joints, $P_0, P_1, \cdots, P_n$ only (finite) shear forces are exerted on the beam.

On the other hand, the Fowler-Wilson method of §2 is not specifically an approximation to nonlinear splines. It is equally closely related to other families of curves which reduce to cubics in the linearized approximation. Another such family is provided by "Euler's spirals" or clothoids. These are plane curves whose governing DE has the simple form

$$\ddot{K} = 0, \quad \text{or} \quad K = a + bs. \tag{24}$$

These evidently include circles (the case $b = 0$), and are thus more advantageous for some applications than "mechanical splines", for which $\mu_i = 0$ above (circles are "elastica" with $\mu_i = \infty$).

The Effect of Tension. If one wants to consider interpolation by general elastica imposing only condition i) on the external forces, one has to decide on the parameter $\lambda$ which is related to bending moment N and tension T at the endpoints $P_0$, $P_n$ through (23). Schweikert [4] discusses this in the linearized limit, and shows that for some finite tension all extraneous inflection points disappear. Similarly, in the nonlinear case a choice of the parameter $\lambda$ can be used to select a "reasonable" interpolation based on criteria other than those contained in the data, which may sometimes be desirable. For instance, the sequence of points $P_0 = (1,0)$, $P_1 = (2,0)$, $P_2 = (0,2)$, $P_3 = (0,1)$ for $n = 3$ permits "reasonable" interpolation for a wide range of positive $\lambda$,* but no solution satisfying both i) and iii).

## ACKNOWLEDGEMENT

We would like to thank Carl de Boor for his contributions to our work, especially for suggesting equation [illegible]).

* including a circle ($\lambda = \infty$)

## REFERENCES

1. A. H. Fowler and C. W. Wilson, "Cubic spline, a curve fitting routine", Report Y-1400, Oak Ridge, 1963.

2. A. E. H. Love, "A treatise on the mathematical theory of elasticity", fourth edition, 1927.

3. D. H. MacLaren, "Formulas for fitting a spline curve through a set of points", Boeing Appl. Math. Report #2, 1958.

4. Daniel G. Schweikert, "An interpolation curve using a spline in tension" (thesis), to appear.