Copyright © 1983 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 62, No. 1, January 1983
Printed in U.S.A.

Smoothing With Periodic Cubic Splines

By N. Y. GRAHAM

(Manuscript received May 7, 1982)

In this paper we present a mathematical algorithm for constructing a smoothing cubic spline with periodic end conditions and a predetermined 'closeness of fit' to a given set of points in the plane. In addition to providing a mathematical tool for smoothing raw data in which the underlying function is known to be periodic, this algorithm has special significance in computer graphics, because the use of smoothing functions with periodic end conditions is essential for producing visually acceptable, smooth, closed curves. Sample plots are included to illustrate the power and flexibility of this algorithm.

I. INTRODUCTION

Although "natural" splines are used extensively and are quite appropriate for smoothing many types of data, they often produce less than satisfactory results when used to smooth data points that belong to a periodic function. The inappropriateness of using "natural" splines to approximate periodic data is especially evident in graphics applications. In particular, when the data points represent a closed curve, smoothing (parametrically) with "natural" splines will lead to unacceptable results because the "natural" end conditions will cause the curve either to close up with a noticeable discontinuity, or to not close up at all (see Fig. 1).

Existing methods for constructing smoothing splines with a predetermined closeness of fit all lead to splines with "natural" end conditions.[1,2] A method developed by Spath[3] produces a smoothing spline with periodic end conditions, but the closeness of fit cannot be determined in advance. In this paper we will describe a method for constructing a smoothing cubic spline that has periodic end conditions and that also satisfies a predetermined closeness of fit to a given set of data points.

This algorithm has potentially wide applicability, especially in the realm of interactive graphics. It makes possible the computer genera-

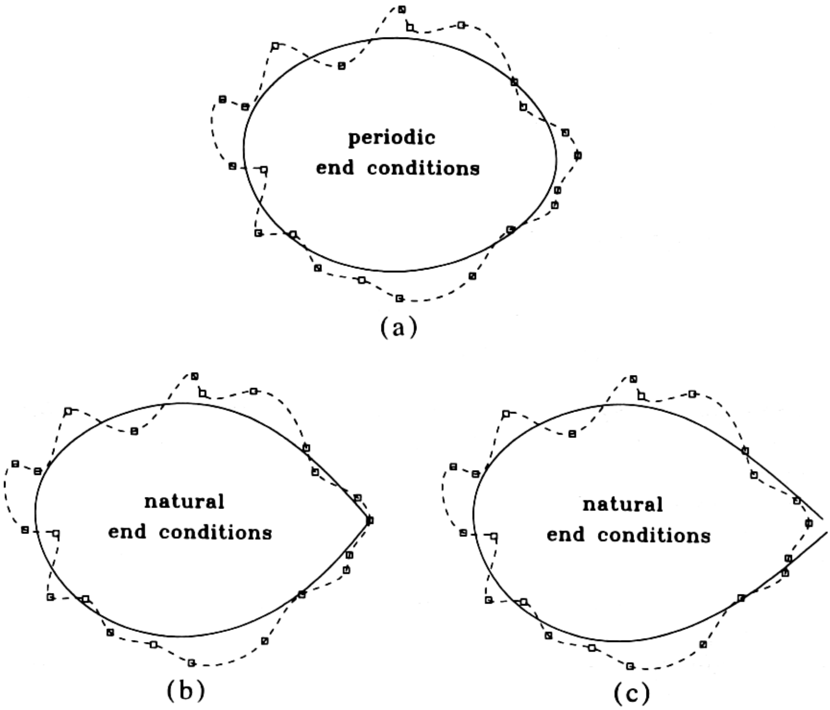

Fig. 1—Comparison of periodic versus natural spline smoothing. (a) Periodic cubic spline smoothing with uniform weights. (b) Natural cubic spline smoothing with small weights at end points. (c) Natural cubic spline smoothing with uniform weights.

tion of free-form, smooth, closed curves by merely specifying the approximate locations of as few as three distinct points. The shape of the curve can be controlled easily by moving one or more points, or by adjusting the weighting factors associated with some or all of the points.

An efficient program based on this new algorithm has been written and tested. Sample plots illustrating this method are included.

II. TERMINOLOGY

Let $P_k = (x_k, y_k)$, $k = 1, n$, be n points in the plane. A "cubic spline" on $[x_1, x_n]$ with knots at $x_1, \cdots, x_n$, is a function f that coincides with a third-order polynomial f_k on each sub-interval $[x_k, x_{k+1}]$, $k = 1, n - 1$, and such that f is continuous and has continuous first and second derivatives over the entire interval $[x_1, x_n]$.

In other words, f is a cubic spline on $[x_1, x_n]$ if, for each $k = 1, n - 1$ there exist real numbers a_k, b_k, c_k, d_k (the "spline coefficients" of f) such that for every x in $[x_k, x_{k+1}]$,

$$f(x) = f_k(x) = a_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3. \qquad (1)$$

Furthermore, the continuity of f, f', and f'' on $[x_1, x_n]$ implies that at each interior knot x_k, $k = 2, n - 1$,

$$f_{k-1}(x_k) = f_k(x_k), \qquad (2)$$

$$f'_{k-1}(x_k) = f'_k(x_k), \qquad (3)$$

$$f''_{k-1}(x_k) = f''_k(x_k). \qquad (4)$$

The cubic spline f is said to be "periodic" if it satisfies the following additional conditions (known as "periodic end conditions"):

$$f(x_n) = f(x_1), \qquad (5)$$

$$f'(x_n) = f'(x_1), \qquad (6)$$

$$f''(x_n) = f''(x_1). \qquad (7)$$

A "natural" cubic spline differs from a "periodic" cubic spline in that it satisfies the so-called "natural end conditions": $f''(x_1) = f''(x_n) = 0$.

III. FORMAL STATEMENT OF THE PROBLEM

Let $P_k = (x_k, y_k)$, $k = 1, n$, be n points in the plane, with $x_1 < x_2 < \cdots < x_n$. Let w_k, $k = 1, n$, be positive real numbers ("weighting factors") associated with P_k, $k = 1, n$, respectively. (Assume $y_n = y_1$, $w_n = w_1$.) Given an arbitrary constant, $M > 0$, the problem is to determine the set of $4(n - 1)$ coefficients of the periodic cubic spline f on $[x_1, x_n]$ with knots at $x_1, \cdots, x_n$, such that

(*i*) f has minimal "total curvature" $G(f) = \int_{x_1}^{x_n} f''(x)^2 dx$, and

(*ii*) f satisfies the following weighted, distance-squared constraint, or "closeness of fit," with respect to the given points:

$$H(f) = \sum_{k=1}^{n} \left[\frac{f(x_k) - y_k}{w_k}\right]^2 \leq M.$$

Note that the weighting factors give inverse importance to the points. Note also that M controls the degree of smoothness, so that increasing the value of M for a fixed set of weighting factors will lead to a smoother, or flatter, spline. Conversely, choosing a sufficiently small value of M will lead to a spline that closely approximates an interpolating spline. In general, an appropriate choice for the value of M will depend on the values chosen for the weighting factors. If, for example, the weighting factors are chosen so that w_k is the standard deviation at x_k, then a suitable choice for M would be some value between the confidence limits, $n - \sqrt{2n}$ and $n + \sqrt{2n}$.

IV. GENERAL APPROACH TO THE SOLUTION

To minimize $G(f)$ subject to the constraint $H(f) \leq M$, we introduce an auxiliary variable z and a non-negative Lagrange multiplier p and minimize

$$F(f) = G(f) + p[H(f) + z^2 - M].$$

This is the approach used by Reinsch in Ref. 2 to solve the analogous problem for "natural" cubic splines. However, Reinsch minimizes F over the class of all continuous functions with continuous first and second derivatives on $[x_1, x_n]$, which leads to a "natural" cubic spline as the solution. We will restrict the class of admissible functions in the minimization of F to periodic cubic spline splines on $[x_1, x_n]$ with knots at $x_1, \cdots, x_n$, obtaining a direct solution to our problem.

V. LINEAR SYSTEM RESULTING FROM CONTINUITY AND PERIODICITY CONDITIONS

Let f be an arbitrary periodic cubic spline on $[x_1, x_n]$, with spline coefficients a_k, b_k, c_k, d_k, $k = 1, n-1$. Then, for x in $[x_k, x_{k+1}]$, $f(x)$ is given by (1) above, and $f'(x)$ and $f''(x)$ are given below:

$$f'(x) = f_k'(x) = b_k + 2c_k(x - x_k) + 3d_k(x - x_k)^2, \tag{8}$$

$$f''(x) = f_k''(x) = 2c_k + 6d_k(x - x_k). \tag{9}$$

Expressing f, f', and f'' explicitly in eqs. (1), (8), and (9), respectively, allows us to derive linear relationships among the spline coefficients. From the continuity of f' at the interior knots of (3) and the periodicity of f' on $[x_1, x_n]$ in (6), it follows that:

$$2c_k h_k = b_{k+1} - b_k - 3d_k h_{k,}^2 \qquad \text{for} \quad k = 1,\ n-1, \tag{10}$$

where $h_k = x_{k+1} - x_k$ for $k = 1, n-1$, and b_n denotes b_1.

From the continuity of f at the interior knots of (2) and the periodicity of f on $[x_1, x_n]$ in (5), the first-order coefficients may be expressed in terms of the constant, second-order, and third-order coefficients:

$$b_k = (a_{k+1} - a_k)/h_k - c_k h_k - d_k h_{k,}^2 \qquad \text{for} \quad k = 1,\ n-1, \tag{11}$$

where a_n denotes a_1.

From the continuity of f'' at the interior knots of (4) and the periodicity of f'' on $[x_1, x_n]$ in (7), the third-order coefficients may be expressed as a function of the second-order coefficients:

$$d_k = (c_{k+1} - c_k)/3h_k, \qquad \text{for} \quad k = 1,\ n-1, \tag{12}$$

where c_n denotes c_1.

Using (11) and (12) to eliminate the b_k's and d_k's from (10) leads to a system of linear equations in the a_k's and c_k's, given in matrix notation as follows:

$$\mathbf{Sc} = 3\mathbf{Qa}, \tag{13}$$

where **S** and **Q** are symmetric, cyclic-tridiagonal matrices of order $n - 1$, and **c**, **a** are the column vectors $(c_1, \cdots, c_{n-1})^T$, $(a_1, \cdots, a_{n-1})^T$, respectively.

The non-zero entries of **S** and **Q** are expressed in terms of the distances between successive knots:

$$\mathbf{S}(k, k) = 2(h_{k-1} + h_k) \qquad \text{for} \quad k = 1, n-1$$

$$\mathbf{S}(k, k+1) = \mathbf{S}(k+1, k) = h_k \qquad \text{for} \quad k = 1, n-2$$

$$\mathbf{S}(1, n-1) = \mathbf{S}(n-1, 1) = h_{n-1}$$

$$\mathbf{Q}(k, k) = -1/h_{k-1} - 1/h_k \qquad \text{for} \quad k = 1, n-1$$

$$\mathbf{Q}(k, k+1) = \mathbf{Q}(k+1, k) = 1/h_k \qquad \text{for} \quad k = 1, n-2$$

$$\mathbf{Q}(1, n-1) = \mathbf{Q}(n-1, 1) = 1/h_{n-1},$$

where h_0 denotes h_{n-1}. By Gershgorin's Theorem[4] it can be shown that **S** is positive definite (and therefore non-singular), while **Q** is positive semi-definite and singular with rank $n - 2$.

VI. LINEAR SYSTEM RESULTING FROM MINIMIZING *F* WITH RESPECT TO THE CONSTANT COEFFICIENTS

Note that (13) is a system of $n - 1$ linear equations in $2(n - 1)$ unknowns: the constant coefficients and the second-order coefficients. We shall derive a second system of $n - 1$ linear equations in these unknowns. We proceed by first showing that H and G can be expressed as functions of the constant coefficients only.

From the spline representation of f in (1), it follows immediately that

$$H = \sum_{k=1}^{n} \left[\frac{f(x_k) - y_k}{w_k}\right]^2 = \sum_{k=1}^{n} \left(\frac{a_k - y_k}{w_k}\right)^2$$

is a function of $a_1, \cdots, a_n$. And since the periodicity of f implies

$$a_n = f(x_n) = f(x_1) = a_1,$$

H is a function of the constant spline coefficients $a_1, \cdots, a_{n-1}$.

From the explicit representation of f'' in (9), the "total curvature" G can be expressed as

$$G = \int_{x_1}^{x_n} f''(x)^2 dx = \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} f''(x)^2 dx$$

$$= \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} [2c_k + 6d_k(x - x_k)^2 dx.$$

Evaluating the integrals over each sub-interval directly and eliminating the d_k's with (12) lead to the following:

$$G = \sum_{k=1}^{n-1} \frac{4}{3} h_k(c_k^2 + c_k c_{k+1} + c_{k+1}^2).$$

Rewriting in matrix notation and applying (13), we have:

$$G = (2/3)\mathbf{c}^T\mathbf{S}\mathbf{c} = (2/3)(3\mathbf{S}^{-1}\mathbf{Q}\mathbf{a})\mathbf{S}(3\mathbf{S}^{-1}\mathbf{Q}\mathbf{a}) = 6\mathbf{a}^T\mathbf{Q}\mathbf{S}^{-1}\mathbf{Q}\mathbf{a}.$$

Thus, G also can be expressed in terms of the constant spline coefficients. Note that from this representation of G,

$$\frac{\partial G}{\partial a_k} = 12\mathbf{Q}_k\mathbf{S}^{-1}\mathbf{Q}\mathbf{a}$$

for $k = 1, n - 1$, where $\mathbf{Q}_k$ is the kth row of $\mathbf{Q}$.

Since G and H are functions of $a_1, \cdots, a_{n-1}$, then F is a function of the independent variables $a_1, \cdots, a_{n-1}, p$, and z. In order for F to be minimized, the partial derivative of F with respect to each of its independent variables must vanish. Thus, for each $k = 1, n - 1$, differentiating F with respect to a_k yields:

$$\frac{\partial F}{\partial a_k} = \frac{\partial G}{\partial a_k} + p\frac{\partial H}{\partial a_k} = 12\mathbf{Q}_k\mathbf{S}^{-1}\mathbf{Q}\mathbf{a} + 2p\left(\frac{a_k - y_k}{w_k^2}\right) = 0.$$

Rewriting this set of $n - 1$ linear equations in matrix notation and using (13) to replace $\mathbf{S}^{-1}\mathbf{Q}\mathbf{a}$ with $\mathbf{c}/3$ lead to:

$$4\mathbf{Q}\mathbf{c} + 2p\mathbf{W}^{-2}(\mathbf{a} - \mathbf{y}) = \mathbf{0}, \tag{14}$$

where $\mathbf{y}$ is the column vector $(y_1, \cdots, y_{n-1})^T$. Combining (13) and (14), we have the following linear system in $\mathbf{c}$:

$$(p\mathbf{S} + 6\mathbf{Q}\mathbf{W}^2\mathbf{Q})\mathbf{c} = 3p\mathbf{Q}\mathbf{y}. \tag{15}$$

The matrix $\mathbf{A}_p = (p\mathbf{S} + 6\mathbf{Q}\mathbf{W}^2\mathbf{Q})$ is symmetric, five-banded with three non-zero entries in the upper right and lower left corners. It can be shown that, for all positive values of p, $\mathbf{A}_p$ is positive definite. Thus, for each $p > 0$, (15) has a unique solution in $\mathbf{c}$.

Note that for each non-zero value of p, $\mathbf{a} = \mathbf{y} - (2/p)\mathbf{W}^2\mathbf{Q}\mathbf{c}$ from (14). Note also that from (12) $\mathbf{d}$ is uniquely determined by $\mathbf{c}$, and from (11) $\mathbf{b}$ is uniquely determined by $\mathbf{a}$, $\mathbf{c}$, and $\mathbf{d}$. Thus, to each positive value of p corresponds a unique periodic cubic spline on $[x_1, x_n]$, whose coefficients are given in the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$.

VII. CONSEQUENCES OF MINIMIZING *F* WITH RESPECT TO *p* AND *z*

Minimizing $F = G + p(H + z^2 - M)$ with respect to the Lagrange multiplier p leads to:

$$\frac{\partial F}{\partial p} = H + z^2 - M = 0,$$

which merely states that the distance constraint on H (expressed as an equality in terms of the auxiliary variable z) must be satisfied when the minimal value of F is attained.

On the other hand, minimizing F with respect to z yields:

$$\frac{\partial F}{\partial z} = 2pz = 0,$$

which implies that at least one of the two variables, p or z, must be equal to 0 when the minimal value of F is attained.

Note that if $p = 0$ when F is minimized, then $F = G$. Since $G = (2/3)\mathbf{c}^T\mathbf{S}\mathbf{c}$ and $\mathbf{S}$ is positive definite, then G is minimized when $\mathbf{c} = \mathbf{0}$. This in turn implies $\mathbf{d} = \mathbf{0}$, so that the second- and third-order coefficients vanish, resulting in a piecewise linear minimizing spline. The properties of being piecewise linear and having a continuous first derivative together imply that the minimizing spline is a straight line. Furthermore, periodicity of the spline implies that the straight line is in fact horizontal.

On the other hand, if $p > 0$ when F is minimized, then $z = 0$, so that $H = M$. Since $\mathbf{a} = \mathbf{y} - (2/p)\mathbf{W}^2\mathbf{Q}\mathbf{c}$ from (14), and $\mathbf{c} = 3p\mathbf{A}_p^{-1}\mathbf{Q}\mathbf{y}$ from (15), H can be expressed as a function of p. Thus, if minimization of F occurs for a positive value of p, it remains to determine the value of p for which $H(p) = M$.

VIII. PROPERTIES OF H AS A FUNCTION OF p

The following facts can be established: for all positive values of p, $H(p)$ is a continuous, convex function of p with negative slope. Furthermore, as p approaches zero from the right, $H(p)$ becomes arbitrarily large.

IX. ALGORITHM FOR DETERMINING SPLINE COEFFICIENTS

We can now state the following algorithm for determining the minimizing spline. Compute the equation of the horizontal straight line with the least-squares fit to the given data points:

$$f(x) = \left(\sum_{k=1}^{n} \frac{y_k}{w_k^2}\right) \Big/ \left(\sum_{k=1}^{n} \frac{1}{w_k^2}\right).$$

Determine if this line satisfies the distance constraint on $H(f)$. If it does, we are done. If it does not satisfy the distance constraint, start with some positive value of p and search for the value of p for which $H(p) = M$, using a combination of Newton's method when moving to the right and a binary search when moving to the left (or any applicable search). Insert this value of p in (15), solve the linear system for $\mathbf{c}$, and compute the related values of $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{d}$ using (14), (11), and (12). The periodic cubic spline associated with this value of p will be our solution.

X. PRACTICAL CONSIDERATIONS IN SOLVING THE LINEAR SYSTEM

Since the matrix $\mathbf{A}_p$ is positive definite for each $p > 0$, it can be decomposed into the product of a lower triangular matrix $\mathbf{R}$ and its transpose, using the square-root (Cholesky's) method.[5] The linear system $\mathbf{A}_p\mathbf{c} = 3p\mathbf{Q}\mathbf{y}$ can then be solved efficiently in two steps, by applying forward substitution to the lower triangular system $\mathbf{R}\mathbf{v} = 3p\mathbf{Q}\mathbf{y}$, followed by backward substitution to the upper triangular system $\mathbf{R}^T\mathbf{c} = \mathbf{v}$.

Furthermore, since $\mathbf{A}_p$ is symmetric and five-banded with three non-zero entries in two corners, its decomposition $\mathbf{R}$ will consist of three non-zero bands (the main diagonal and the two diagonals below it) and two non-zero rows along the bottom, so that $\mathbf{R}$ can be stored in fewer than $5n$ locations. (The entries of $\mathbf{A}_p$ need not be stored; they may be computed as needed.)

The sparseness of the matrix $\mathbf{R}$ and its structure described above lead not only to its compact storage, but also to the linear time solution of the upper and lower triangular systems, and hence to the linear time solution of the system $\mathbf{A}_p\mathbf{c} = 3p\mathbf{Q}\mathbf{y}$, for each non-zero value of p.

It should be pointed out that, unless the number of data points to be smoothed is rather limited (approximately 30 or fewer), the straightforward application of Cholesky's method to decompose $\mathbf{A}_p$ will encounter underflow problems. This is due to the fact that, as the dimension of $\mathbf{A}_p$ increases, entries with exponentially decreasing magnitudes will appear in its decomposition $\mathbf{R}$. This difficulty can be circumvented by truncating sufficiently small values to zero, while still retaining single-precision accuracy in the solution of the triangular systems. (Truncation has the additional advantage of significantly reducing the number of arithmetic operations required in computing the entries of $\mathbf{R}$ when the number of data points is large.)

The program implementing this algorithm has been tested on an IBM-370 computer using single-precision arithmetic, and has successfully smoothed up to 250 points before encountering detectable round-off errors. On the average, the Newton and binary search converged after six to eight iterations, independent of the number of data points.

XI. SAMPLE PLOTS

The data points in Figs. 1 and 2 were generated by adding random noise to an ellipse. The dotted curves represent cubic spline interpo-

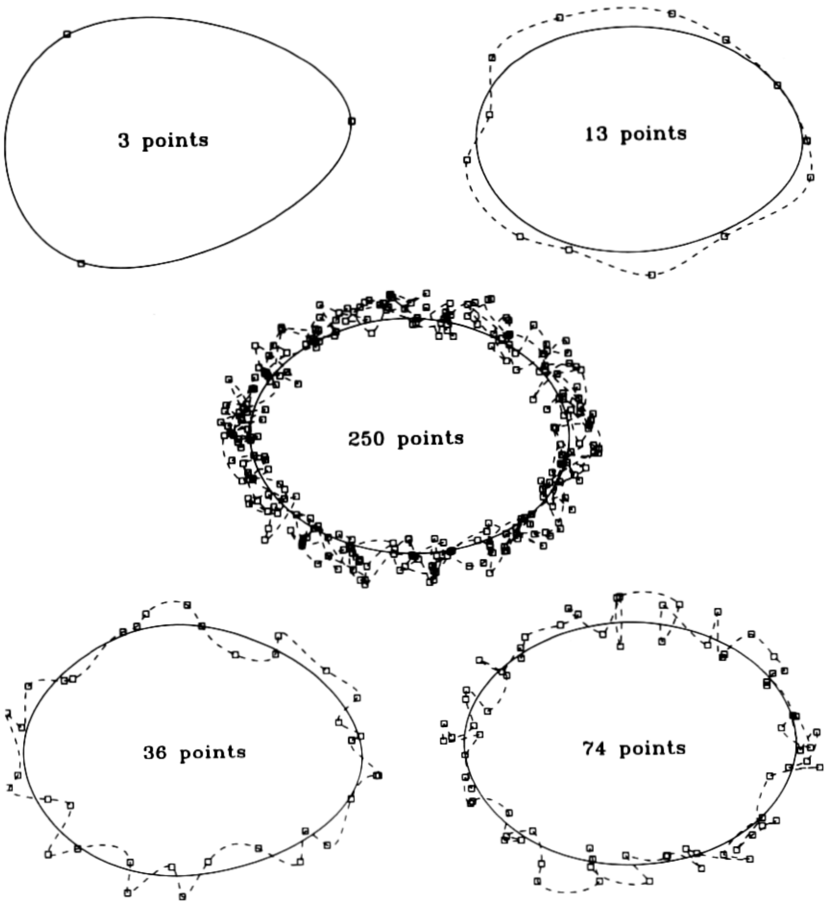

Fig. 2—Periodic cubic spline smoothing for a varying number of data points with uniform weights.

lation of the data (using periodic splines), while the solid curves represent cubic spline smoothing of the same data. In each case a parameter was introduced so that the curve could be represented as two separate single-valued functions of the parameter. Then smoothing was performed twice, once with the x values as a function of the parameter and then again with the y values as a function of the parameter. The smoothed x and smoothed y values were then plotted against the parameter to produce the closed curve. Figure 2 illustrates the algorithm with a varying number of data points, from only three distinct points to 250 points. In each case, uniform weights were used. A "tight" fit was chosen in the example with three points to show how the method can be used to simulate periodic interpolation of the points.

REFERENCES

1. Carl de Boor, *A Practical Guide to Splines*, New York: Springer-Verlag, 1978.
2. Christian H. Reinsch, "Smoothing by Spline Functions," Numerische Mathematik, *10* (1967), pp. 177–83.
3. H. Spath, "Spline Algorithms for Curves and Surfaces," Winnepeg: Utilitas Mathematica Publishing Inc., 1974.
4. David I. Steinberg, *Computational Matrix Algebra*, New York: McGraw-Hill, Inc., 1974, p. 251.
5. V. N. Faddeeva, *Computational Methods of Linear Algebra*, New York: Dover Publications, Inc., 1959.