Contours for <PLOT>

Harold V. McIntosh
Departamento de Aplicación de Microcomputadoras
Instituto de Ciencias, Universidad Autónoma de Puebla
Apartado Postal 461, (72000) Puebla, Puebla, México.

March 18, 1989
revised January 14, 2001

Abstract

Techniques used to plot contours in the package <PLOT> are described. Aside from its using traditional interpolation formulas, <PLOT> is characterized by taking pairs of triangles rather than rectangles for primitive domains, and for tracing out one entire contour before starting another, to minimize pen raising and lowering.

Contents

1 Interpolation

Almost all the interpolation formulas used in numerical analysis which depend upon linear quantities such as the values, derivatives, finite differences, or the like, of data or functions can be summarized in a simple determinantal formula. For simplicity, let us consider the representation of a polynomial of second degree as a function of a single variable. We might write

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x \\ {x_1}^2 & {x_2}^2 & {x_3}^2 & x^2 \\ y_1 & y_2 & y_3 & y \end{vmatrix} = 0. \tag{1}$$

By inspection, it can be verified that the equation is that of a second degree polynomial, and that by well known properties of determinants, the equation is satisfied for the three pairs of points (x_1, y_1), (x_2, y_2), and (x_3, y_3).

For a higher degree polynomial it is only necessary to add more rows containing powers of the variable, and more columns with additional data points. Indeed, the method applies equally well to a function of several variables; it is only necessary to have rows for each of the basis monomials and an equal number of columns.

By using Laplace's expansion for determinants, this equation can be given a polynomial form with explicit coefficients. However, it is easy to derive a very compact form of the equation by using a little matrix algebra. Recall that the determinant of a matrix is zero if and only if there is a nonzero vector whose product with the matrix vanishes. It is convenient to partition the matrix and the corresponding vector in the following way:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x \\ {x_1}^2 & {x_2}^2 & {x_3}^2 & x^2 \\ \hline y_1 & y_2 & y_3 & y \end{array}\right] \left[\begin{array}{c} u_1 \\ u_2 \\ u_3 \\ \hline v \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline 0 \end{array}\right].$$

If suitable symbols are introduced for the submatrices, we have

$$\begin{bmatrix} M & X \\ Y^T & y \end{bmatrix} \begin{bmatrix} U \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

which can be decomposed into two separate equations,

$$\begin{aligned} MU + Xv &= 0, \\ Y^T U + yv &= 0. \end{aligned}$$

If $v = 0$ then U cannot be zero (we require a nonzero vector overall), and M would have to be singular, by the first equation. However M is the well known

Vandermonde matrix for a polynomial basis, whose determinant cannot vanish when its data points are distinct. In the general case M is a generalization of the Vandermonde matrix, to which a similar conclusion applies. Thus it is safe to assume that M and v are both invertible.

Therefore we can rearrange the first equation to get

$$U = -M^{-1}Xv \tag{2}$$

and substitute it into the second to obtain

$$y = Y^T M^{-1} X. \tag{3}$$

The components of the row vector $Y^T M^{-1}$ depend on products of two factors, the first of which is the value of the function to be represented. The second is a geometric factor, which depends only on the location of the data points and not at all on the value of the function at those points. Combined they present an explicit formula for the coefficients of the interpolating polynomial. The Vandermonde matrix has to be evaluated only once for any given set of data points.

2 The Vandermonde matrix

The Vandermonde matrix and its determinant have such a regular structure that it is worth examining it in further detail. One starting point is the matrix

$$D = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -c & -b & -a \end{bmatrix},$$

whose column eigenvectors satisfy the equation

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -c & -b & -a \end{bmatrix} \begin{bmatrix} 1 \\ u \\ v \end{bmatrix} = \begin{bmatrix} u \\ v \\ -c - bu - av \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ u \\ v \end{bmatrix}.$$

Accordingly the components of the eigenvector satisfy

$$\begin{aligned} u &= \lambda, && (4) \\ v &= \lambda u, && (5) \\ -c - bu - av &= \lambda v. && (6) \end{aligned}$$

which means that λ is a root of the equation

$$\chi(\lambda) = \lambda^3 + a\lambda^2 + b\lambda + c = 0, \tag{7}$$

and that the eigenvectors are columns of the Vandermonde matrix formed from the eigenvalues. In turn it follows that the coefficients in the bottom row of D are sums of homogeneous products of the eigenvalues, as befits the coefficients of a polynomial.

The advantage of this discovery is that it is equally easy to deduce the row eigenvectors of D, which is a painless way to invert the Vandermonde matrix, due to the biorthogonality of the two sets of eigenvectors..

$$\begin{bmatrix} u & v & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -c & -b & -a \end{bmatrix} = \begin{bmatrix} -c & u-b & v-a \end{bmatrix} = \lambda \begin{bmatrix} u & v & 1 \end{bmatrix}$$

This time the components of the eigenvector satisfy

$$\begin{aligned} -c &= \lambda u, && (8) \\ u-b &= \lambda v, && (9) \\ v-a &= \lambda. && (10) \end{aligned}$$

Again the eigenvector components are defined by a recursive process which can be followed if the constants are understood to be homogeneous product sums of the eigenvalues. Thus c, which is the product of all the eigenvalues, is divided by one of them (say λ_i) to obtain u. To obtain v, the product from which λ_i was just omitted is subtracted from the sum of all products in which one factor at a time is omitted, leaving products from which the factor λ_i can be divided out. The process continues in similar fashion to produce a vector whose components are the coefficients of the polynomial

$$\begin{aligned} \xi_i(\lambda) &= \frac{\chi(\lambda)}{(\lambda-\lambda_i)}, && (11) \\ &= \xi_{i,3}\lambda^2 + \xi_{i,2}\lambda + \xi_{i,1}. && (12) \end{aligned}$$

The inverse of the Vandermonde matrix requires normalized row eigenvectors, obtainable by dividing the i-th row by $\xi_i(\lambda_i)$, which turns out to be $\chi'(\lambda_i)$. The polynomials π_i, whose coefficients form the rows of the inverse matrix, are usually written in factored form and are widely known as *Lagrange interpolation polynomials*; they are now written as a sum of powers to fit the matrix format. For $n+1$ data points,

$$\begin{aligned} \pi_i(\lambda) &= \frac{(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_{n+1})}{(\lambda_i-\lambda_1)(\lambda_i-\lambda_2)\cdots(\lambda_i-\lambda_{n+1})} && (13) \\ &= \pi_{i,n+1}\lambda^n + \pi_{i,n}\lambda^{n-1} + \ldots + \pi_{i,1}. && (14) \end{aligned}$$

For the quotient to make sense, it is understood that the i-th factor is omitted from both numerator and denominator. The terms $\pi_{i,j}$ are directly the elements of the inverse Vandermonde matrix; for our example

$$\begin{aligned} M^{-1} &= \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix} && (15) \\ &= \begin{bmatrix} \frac{x_2x_3}{(x_1-x_3)(x_2-x_1)} & \frac{x_2+x_3}{(x_1-x_3)(x_2-x_1)} & \frac{1}{(x_1-x_3)(x_2-x_1)} \\ \frac{x_3x_1}{(x_1-x_3)(x_2-x_1)} & \frac{x_3+x_1}{(x_1-x_3)(x_2-x_1)} & \frac{1}{(x_1-x_3)(x_2-x_1)} \\ \frac{x_1x_2}{(x_1-x_3)(x_2-x_1)} & \frac{x_1+x_2}{(x_1-x_3)(x_2-x_1)} & \frac{1}{(x_1-x_3)(x_2-x_1)} \end{bmatrix}. && (16) \end{aligned}$$

Of course, inserting these results into $y = Y^T M^{-1} X$ yields Lagrange's formula,

$$y = \sum_{i=1}^{n} y_i \pi_i(x). \tag{17}$$

3 Confluence

Suppose that in the first example,

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x \\ {x_1}^2 & {x_2}^2 & {x_3}^2 & x^2 \\ y_1 & y_2 & y_3 & y \end{vmatrix} = 0, \tag{18}$$

$x_2 = x_1$, which could happen if the data points were not distinct. By repeating a column, the determinant would vanish, invalidating the ensuing derivation when it came time to divide by the Vandermonde determinant. Foreseeing such a coincidence, the first column could be subtracted from the second without altering the determinant's value. Then,

$$\begin{vmatrix} 1 & 0 & 1 & 1 \\ x_1 & x_2 - x_1 & x_3 & x \\ {x_1}^2 & {x_2}^2 - {x_1}^2 & {x_3}^2 & x^2 \\ y_1 & y_2 - y_1 & y_3 & y \end{vmatrix} = 0. \tag{19}$$

Recognizing $(y_2 - y_1)/(x_2 - x_1)$ as an approximation to a derivative, the common factor $(x_2 - x_1)$ could be removed from the second column and discarded, inasmuch as the vanishing of the surviving determinant will now take place on its own merits.

The result will be

$$\begin{vmatrix} 1 & 0 & 1 & 1 \\ x_1 & 1 & x_3 & x \\ {x_1}^2 & 2x_1 & {x_3}^2 & x^2 \\ y_1 & {y_1}' & y_3 & y \end{vmatrix} = 0. \tag{20}$$

Given clusters of data points, it is not hard to see that higher derivatives can be used to remove the redundancies. At one extreme, all the points coincide so all the successive derivatives would be used, resulting in a *taylor series.*

The result would be

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ a & 1 & 0 & x \\ a^2 & 2a & 2 & x^2 \\ y(a) & y'(a) & y''(a) & y \end{vmatrix} = 0. \tag{21}$$

which works out to be

$$y(x) = \begin{bmatrix} y(a) & y'(a) & y''(a) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ a^2 & 2a & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \tag{22}$$

$$= \frac{1}{2}\left[\begin{array}{ccc} y(a) & y'(a) & y''(a) \end{array}\right]\left[\begin{array}{ccc} 2 & 0 & 0 \\ -2a & 1 & 0 \\ a^2 & -2a & 1 \end{array}\right]\left[\begin{array}{c} 1 \\ x \\ x^2 \end{array}\right] \quad (23)$$

$$= \left[\begin{array}{ccc} y(a) & y'(a) & y''(a) \end{array}\right]\left[\begin{array}{c} 1 \\ x-a \\ \frac{1}{2}(x-a)^2 \end{array}\right], \quad (24)$$

generalizing to the well-known formula

$$y(x) = \sum_{i=0}^{n} \frac{y^{(i)}(a)}{i!}(x-a)^i. \quad (25)$$

A more moderate arrangment would coalesce the data in distinct pairs to define polynomials in terms of their values and derivatives at selected points, something known as *Hermite interpolation*.

4 Specialized data points

The selection of particular collections of data points leads to specialized forms of Lagrange's interpolation formula, many of which are of common enough ocurrence to deserve special names and treatments.

By choosing $x_i = -i$, the interpolation points become integers, running backwards from zero. For N points the master polynomial will be

$$x^{(N)} = x(x+1)\cdots(x+N-1), \quad (26)$$

the subscript intended to imply something like a power, but more like factorial, but with increasing factors. In common with powers, the exponent N tells the number of terms in the product, and $x^{(0)}$ can be taken as 1. Evidently, these "factorials" can be rewritten as polynomials,

$$x^{(N)} = \sum_{i=0}^{i=N} S_i^N x^i, \quad (27)$$

Their coefficients are called *Stirling numbers of the first kind.* The inverse relation,

$$x^N = \sum_{i=0}^{i=N} \mathcal{S}_i^N x^{(i)}, \quad (28)$$

defines *Stirling numbers of the second kind.* The two kinds of numbers evidently satisfy an orthogonality relation, expressing the fact that they are elements of mutually inverse matrices.

By defining a difference operator for functions:

$$\Delta f(x) \quad = \quad f(x) - f(x-1), \quad (29)$$

we can observe that

$$\begin{aligned}\Delta x^{(N)} &= x^{(N)} - (x-1)^{(N)} \\ &= x(x-1)\cdots(x+N-1) - (x-1)x\cdots(x-1+N-1) \\ &= x(x-1)\cdots(x+N-1-1)((x+N-1)-(x-1)) \\ &= Nx^{(N-1)}, \end{aligned} \tag{30}$$

which in remarkably similar to the formula for the derivative of powers. Note that the operator takes *backward* differences, subtracting the function at a lesser value of its argument from its present value.

Another useful identity, which is a direct consequence of the definition, is

$$x^{(N+1)} = x^{(N)}(x+n), \tag{31}$$

It is convenient to introduce a second kind of factorial,

$$x_{(N)} = x(x-1)\cdots(x-N+1), \tag{32}$$

wherein x is decremented rather than incremented in successive factors. Its relation to the first is

$$x_{(N)} = (-1)^N(-x)^{(N)}, \tag{33}$$

The Stirling matrix relates the "factorial powers" to normal powers, but a different relationship is called for when the interpolation polynomials for a fixed number of points are referred to normal polynomials. The point is that in this context, what is required is the inverse of the Vandermonde matrix, whose elements are coefficients of the individual Lagrange interpolation polynomials, not the coefficients of the one single master polynomial. Thus we would need the coefficients of

$$\begin{aligned} v_i(x) &= \frac{(x-0)}{(i-0)}\frac{(x-1)}{(i-1)}\frac{(x-2)}{(i-2)}\cdots\frac{(x-i)}{(i-i)}\cdots\frac{(x-n+2)}{(i-n+2)}\frac{(x-n+1)}{(i-n+1)}\frac{(x-n)}{(i-n)} \\ &= \frac{x_{(i)}}{i_{(i)}}\frac{(x-(i+1))_{(n-(i+1))}}{(-1)_{(n-(i+1))}}. \end{aligned} \tag{34}$$

5 Three-point two-dimensional approximation

The fundamental hypothesis of the <PLOT> contour programs is that linear interpolation will suffice. Since the monomials 1, x, and y form a basis or first degree functions of two variables, the Vandermonde matrix for the combination will have dimension 3. Usually planar data is taken from a square or rectangular grid; if the corners of the grid squares are used as base points, there will be one too many. Rather than worry about all the ways that contour lines can cross a square, <PLOT> divides each square into two triangles, which are then treated separately. The three vertices of each half square fit nicely into the interpolation formula. When the approximation of linearity is grossly erroneous, the existence of a preferred diagonal becomes quite visible in the erratic wanderings of the contour.

For this case the basic equation has the form

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x \\ y_1 & y_2 & y_3 & y \\ z_1 & z_2 & z_3 & z \end{vmatrix} = 0.$$

The Vandermonde matrix is

$$M = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}.$$

The inverse matrix is

$$M^{-1} = \frac{1}{V} \begin{bmatrix} x_2y_3 - y_2x_3 & y_3 - y_2 & x_3 - x_2 \\ x_3y_1 - y_3x_1 & y_1 - y_3 & x_1 - x_3 \\ x_1y_2 - y_1x_2 & y_2 - y_1 & x_2 - x_1 \end{bmatrix},$$

with Vandermonde determinant

$$V = x_2y_3 + x_3y_1 + x_1y_2 - y_2x_3 - y_3x_1 - y_1x_2.$$

However it is arrived at, the equation can be exhibited in the explicit form

$$\begin{vmatrix} 1 & 1 & 1 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} x + \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ z_1 & z_2 & z_3 \end{vmatrix} y + \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} z + \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} = 0.$$

For the purposes of contouring, it is fortunate that the basic equation is an implicit equation of considerable symmetry. For purposes of interpolation, z can be expressed as a function of x and y, but it is equally possible to assign z a value and solve for y as a function of x.

Even so, the equations appear rather formidable. If a specific contour value z_0 has been chosen, the term in z can be combined with the constant term obtain an equation for a line in a plane:

$$\begin{vmatrix} 1 & 1 & 1 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} x + \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ z_1 & z_2 & z_3 \end{vmatrix} y + \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 - z_0 & z_2 - z_0 & z_3 - z_0 \end{vmatrix} = 0.$$

For uniformity of appearance, z_0 could be subtracted from the last rows of the two left hand determinants. Further consolidation results from selecting a specific value for one of the remaining variables, say x_0. Then the formula for y would be

$$\begin{vmatrix} 1 & 1 & 1 \\ x_1 - x_0 & x_2 - x_0 & x_3 - x_0 \\ z_1 - z_0 & z_2 - z_0 & z_3 - z_0 \end{vmatrix} y + \begin{vmatrix} x_1 - x_0 & x_2 - x_0 & x_3 - x_0 \\ y_1 & y_2 & y_3 \\ z_1 - z_0 & z_2 - z_0 & z_3 - z_0 \end{vmatrix} = 0;$$

still, this and similar results follow directly from original determinant.

6 Four-point two-dimensional approximation

Interpolation over triangles to get a linear approximation in two dimensions conflicts with the rectangular layout over a grid characteristic of most data sets, although the problem is readily resolved by splitting the rectangles into a pair of triangles.

Another approach to interpolating rectangular data is to use the cartesian product of coordinatewise linear interpolation. It has the further advantage that the corners of the coordinate rectangles can be the four interpolation points which the method requires.

Consider the Vandermonde determinant

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_1 & x_2 & x \\ y_1 & y_1 & y_2 & y_2 & y \\ x_1y_1 & x_2y_1 & x_1y_2 & x_2y_2 & xy \\ z_{11} & z_{21} & z_{12} & z_{22} & z \end{vmatrix} = 0. \qquad (35)$$

The weight matrix is a Kronecker product

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_1 & x_2 \\ y_1 & y_1 & y_2 & y_2 \\ x_1y_1 & x_2y_1 & x_1y_2 & x_2y_2 \end{bmatrix} \qquad (36)$$

$$= \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ y_1 & y_2 \end{bmatrix} \qquad (37)$$

which therefore has the inverse

$$M^{-1} = \frac{1}{x_2 - x_1} \begin{bmatrix} x_2 & -1 \\ -x_1 & 1 \end{bmatrix} \otimes \frac{1}{y_2 - y_1} \begin{bmatrix} y_2 & -1 \\ -y_1 & 1 \end{bmatrix} \qquad (38)$$

$$= \frac{1}{(y_2 - y_1)(x_2 - x_1)} \begin{bmatrix} x_2y_2 & -y_2 & -x_2 & 1 \\ -x_1y_2 & y_2 & x_1 & -1 \\ -x_2y_1 & y_1 & x_2 & -1 \\ x_1y_1 & -y_1 & -x_1 & 1 \end{bmatrix} \qquad (39)$$

This time the interpolation formula $z = Y^T M^{-1} X$ uses the function values at the corners of the rectangle,

$$Y^T = [z_{11}, z_{21}, z_{12}, z_{22}]$$

as a row of coefficients, with the variables {1, x, y, xy} constituting X. For purposes of contouring, however, the Vandermonde determinant would have been expanded differently, making use of a Laplace expansion according to the last column to get

$$A - Bx + Cy - Dxy + Ez = 0 \qquad (40)$$

and thence

$$y = -\frac{A - Bx + Ez}{C - Dx} \qquad (41)$$

The pole in this formula is nice because it is capable of accomodating a saddle point in the contours, although the formula will still have problems with maxima and minima.

7 Two dimensional quadratic approximation

Failure of a linear approximation to a surface will occur whenever the triangulation of the coordinate grid is insufficiently fine, which is normally compensated by reducing the grid size. Near critical points, however, a smaller grid only makes matters worse. The remedy is to use a quadratic approximation, unless the critical point is of still higher order. The most general second degree polynomial in two variables has the form

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F,$$

which implies a basis composed of six monomials and therefore a 7×7 basic determinant

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y \\ {x_1}^2 & {x_2}^2 & {x_3}^2 & {x_4}^2 & {x_5}^2 & {x_6}^2 & x^2 \\ x_1y_1 & x_2y_2 & x_3y_3 & x_4y_4 & x_5y_5 & x_6y_6 & xy \\ {y_1}^2 & {y_2}^2 & {y_3}^2 & {y_4}^2 & {y_5}^2 & {y_6}^2 & y^2 \\ z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z \end{vmatrix} = 0.$$

It would not seem to be advisable to try to express such large Vandermonde determinants in symbolic form, so any particular instance should be inverted numerically to obtain the row of coefficients $Y^T M^{-1}$. An inhomogeneous second order polynomial can be written as a quadratic form,

$$z = U^T Q U,$$

where

$$U^T = \left[\begin{array}{ccc} 1 & x & y \end{array}\right],$$

and

$$Q = \left[\begin{array}{ccc} F & E & D \\ E & C & B \\ D & B & A \end{array}\right].$$

The contour lines for z can be represented by y as a function of x if the terms in y are gathered together and expressed as a quadratic form. Write the previous equation in the form

$$Cy^2 + 2(Bx + E)y + (Ax^2 + 2Dx + F - z) = 0,$$

using the coefficients gotten from inverting the Vandermonde matrix, and solve it for y using the quadratic formula,

$$y = \frac{-(Bx + E) + \sqrt{(Bx + E)^2 - C(Ax^2 + 2Dx + F - z)}}{C}.$$

The other half of the curve could be traced out with the negative square root.

Nevertheless there is a better way to procede. Traditionally the cross term is eliminated by rotating x and y,

$$x = x' \cos\ \theta + y' \sin\ \theta, \quad (42)$$

$$y = -x' \sin\ \theta + y' \cos\ \theta. \quad (43)$$

Substituting these values into $2Bxy$ then leaves, after some trigonometric manipulation, a coefficient for $x'y'$,

$$(A - C) \sin\ 2\theta + 2B \cos\ 2\theta,$$

which will vanish if θ is chosen to be

$$2\theta = \tan^{-1} \frac{2B}{C - A}.$$

The expression becomes

$$(A \cos^2\ \theta - 2B \sin\ \theta \cos\ \theta + C \cos^2\ \theta) x'^2 + 2(D \cos\ \theta - E \sin\ \theta) x' +$$

$$(A \cos^2\ \theta + 2B \sin\ \theta \cos\ \theta + C \cos^2\ \theta) y'^2 + 2(D \sin\ \theta + E \cos\ \theta) y' + F.$$

With further trigonometric manipulation,

$$\sin\ 2\theta = \frac{2B}{\sqrt{A^2 + 4B^2 + C^2 - 2AC}}, \quad (44)$$

$$\cos\ 2\theta = \frac{C - A}{\sqrt{A^2 + 4B^2 + C^2 - 2AC}}, \quad (45)$$

$$(46)$$

from which the half-angle formulas lead to

$$\sin\ \theta = \sqrt{\frac{1 - \cos\ 2\theta}{2}} \quad (47)$$

$$\cos\ \theta = \sqrt{\frac{1 + \cos\ 2\theta}{2}}. \quad (48)$$

$$(49)$$

The second step is to complete the squares, using the formula

$$\alpha x^2 + \beta x = \alpha (x + \frac{\beta}{2\alpha})^2 - (\frac{\beta}{2\alpha})^2,$$

to get

$$A' x'^2 + C' y'^2 + F' = 0.$$

Then, we might have

$$(ZU)^T = [\ \xi \quad \eta \quad \zeta\ ]$$

with

$$Z^T Q Z = \Lambda.$$

Q has not been diagonalized because diagonalization is a transformation of the type $Z^{-1}QZ$ rather than Z^TQZ; but the translation which completes the square is not unitary, as would be necessary for the two transformations to coincide.

The two remaining diagonal elements of Λ may have the same sign, or opposite signs, according to whether the critical point is extremal or a saddle point. Accordingly coefficients of their eigenvectors could be chosen according to one of the following two forms,

$$(ZU)^T = \begin{bmatrix} 0 & \cos\ t & \sin\ t \end{bmatrix},$$
$$(ZU)^T = \begin{bmatrix} 0 & \cosh\ t & \sinh\ t \end{bmatrix},$$

depending upon whether the eigenvalues agree or disagree in sign. Returning to the original coordinate system gives contour lines uniformly parameterized by t.

It still remains to ascertain how much of the conic section lies within the simplex that will be contoured; one way to avoid the problem would be to graph it through some parameter range, with a visibility function to determine whether the segments to be graphed lie within the chosen simplex or not.

Another important point concerns the location of the six data points needed for quadratic interpolation; one way to obtain them is to choose two consecutive squares from the coordinate grid, and another would be to calculate supplemental points within the square iteslf. Yet another solution would be to work with a triangular or hexagonal grid. Such alternatives tend to be incompatible with normal data acquisition techniques, but might be feasible whenever data could be generated on demand at specified coordinates.

8 Change of origin, scale

It would appear to be necessary to invert the Vandermonde matrix for each grid square; however, the use of a uniform grid results in so much symmetry that the same Vandermonde matrix can be used throughout the grid. The reason can be seen by returning to the Vandermonde matrix for quadratic interpolation, with a translated origin:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 + a & x_2 + a & x_3 + a & x + a \\ {x_1}^2 + 2ax_1 + a^2 & {x_2}^2 + 2ax_2 + a^2 & {x_3}^2 + 2ax_3 + a^2 & x^2 + 2ax + a^2 \\ y_1 & y_2 & y_3 & y \end{vmatrix} = 0.$$

By straightforward determinantal algebra it can be seen that this determinant is the same as the untranslated determinant, but factoring the determinant is more informative.

$$V = \begin{vmatrix} 1 & 0 & 0 & 0 \\ a & 1 & 0 & 0 \\ a^2 & 2a & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x \\ {x_1}^2 & {x_2}^2 & {x_3}^2 & x^2 \\ y_1 & y_2 & y_3 & y \end{vmatrix}.$$

The determinant on the left, having the value 1, is nonsingular; translation preserves the validity of the interpolation equation even as it rearranges its coefficients. Other changes of basis should be handled the same way, by factoring out the responsible matrix. Nor are admissible changes confined solely to the independent variables; for example the zero point for the the dependent variable could be changed. Not often do we find transformations which mix the dependent variable with the independent variables, although they are just as admissible as any others within the framework of the theory.

Although it is not performed as frequently, changing scale follows the same process as translating the origin. We need the matrix

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ sx_1 & sx_2 & sx_3 & sx \\ (sx_1)^2 & (sx_2)^2 & (sx_3)^2 & (sx)^2 \\ y_1 & y_2 & y_3 & y \end{vmatrix} = 0.$$

This time the factorization is much easier.

$$V = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & s & 0 & 0 \\ 0 & 0 & s^2 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x \\ {x_1}^2 & {x_2}^2 & {x_3}^2 & x^2 \\ y_1 & y_2 & y_3 & y \end{vmatrix}.$$

As a general proposition the expansion of a polynomial according to a prescribed basis with determined coefficients can be considered to be the inner product of two vectors. The row, say, carries the coefficients as coordinates while the column is outfitted with the members of the basis. Accordingly, whenever a change of basis is made, such as referring the set of points to a new origin, the inverse matrix of the change can be applied to the coefficients to get a new expansion formula.

The case of translation, which we are examining, has an inverse matrix of typical form

$$T^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -a & 1 & 0 & 0 \\ a^2 & -2a & 1 & 0 \\ -a^3 & 3a^2 & -3a & 1 \end{bmatrix}, \tag{50}$$

$-a$ appearing in place of a because it represents a translation in the opposite direction.

Starting from Maclaurin's series

$$p(x) = \sum_{i=0}^{n} \frac{p^{(i)}(0)}{i!} x^i \tag{51}$$

we could get to the Taylor's series

$$p(x) = \sum_{i=0}^{n} \frac{p^{(i)}(a)}{i!} (x-a)^i \tag{52}$$

(bearing in mind that here superscript powers signify derivatives, not Newton products) by consulting the pair of transformations

$$\begin{bmatrix} 1 \\ x-a \\ (x-a)^2 \\ (x-a)^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -a & 1 & 0 & 0 \\ a^2 & -2a & 1 & 0 \\ -a^3 & 3a^2 & -3a & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} \tag{53}$$

$$\begin{bmatrix} p(a) \\ p'(a) \\ \frac{1}{2}p''(a) \\ \frac{1}{6}p'''(a) \end{bmatrix} = \begin{bmatrix} 1 & a & a^2 & a^3 \\ 0 & 1 & 2a & 3a^2 \\ 0 & 0 & 1 & 3a \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p(0) \\ p'(0) \\ \frac{1}{2}p''(0) \\ \frac{1}{6}p'''(0) \end{bmatrix} \qquad (54)$$

9 Standardized interpolation formulas

The formulas for quadratic interpolation imply a passage from a linear form in quadratic variables to a quadratic form in linear variables, whose behavior under a change of basis ought to be analyzed. Although this change of emphasis has its merit, not all of the symmetry inherent in the determinantal version is immediately evident in the new form.

In this one dimensional example, we would write

$$z = U^T Q U$$

with

$$U^T = \begin{bmatrix} 1 & x \end{bmatrix}$$

and

$$Q = \begin{bmatrix} C & B \\ B & A \end{bmatrix}.$$

The translation Z,

$$Z = \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix}.$$

transforms Q into $Z^T Q Z$, which is namely

$$\begin{aligned} Z^T Q Z &= \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C & B \\ B & A \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix} && (55) \\ &= \begin{bmatrix} C + 2aB + a^2 A & B + aA \\ B + aA & A \end{bmatrix}. && (56) \end{aligned}$$

The choice of $a = -B/A$ removes the off-diagonal elements from $Z^T Q Z$, providing a matricial version of completing the square.

Since the interpolation equations are the same for all grid sites, and since the scale and origin of the equations can be changed so easily, it is reasonable to tabulate the interpolation coefficients for different coordinate grids, particularly for the commonly used square grid.

- two dimensional linear interpolation

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & x \\ 0 & 0 & 1 & y \\ z_{0,0} & z_{0,1} & z_{1,1} & z \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$z = z_{0,0} + (z_{0,1} - z_{0,0})x + (z_{1,1} - z_{0,1})y$$

- two dimensional quadratic interpolation

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & x \\ 0 & 1 & 0 & 1 & 0 & 1 & y \\ 0 & 1 & 4 & 0 & 1 & 4 & x^2 \\ 0 & 1 & 0 & 0 & 0 & 2 & xy \\ 0 & 1 & 0 & 1 & 0 & 1 & y^2 \\ z_{0,0} & z_{0,1} & z_{0,2} & z_{1,0} & z_{1,1} & z_{1,2} & z \end{bmatrix}$$

$$M^{-1} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

- three dimensional linear interpolation
- three dimensional quadratic interpolation

10 Three dimensional linear approximation

The representation of functions of three variables through contours in a space of three dimensions follows the same lines that have already been discussed, once the dimension of the pertinent sets of basis monomials has been taken into account. Thus four monomials, $\{1, x, y, z\}$ are required in three dimensions, with the basic equation taking the form of a 5×5 Vandermonde determinant

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 & x \\ y_1 & y_2 & y_3 & y_4 & y \\ z_1 & z_2 & z_3 & z_4 & z \\ w_1 & w_2 & w_3 & w_4 & w \end{vmatrix} = 0.$$

Not only does every increase in dimension make a symbolic solution of this equation less tractable, the problems of graphical representation become increasingly more difficult. If we would represent segments of two dimensional contours by line segments, we should represent segments of three dimensional contours by fragments of planes, which is a much more complicated process; we would probably represent such a fragment by a set of lines, such as those forming the border where it intersects the coordinate grid.

As always, the question of where the data points lie with respect to the coordinate grid arises; not all four points should lie on the same face of a cube, to avoid linear dependencies in the Vandermonde matrix.

The coordinate cubes have eight vertices, which is double the required number of points. One solution is to decompose the cube into six tetrahedra just as the coordinate square was decomposed into two triangles in two dimensions. Even though the work is multiplied by six, the greater part of the classification into cases according to how the contour surfaces cross the edges and faces of the cube is eliminated.

Generally the intersection of a plane with a tetrahedron is a triangle, but it has to be remembered that quadrilateral sections are also possible—for example when the interecting plane is parallel or nearly parallel to a pair of opposite edges of the tetrahedron.

Other approaches are possible if it is not required that the contour planes be represented by lines passing through the grid lines. For example, a circle could be traced out tangent to the grid lines, so that its counterimage on the contour plane could be plotted. The result would be a network of tangent ellipses which might cover the contour surface fairly well.

Determining the parameters of circles and calculating their points—even by table lookup—would consume additional time; a simpler approach would to be to trace the counterprojection of the edge-dual network formed from the coordinate grid. This network would consist of lines joining the midpoints of the edges of the faces of the coordinate cubes.

Of course this discussion makes no mention at all of visibility. Nothing need be said if a simple wire frame model is all that is desired. But, if the surface is assumed to be opaque or semitransparent, it is inevitable that some parts of it would be obscured by others. Then, further calculations would be required to decide whether, or how much, of the surface patches would be visible.

11 Three dimensional quadratic approximation

The quadratic approximation in three dimensions involves ten basis monomials, and thus an 11×11 Vandermonde determinant.

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & x \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y \\ z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z_7 & z_8 & z_9 & z_{10} & z \\ {x_1}^2 & {x_2}^2 & {x_3}^2 & {x_4}^2 & {x_5}^2 & {x_6}^2 & {x_7}^2 & {x_8}^2 & {x_9}^2 & {x_{10}}^2 & x^2 \\ x_1y_1 & x_2y_2 & x_3y_3 & x_4y_4 & x_5y_5 & x_6y_6 & x_7y_7 & x_8y_8 & x_9y_9 & x_{10}y_{10} & xy \\ x_1z_1 & x_2z_2 & x_3z_3 & x_4z_4 & x_5z_5 & x_6z_6 & x_7z_7 & x_8z_8 & x_9z_9 & x_{10}z_{10} & xz \\ {y_1}^2 & {y_2}^2 & {y_3}^2 & {y_4}^2 & {y_5}^2 & {y_6}^2 & {y_7}^2 & {y_8}^2 & {y_9}^2 & {y_{10}}^2 & y^2 \\ y_1z_1 & y_2z_2 & y_3z_3 & y_4z_4 & y_5z_5 & y_6z_6 & y_7z_7 & y_8z_8 & y_9z_9 & y_{10}z_{10} & yz \\ {z_1}^2 & {z_2}^2 & {z_3}^2 & {z_4}^2 & {z_5}^2 & {z_6}^2 & {z_7}^2 & {z_8}^2 & {z_9}^2 & {z_{10}}^2 & z^2 \\ w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 & w_9 & w_{10} & w \end{vmatrix} = 0.$$

The quadratic form Q now requires a 4×4 matrix

$$Q = \begin{bmatrix} J & I & H & G \\ I & F & E & D \\ H & E & C & B \\ G & D & B & A \end{bmatrix}.$$

If it is given the same treatment as the matrix for two dimensional quadratic approximations, we can make one of its eigenvalues vanish, leaving three others to contend with. There will be many more sign combinations to contend with, but in principle a choice such as

$$(ZU)^T = \left[\begin{array}{cccc} 0 & \alpha t & \cos\, t & \sin\, t \end{array} \right]$$

will result in a spiral running over the surface of whatever sphere or hyperboloid corresponds to the diagonal form of the matrix, and thus a means of visualizing the contour surface.

Since the coordinate cubes have only eight vertices, two additional data points have to be found to make up the required ten data points. The most symmetric points in the coordinate grid would be a pair of opposite points lying on the extension of one of the body diagonals of the cube.

12 Fitting by least squares

As is often the case, a contour plot may not be judged so much for its accuracy as for its neatness. For example, critical points could be expected to be relatively rare and so it might be thought that the crudeness of the linear approximation in the vicinity of a critical point would go unnoticed. However, exactly the opposite is true, to the extent that contour maps are often prepared for the sole purpose of locating the critical points, leaving the remainder of the diagram as the part likely to be ignored.

It is also true that a least squares fitting of a linear or quadratic approximation obtained by combining the data points of several simplices may sometimes be preferred to the more detailed representation derived from the individual simplices. It is then necessary to construct the normal equations from the data, and invert their coefficient matrix, to obtain the parameters to represent the contour, with the added complication of having to join the contours smoothly as they pass from one region to another. Moreover, the coefficient matrix would actually have to be inverted for each region, which can be avoided when interpolating through the individual simplices.

13 The sequence of triangles

Once suitable interpolation formulas have been deduced to obtain the contour lines in <PLOT>, attention must be given to some practical aspects of the contouring programs. When the contour maps were prepared with pen plotters, the overriding consideration was the extremely long time needed to raise and lower the pen, relative to the time required for horizontal or vertical movements. Thus it was advantageous to trace out an entire contour line, until it either closed on itself or reached the boundary of the contoured region. Electrostatic plotters or displays on video monitors do not impose such a penalty, so that all the contours desired for a given triangle can be computed and displayed before moving on to the next triangle.

Since the completion of one contour before commencing another could lead the pen through triangles that would eventually be scanned as possible starting points, a bitmap was constructed in which a flag indicates whether a triangle has previously been traversed or not. The same information prevents retracing a closed contour once the pen has returned to its starting point. Much space can be saved by programming the bitmap in assembly language, but confining <PLOT> exclusively to a high level language such as Fortran or "C" encourages wasting an entire word of computer memory per bit.

The use of simplices instead of the coordinate grid leads to several special cases for the interpolation formulas. By permuting the vertices of the simplices, a single subroutine for the interpolation suffices. Passage from one the vertices of one simplex to the next can be accomplished by reflection in the face of the simplex, once again leading to a uniform formula for the process. It is unfortunate that space cannot be filled by any kind of uniform simplices.

14 Checklist

- this is now a third draft
- program 2D quadratic interpolation
- program 3D linear interpolation
- program 3D linear interpolation extension
- program 3D quadratic interpolation
- program least square fitting
- think about embedding contour lines on hidden surface drawings

References

[1] Milton Abramowitz snd Irene A. Stegun (editors), *Handbook of Mathematical Functions*, U. S. Government Printing Office, Washington, D. C., 1964.

[2] Wolfgang K. Giloi, *Interactive Computer Graphics*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978 (ISBN 0-13-469189-X).

[3] D. C. Sutcliffe, "Contouring over Rectangular and Skewed Rectangular Grids - an Introduction," *Mathematical Methods in Computer Graphics and Design*, (K. W. Brodlie, Editor), Academic Press, Inc, New York, 1980 (ISBN 0-12-134880-0).

January 14, 2001