NASA
Technical
Paper
2708

June 1987

An Algorithm for Surface Smoothing With Rational Splines

James R. Schiess

(NASA-TP-2708) AN ALGORITHM FOR SURFACE SMOOTHING WITH RATIONAL SPLINES (NASA) 17 p Avail: NTIS HC A02/MF A01 CSCL 12A

N87-22447

Unclas

H1/64 0074199

NASA

NASA
Technical
Paper
2708

1987

An Algorithm for Surface Smoothing With Rational Splines

James R. Schiess

Langley Research Center
Hampton, Virginia

National Aeronautics
and Space Administration

**Scientific and Technical
Information Office**

Introduction

The evolution of high-speed computers and sophisticated display devices has encouraged the development of advanced algorithms for manipulating and displaying multidimensional data. In particular, the area of computer-aided geometric modeling (CAGM) has advanced significantly in recent years. In CAGM, computational geometry and computer graphics are combined to give mathematical and graphical representations of curves, surfaces, and volumes. A wide variety of applications may be found in the mathematical representation of physical phenomena, such as meteorological data, and in the design of aircraft.

Often it is desirable to smooth three-dimensional surface data consisting of two independent variables (x and y) and one dependent variable which contains random noise because of errors in measurement or calibration. The purpose of smoothing the surface is to obtain statistically representative values of the dependent variable. One typical approach to smoothing surfaces with splines is to first smooth along rows of data and then smooth along columns of data (ref. 1). If, however, the data are not aligned in rows and columns, then additional procedures must be applied, such as the use of a triangularization method to interpolate to a rectangular grid before smoothing. A difficulty that can arise with either method is that changes in trends in the data can induce undesirable oscillations in the smoothed spline surface.

One way to reduce or eliminate the oscillations in the spline surface is to apply tension to the surface. Applying mathematical tension to a spline surface is analogous to grasping the opposite edges of a membrane and stretching the membrane to remove wrinkles. In reference 2, Späth developed the rational spline for both curve and surface interpolation. The rational spline is a cubic function which has tension parameters in the denominators of the cubic terms. More recently, Frost and Kinzel (ref. 3) developed an algorithm for automatically adjusting the tension parameters in curve-interpolating rational splines. The Frost and Kinzel method was combined with constrained least squares by Schiess and Kerr (ref. 4) to give an algorithm for rational-spline smoothing of curves. Finally, Schiess (ref. 5) developed two algorithms for rational-spline interpolation of surface data given on a rectangular grid.

The present paper presents an algorithm for smoothing surface data with bivariate rational splines having multiple tension parameters. The multiple tension parameters allow for local control of tension on the smoothing surface. Equations are derived to ensure continuity of the derivatives at the knots. Smoothing is accomplished by finding the weighted least-squares estimates of the bivariate rational-spline coefficients. The capabilities of the rational-spline smoothing algorithm are demonstrated on terrain elevation data.

Symbols

$\mathbf{A}_{ij}$	4 by 4 matrix of coefficients
a_{ijkl}	coefficients of multivariate rational spline in the subregion R_{ij} $(k, l = 1, 2, 3, 4)$
CX_i	coefficients in equation for $\overline{FX}_{ij}$
CY_j	coefficients in equation for $\overline{FY}_{ij}$
c_{ik}	coefficients in univariate rational spline on interval i $(k = 1, 2, 3, 4)$
$\mathbf{D}$	$2MN$ by $4MN$ matrix of constraint cofactors
dx_i	difference between consecutive values of x-coordinate of knot, $\bar{x}_{i+1} - \bar{x}_i$
dy_j	difference between consecutive values of y-coordinate of knot, $\bar{y}_{j+1} - \bar{y}_j$
$\mathbf{E}$	m by $4MN$ matrix of rational-spline cofactors
F_r	function value at the data point (x_r, y_r)
$\overline{F}_{ij}$	function value at $(\bar{x}_i, \bar{y}_j)$
$\overline{FX}_{ij}$	partial derivative of $f_{ij}(x, y)$ with respect to x evaluated at $(\bar{x}_i, \bar{y}_j)$
$\overline{FY}_{ij}$	partial derivative of $f_{ij}(x, y)$ with respect to y evaluated at $(\bar{x}_i, \bar{y}_j)$
$\overline{FXY}_{ij}$	partial derivative of $f_{ij}(x, y)$ with respect to x and y evaluated at $(\bar{x}_i, \bar{y}_j)$
$f_{ij}(x, y)$	rational spline on the subregion R_{ij}
$\mathbf{G}_i$	4 by 4 matrix of functions $g_{ik}(x)$
$g_{ik}(x)$	functions of x used in rational-spline representation on interval i $(k = 1, 2, 3, 4)$
$\mathbf{H}_j$	4 by 4 matrix of functions $h_{jl}(y)$
$h_{jl}(y)$	functions of y used in rational-spline representation on interval j $(l = 1, 2, 3, 4)$
$\mathbf{L}$	$2MN$-element column vector of Lagrange multipliers
M	number of knots along x-axis

m	number of data points
N	number of knots along y-axis
p_i	tension factor for interval i
q_j	tension factor for interval j
R_{ij}	rectangular subregion in xy-plane defined by $\bar{x}_i \leq x \leq \bar{x}_{i+1}$ and $\bar{y}_j \leq y \leq \bar{y}_{j+1}$
$\mathbf{S}_{ij}$	4 by 4 matrix of function and derivative values
$\mathbf{s}_{ij}$	16-element column vector of unknowns
r, s, t, u	variables used to define rational spline
$\mathbf{v}_i(x)$	4-element column vector of functions of x
$\mathbf{vw}_{ij}$	16-element vector of functions of x and y
V	estimated variance of measurement error
$\mathbf{W}$	m by m diagonal weighting matrix
$\mathbf{w}_j(y)$	4-element column vector of functions of y
x, y	independent variables
$\mathbf{Z}$	m-element column vector of function values F_r
$\bar{\mathbf{Z}}$	$4MN$-element column vector of unknown parameters
z	dependent variable in univariate rational spline
Δ_x, Δ_y	differences with respect to x and y variables

Subscripts:

i, j	quantity at the knot $(\bar{x}_i, \bar{y}_j)$
k, l	general indices
r	data point index

Superscripts:

T	matrix transpose
-1	matrix inverse

A prime indicates first derivative with respect to the independent variable. A double prime indicates second derivative with respect to the independent variable. A bar over a quantity denotes that quantity at a knot.

Problem Statement

Let the three-dimensional data $(x_r,\ y_r,\ F_r)$ be given, where $r = 1, 2, ..., m$. The variables x and y are the independent variables. The paired values $(x_r,\ y_r)$ lie in a finite, bounded region of the xy-plane but do not need to be uniformly scattered in the region or lying on grid lines. The values F_r of the dependent variable represent values of a function measured at the points $(x_r,\ y_r)$; the measured values are assumed to be corrupted by random error of unknown statistical characteristics.

The objective is to fit, in the weighted least-squares sense, a bivariate rational spline to the data. This requires the selection of M knot locations along the x-axis and N knot locations along the y-axis so that $\bar{x}_1 < \bar{x}_2 < ... < \bar{x}_M$ and $\bar{y}_1 < \bar{y}_2 < ... < \bar{y}_N$. The knot locations need not be equally spaced. All the data must lie in the region defined by $\{\bar{x}_1 \leq x \leq \bar{x}_M,\ \bar{y}_1 \leq y \leq \bar{y}_N\}$, and several data points should lie in each subregion $R_{ij} = \{\bar{x}_1 \leq x \leq \bar{x}_{i+1},\ \bar{y}_j \leq y \leq \bar{y}_{j+1}\}$.

Rational Splines

In this section both the univariate and bivariate rational splines are described. Because the bivariate rational spline is the tensor product of two univariate rational splines, the characteristics of the univariate rational spline are discussed first.

Univariate Rational Spline

Let $\bar{x}_i$ be the abscissas of the knots of the univariate rational spline, where $i = 1, 2, ..., M$ and $\bar{x}_1 < x_2 < ... < \bar{x}_M$, and let z be the value of the spline at x. The rational spline on interval i $(i = 1, 2, ..., M-1)$ is defined in references 2 and 3 to be

$$z = \sum_{k=1}^{4} c_{ik}\ g_{ik}(x) \qquad (\bar{x}_i \leq x \leq \bar{x}_{i+1}) \qquad (1)$$

where c_{ik} are unknown coefficients,

$$g_{i1}(x) = u \qquad g_{i3}(x) = \frac{u^3}{p_i t + 1}$$

$$g_{i2}(x) = t \qquad g_{i4}(x) = \frac{t^3}{p_i u + 1}$$

where

$$u = \frac{\bar{x}_{i+1} - x}{dx_i}$$

$$t = \frac{x - \bar{x}_i}{dx_i} = 1 - u$$

$$dx_i = \bar{x}_{i+1} - \bar{x}_i$$

and p_i is the tension parameter for interval i.

Equation (1) is defined for all values of the independent variable x in the data range if the tension parameter p_i is restricted to $p_i > -1$. If p_i is set to zero, equation (1) reduces to a cubic-spline function. As p_i increases from zero, the cubic terms decrease in magnitude and the function tends to the equation of the line joining the knots at $\bar{x}_i$ and $\bar{x}_{i+1}$. Because a distinct, independent tension parameter is associated with each interval, the behavior of the function in each interval may be locally controlled.

Evaluation of equation (1) for each subinterval requires knowledge of the four coefficients c_{i1}, c_{i2}, c_{i3}, and c_{i4}. Thus, for M data points (equivalently, for $M - 1$ subintervals), $4M - 4$ coefficients must be determined. Späth (ref. 2) reduces the magnitude of this problem by writing the coefficients in terms of the values of the function and its first derivative at the knots. End conditions are applied to the first derivative, and equations ensuring the continuity of the second derivative at the interior knots are derived. For the interpolation problem, this derivation yields a system of $M - 2$ equations for the $M - 2$ unknown interior first derivatives. Frost and Kinzel (ref. 3) extend Späth's approach by allowing for three different end conditions and by developing an iterative method for determining the tension parameters. Tension parameters are found so that the interpolating rational spline deviates from the line joining knots by a prescribed value.

Another approach to determining a smooth fit to the data has been presented by Schiess and Kerr (ref. 4) in deriving a least-squares univariate rational-spline approximation. The rational spline is reformulated in terms of the unknown spline function and its second derivative at the knots. Smoothness is ensured by imposing the constraints that the first derivatives are continuous at the interior knots. This approach leads to a constrained least-squares problem in the $2M$ values of the unknown function and its second derivative at the knots.

Bivariate Rational Spline

Let a given set of M by N knot points in three dimensions be represented by ($\bar{x}_i$, $\bar{y}_j$, $\bar{F}_{ij}$), where $i = 1, 2, ..., M$ and $j = 1, 2, ..., N$. The independent variables are assumed to be ordered ($\bar{x}_1 < \bar{x}_2 < ... < x_M$ and $\bar{y}_1 < \bar{y}_2 < ... < \bar{y}_N$) and form a rectangular grid, but are not necessarily equally spaced.

The multiple-tension-parameter bivariate rational spline on the subregion R_{ij} defined by $\bar{x}_i \leq x \leq \bar{x}_{i+1}$ ($i = 1, 2, ..., M - 1$) and $\bar{y}_j \leq y < \bar{y}_{j+1}$ ($j = 1, 2, ..., N - 1$) is defined in reference 2 by

$$f_{ij}(x, y) = \sum_{k=1}^{4} \sum_{l=1}^{4} a_{ijkl}\, g_{ik}(x)\, h_{jl}(y) \qquad (2)$$

where $g_{ik}(x)$ and p_i are the same as for the univariate spline, a_{ijkl} are unknown coefficients,

$$h_{j1}(y) = s \qquad h_{j3}(y) = \frac{s^3}{q_j r + 1}$$

$$h_{j2}(y) = r \qquad h_{j4}(y) = \frac{r^3}{q_j s + 1}$$

where

$$s = \frac{\bar{y}_{j+1} - y}{dy_j}$$

$$r = \frac{y - \bar{y}_j}{dy_j} = 1 - s$$

$$dy_j = \bar{y}_{j+1} - \bar{y}_j$$

and q_j is the tension parameter for interval j.

As defined by equation (2), the bivariate rational spline on each subregion R_{ij} is a function of 2 tension parameters (p_i and q_j) and 16 coefficients (a_{ijkl}). The coefficients are to be determined so that the rational spline and its first and second derivatives are continuous over the entire region. The $M + N - 2$ tension parameters may be adjusted individually; each parameter affects the behavior of the rational spline in a strip parallel to either the x-axis (for q_j) or the y-axis (for p_i).

Since 16 coefficients are needed on each subregion, a total of 16 $(M - 1)(N - 1)$ coefficients must be determined to define the entire rational spline. For example, for a 30 by 30 grid ($M = N = 30$), a total of 13 456 coefficients are needed. Späth (ref. 2) reduces the actual number of unknown quantities by writing the coefficients as linear combinations of the values of the function and its derivatives at the grid points. A similar approach is taken in this paper.

Let $\overline{FX}_{ij}$ and $\overline{FY}_{ij}$ be the first derivatives of $f_{ij}(x, y)$ with respect to x and y, respectively, and $\overline{FXY}_{ij}$ be the cross derivatives, all evaluated at the point $(\bar{x}_i, \bar{y}_j)$. For the subregion R_{ij}, define the 4 by 4 matrices

$$\mathrm{S}_{ij} = \begin{bmatrix} \overline{F}_{ij} & \overline{FY}_{ij} & \overline{F}_{i(j+1)} & \overline{FY}_{i(j+1)} \\ \overline{FX}_{ij} & \overline{FXY}_{ij} & \overline{FX}_{i(j+1)} & \overline{FXY}_{i(j+1)} \\ \overline{F}_{(i+1)j} & \overline{FY}_{(i+1)j} & \overline{F}_{(i+1)(j+1)} & \overline{FY}_{(i+1)(j+1)} \\ \overline{FX}_{(i+1)j} & \overline{FXY}_{(i+1)j} & \overline{FX}_{(i+1)(j+1)} & \overline{FXY}_{(i+1)(j+1)} \end{bmatrix}$$

$$\mathbf{A}_{ij} = \begin{bmatrix} a_{ij11} & a_{ij12} & a_{ij13} & a_{ij14} \\ a_{ij21} & a_{ij22} & a_{ij23} & a_{ij24} \\ a_{ij31} & a_{ij32} & a_{ij33} & a_{ij34} \\ a_{ij41} & a_{ij42} & a_{ij43} & a_{ij44} \end{bmatrix}$$

$$\mathbf{G}_i = \begin{bmatrix} g_{i1}(\bar{x}_i) & g_{i2}(\bar{x}_i) & g_{i3}(\bar{x}_i) & g_{i4}(\bar{x}_i) \\ g'_{i1}(\bar{x}_i) & g'_{i2}(\bar{x}_i) & g'_{i3}(\bar{x}_i) & g'_{i4}(\bar{x}_i) \\ g_{i1}(\bar{x}_{i+1}) & g_{i2}(\bar{x}_{i+1}) & g_{i3}(\bar{x}_{i+1}) & g_{i4}(\bar{x}_{i+1}) \\ g'_{i1}(\bar{x}_{i+1}) & g'_{i2}(\bar{x}_{i+1}) & g'_{i3}(\bar{x}_{i+1}) & g'_{i4}(\bar{x}_{i+1}) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 1 & 0 \\ -\frac{1}{dx_i} & \frac{1}{dx_i} & -\frac{3+p_i}{dx_i} & 0 \\ 0 & 1 & 0 & 1 \\ -\frac{1}{dx_i} & \frac{1}{dx_i} & 0 & \frac{3+p_i}{dx_i} \end{bmatrix}$$

$$\mathbf{H}_j = \begin{bmatrix} h_{j1}(\bar{y}_j) & h_{j2}(\bar{y}_j) & h_{j3}(\bar{y}_j) & h_{j4}(\bar{y}_j) \\ h'_{j1}(\bar{y}_j) & h'_{j2}(\bar{y}_j) & h'_{j3}(\bar{y}_j) & h'_{j4}(\bar{y}_j) \\ h_{j1}(\bar{y}_{j+1}) & h_{j2}(\bar{y}_{j+1}) & h_{j3}(\bar{y}_{j+1}) & h_{j4}(\bar{y}_{j+1}) \\ h'_{j1}(\bar{y}_{j+1}) & h'_{j2}(\bar{y}_{j+1}) & h'_{j3}(\bar{y}_{j+1}) & h'_{j4}(\bar{y}_{j+1}) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 1 & 0 \\ -\frac{1}{dy_j} & \frac{1}{dy_j} & -\frac{3+q_j}{dy_j} & 0 \\ 0 & 1 & 0 & 1 \\ -\frac{1}{dy_j} & \frac{1}{dy_j} & 0 & \frac{3+q_j}{dy_j} \end{bmatrix}$$

Thus, in matrix notation,

$$\mathbf{S}_{ij} = \mathbf{G}_i \mathbf{A}_{ij} \mathbf{H}_j^T \tag{3}$$

Equation (3) can be verified by differentiating equation (2), as appropriate, and evaluating the results at the corners of the subregion. Note that $\mathbf{G}_i$ and $\mathbf{H}_j$ depend on the grid spacing and tension parameters but not on the function values.

Since the matrices $\mathbf{G}_i$ and $\mathbf{H}_j$ are nonsingular, equation (3) can be solved for the matrix of coefficients:

$$\mathbf{A}_{ij} = \mathbf{G}_i^{-1} \mathbf{S}_{ij} (\mathbf{H}_j^T)^{-1} \tag{4}$$

Therefore, for any subregion the 16 coefficients can be determined from the values of the function, its first derivatives with respect to x and y, and its cross derivatives at the four corners of the subregion. Therefore, a total of $4MN$ function and derivative values are needed to calculate the coefficients.

Rational-Spline Smoothing

In this section an algorithm for finding the surface-smoothing rational spline in terms of the function and its derivatives is presented. This algorithm is for the general case of $M-1$ values of the tension parameters p_i $(i = 1, 2, ..., M-1)$ and $N-1$ values of the tension parameters q_j $(j = 1, 2, ..., N-1)$. An algorithm for the single-tension-parameter rational spline is not presented because that rational spline is a special case of the multiple-parameter spline.

Vector Formulation

Let the knot locations $(\bar{x}_i, \bar{y}_j)$ and tension parameters p_i and q_j (for $i = 1, 2, ..., M$ and $j = 1, 2, ..., N$) be given. The knots do not need to be equally spaced, but they must form a rectangular grid. Let the data $(x_r,\ y_r,\ F_r)$ for $r = 1, 2, ..., m$ also be given. Least-squares estimation of the $4MN$ unknown function and derivative values requires that there be more data points than unknowns, or that $m > 4MN$. Further, there should be several data points in each region R_{ij}.

In order to find least-squares estimates of $\overline{F}_{ij}$, $\overline{FX}_{ij}$, $\overline{FY}_{ij}$, and $\overline{FXY}_{ij}$, it is necessary to reformulate the problem. Using equation (4) in equation (2) results in the rational spline being written as

$$f_{ij}(x, y) = \mathbf{v}_i^T(x)\ \mathbf{S}_{ij} \mathbf{w}_j(y) \tag{5}$$

where

$$\mathbf{v}_i(x) = (\mathbf{G}_i^{-1})^T\ [g_{i1}(x),\ g_{i2}(x),\ g_{i3}(x),\ g_{i4}(x)]^T \tag{6}$$

and

$$\mathbf{w}_j(y) = (\mathbf{H}_j^{-1})^T\ [h_{j1}(y),\ h_{j2}(y),\ h_{j3}(y),\ h_{j4}(y)]^T \tag{7}$$

In terms of individual entries in $\mathbf{v}_i(x)$, $\mathbf{w}_j(y)$, and $\mathbf{S}_{ij}$, equation (5) can be written as

$$f_{ij}(x,y) = \sum_{k=1}^{4}\sum_{l=1}^{4} \mathbf{v}_{ik}(x)\ \mathbf{S}_{ijkl}\ \mathbf{w}_{jl}(y)$$
$$= \sum_{k=1}^{4}\sum_{l=1}^{4} \mathbf{v}_{ik}(x)\ \mathbf{w}_{jl}(y)\ \mathbf{S}_{ijkl} \qquad (8)$$

or

$$f_{ij}(x,y) = (\mathbf{vw}_{ij})^T\ \mathbf{s}_{ij}$$

with the two 16-element column vectors defined by

$$\mathbf{vw}_{ij} = [v_{i1}(x)w_{j1}(y),\ v_{i1}(x)w_{j2}(y),\ v_{i1}(x)\ w_{j3}(y), ..., v_{i4}(x)w_{j4}(y)]^T$$

and

$$\mathbf{s}_{ij} = [S_{ij11},\ S_{ij12},\ S_{ij13},\ S_{ij14},\ S_{ij21}, ...,\ S_{ij44}]^T$$

As presented in equation (8), $f_{ij}(x,y)$ is a linear function of the vector $\mathbf{s}_{ij}$ of unknown function and derivative values at the knots and therefore is amenable to least-squares estimation.

One minor modification which is made to simplify the use of this formulation is to rearrange the entries in $\mathbf{s}_{ij}$ (and corresponding entries in $\mathbf{vw}_{ij}$) so that all the unknowns corresponding to one knot are in consecutive entries. The rearrangement is chosen so that the entries are in the order $\overline{F}_{ij}$, $\overline{FX}_{ij}$, $\overline{FY}_{ij}$, and $\overline{FXY}_{ij}$.

Constraint Equations

In the formulation used herein, the continuity of the first derivatives and cross derivatives at the interior knots is ensured because these derivatives are parameters to be estimated in the rational-spline formulation. It is also desirable that the second derivatives be continuous at the interior knots. Furthermore, for the knots on the boundaries, natural spline conditions can be applied (i.e., the second derivatives are zero on the boundary). Both of these conditions can be added to the least-squares formulation as constraint equations which are linear in the unknown quantities.

For the continuity of the second derivatives at the interior knots, the equations derived in reference 5 are used. Equations (9) and (10) of this report are the same as equations (5) and (6) in reference 5 (with the current notation and some rearrangement). Constraining the second derivatives with respect to x to be continuous at the interior knots results in the $(M-2)N$ equations

$$CX_{i-1}\ \overline{FX}_{(i-1)j} + [\ (2+p_{i-1})CX_{i-1} + (2+p_i)CX_i]\ \overline{FX}_{ij} + CX_i\ \overline{FX}_{(i+1)j} - \frac{3+p_{i-1}}{dx_{i-1}}\ CX_{i-1}\ \Delta_x\overline{F}_{(i-1)j} - \frac{3+p_i}{dx_i}CX_i\ \Delta_x\overline{F}_{ij} = 0$$
$$(i = 2,3,...,M-1;\ j = 1,2,...,N) \qquad (9)$$

where

$$CX_i = \frac{p_i^2 + 3p_i + 3}{[(2+p_i)^2 - 1]dx_i}$$

$$\Delta_x\overline{F}_{ij} = \overline{F}_{(i+1)j} - \overline{F}_{ij}$$

Similarly, continuity of the second derivative with respect to y at the interior knots yields the additional $M(N-2)$ equations

$$CY_{j-1}\ \overline{FY}_{i(j-1)} + [\ (2 + q_{j-1})CY_{j-1} + (2 + q_j)CY_j]\overline{FY}_{ij} + CY_j\overline{FY}_{i(j+1)} - \frac{3+q_{j-1}}{dy_{j-1}}\ CY_{j-1}\ \Delta_y\overline{F}_{i(j-1)} - \frac{3+q_j}{dy_j}\ CY_j\ \Delta_y\overline{F}_{ij} = 0$$
$$(i = 1,2,...,M;\ j = 2,3,...,N-1) \qquad (10)$$

where

$$CY_j = \frac{q_j^2 + 3q_j + 3}{[(2+q_j)^2 - 1]dy_j}$$

$$\Delta_y\overline{F}_{ij} = \overline{F}_{i(j+1)} - \overline{F}_{ij}$$

The natural spline boundary conditions require that the second derivatives with respect to x be zero at the boundaries, conditions which are equivalent to requiring the rational spline be linear on and outside the boundaries. These conditions result in two sets of constraint equations. Along the boundary $x = \bar{x}_1$, the equations are

$$\frac{3 + p_1}{dx_1}\overline{F}_{1j} - \frac{3 + p_1}{dx_1}\overline{F}_{2j} + (2 + p_1)\overline{FX}_{1j} + \overline{FX}_{2j} = 0 \qquad (j = 1, 2, ..., N) \tag{11}$$

Along the boundary $x = \bar{x}_M$, the equations are

$$\frac{3 + p_{M-1}}{dx_{M-1}}\overline{F}_{(M-1)j} - \frac{3 + p_{M-1}}{dx_{M-1}}\overline{F}_{Mj} + \overline{FX}_{(M-1)j} + (2 + p_{M-1})\,\overline{FX}_{Mj} = 0 \qquad (j = 1, 2, ..., N) \tag{12}$$

On the boundaries $y = \bar{y}_1$ and $y = \bar{y}_N$, the second derivative with respect to y is zero. These conditions lead to two additional sets of constraint equations. For the boundary $y = \bar{y}_1$, equation (13) applies:

$$\frac{3 + q_1}{dy_1}\overline{F}_{i1} - \frac{3 + q_1}{dy_1}\overline{F}_{i2} + (2 + q_1)\overline{FY}_{i1} + \overline{FY}_{i2} = 0 \qquad (i = 1, 2, ..., M) \tag{13}$$

Along the boundary $y = \bar{y}_N$, the constraint equations are

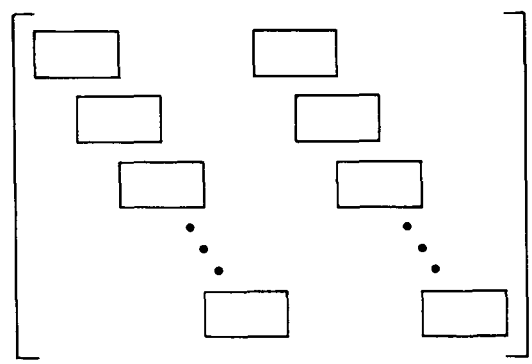
$$\frac{3 + q_{N-1}}{dy_{N-1}}\overline{F}_{i(N-1)} - \frac{3 + q_{N-1}}{dy_{N-1}}\overline{F}_{iN} + \overline{FY}_{i(N-1)} + (2 + q_{N-1})\,\overline{FY}_{iN} = 0 \qquad (i = 1, 2, ..., M) \tag{14}$$

The $(M - 2)N + M(N - 2)$ constraints at the interior knots and $2M + 2N$ boundary constraints yield a total of $2MN$ constraint equations. The derivation of the constraint equations resulting from the boundary conditions is given in the appendix. In the next section these equations are incorporated into the least-squares fit of the rational spline to the data.

Constrained Weighted Least Squares

The rational-spline surface approximation is found through solution of a constrained weighted least-squares problem. The solution to this problem is most easily seen with the following matrices defined. Let $\overline{\mathbf{Z}}$ be the $4MN$-element column vector of unknowns such that $\overline{\mathbf{Z}} = (\overline{F}_{11}, \overline{FX}_{11}, \overline{FY}_{11}, \overline{FXY}_{11}, \overline{F}_{21}, \overline{FX}_{21}, ..., \overline{FXY}_{MN})^T$. Let $\mathbf{Z}$ be the m by 1 vector of known function values F_r at the data points (x_r, y_r) $(r = 1, 2, ..., m)$ such that $\mathbf{Z} = (F_1, F_2, ..., F_m)^T$. Define the m by $4MN$ matrix $\mathbf{E}$ of cofactors $\overline{F}_{ij}$, $\overline{FX}_{ij}$, $\overline{FY}_{ij}$, and $\overline{FXY}_{ij}$ $(i = 1, 2, ..., M$ and $j = 1, 2, ..., N)$ as given in equation (8). From the definition of $\overline{\mathbf{Z}}$, the quadruplets $(\overline{F}_{ij}, \overline{FX}_{ij}, \overline{FY}_{ij}, \overline{FXY}_{ij})$ are ordered so that the subscript i changes faster than the subscript j. Equivalently, the values in $\mathbf{Z}$ are ordered so that all the values in region R_{11} are first, those in region R_{21} are second, and so forth; the ordering within regions is unimportant. With this ordering, $\mathbf{E}$ has the general block structure

$\mathbf{E} =$



where only the entries in blocks are nonzeroes. Each block is eight columns wide. With these definitions, equation (8) evaluated at (x_r, y_r, F_r) $(r = 1, 2, ..., m)$ can be written

$$\mathbf{Z} = \mathbf{E}\overline{\mathbf{Z}} \tag{15}$$

Let $\mathbf{D}$ be the $2MN$ by $4MN$ matrix of cofactors of the constraint equations (eqs. (9) to (14)). Thus, the constraint equations can be written in matrix form as

$$\mathbf{D}\overline{\mathbf{Z}} = 0 \tag{16}$$

Each row of $\mathbf{D}$ has only four (corresponding to eqs. (11) to (14)) or six (eqs. (9) and (10)) nonzero entries.

Finally, let $\mathbf{W}$ be an m by m diagonal weighting matrix containing the weights on the diagonal. Then, the constrained weighted least-squares problem requires minimization of $(\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}})^T\mathbf{W}(\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}})$ with respect to $\overline{\mathbf{Z}}$ such that $\mathbf{D}\overline{\mathbf{Z}} = 0$. With $\mathbf{L}$ defined as a $2MN$-element column vector of Lagrange multipliers, the constrained problem can be rewritten as the unconstrained problem, that is, minimization of $(\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}})^T\mathbf{W}(\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}}) + 2\mathbf{L}^T\mathbf{D}\overline{\mathbf{Z}}$ with respect to $\overline{\mathbf{Z}}$ and $\mathbf{L}$. The solutions of this minimization problem (ref. 6) are

$$\overline{\mathbf{Z}} = (\mathbf{E}^T\mathbf{W}\mathbf{E})^{-1}\,\mathbf{E}^T\mathbf{W}\mathbf{Z} - (\mathbf{E}^T\mathbf{W}\mathbf{E})^{-1}\,\mathbf{D}^T\mathbf{L} \tag{17}$$

and

$$\mathbf{L} = [\mathbf{D}(\mathbf{E}^{\mathrm{T}}\mathbf{W}\mathbf{E})^{-1}\mathbf{D}^{T}]^{-1}\ \mathbf{D}(\mathbf{E}^{\mathrm{T}}\mathbf{W}\mathbf{E})^{-1}\ \mathbf{E}^{\mathrm{T}}\mathbf{W}\mathbf{Z} \tag{18}$$

The solutions in equations (17) and (18) exist and are unique if the matrix inverses in the equations exist. Furthermore, a sufficient condition such that the solutions in these two equations yield a unique global minimum of $(\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}})^T\mathbf{W}(\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}})$ subject to the constraint $\mathbf{D}\overline{\mathbf{Z}} = 0$ is $\mathbf{Z}^T\mathbf{E}^T\mathbf{W}\mathbf{E}\mathbf{Z} > 0$ for all $\mathbf{Z} > 0$ with $\mathbf{D}\mathbf{Z} = 0$ (ref. 6).

For the user interested in an overall measure of the error of the fit to the data, the estimated variance of the measurement error is easily calculated. The estimated variance V is defined by (ref. 6)

$$V = \frac{1}{m - 4MN}\ (\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}})^T\ \mathbf{W}(\mathbf{Z} - \mathbf{E}\overline{\mathbf{Z}}) \tag{19}$$

Selection of Tension Parameters

Although an automated procedure for adjusting tension parameters is available for univariate rational splines (refs. 3 and 4), no such procedure has been devised for bivariate rational splines. Instead, it is recommended that the user visually examine three-dimensional or contour plots of the data and the rational-spline smoothed surface and then use engineering judgment to select tension-parameter values. This approach is outlined here.

First, the user should obtain a plot of the original data. This plot will show both general trends and local anomalies of the data. Second, the user should compute the smoothing bicubic-spline surface by calculating the smoothing rational spline with all the tension parameters set to zero. Comparison of the data and the bicubic-spline plots will indicate regions where the cubic spline exhibits undesirable or exaggerated hills and valleys. Using small to moderate tension-parameter values (say, from 1 to 10) for those regions, the user can recalculate a smooth rational spline. In this way, after a few trials with different tension values, a rational-spline surface can be obtained which the user considers representative of the data.

Implementation Considerations

In implementing the rational-spline surface smoother, the number and locations of the knots need to be considered. For the given m data points, the number of knots chosen (M and N) is restricted by the relationship

$$4MN < m \tag{20}$$

since there are $4MN$ unknowns. The knot locations need to be selected so that all the data lie within the region $\{\bar{x}_1 \leq x \leq \bar{x}_m,\ \bar{y}_1 \leq y \leq \bar{y}_M\}$ or on the boundaries. A few data points must lie in each subregion R_{ij}. Since there are $(M-1)(N-1)$ subregions, dividing each side of equation (20) by $(M-1)(N-1)$ indicates that there must be more than $4MN/(M-1)(N-1)$ data points either in each subregion or on the boundaries of the subregion. Note that points on a boundary between two subregions count towards the total number of data points for each of the two subregions. In the terrain elevation example given later, $m = 121$ and $M = N = 5$. Thus, equation (20) becomes $100 < 121$ and there must be more than $(4)(5)(5)/(4)(4) = 6.25$ data points per subregion; that is, each subregion must contain at least 7 data points. As constructed, there are 9 data points within or on the boundaries of each subregion.

The two matrices to be inverted in the solution of equations (17) and (18) can be large—$\mathbf{E}^T\mathbf{W}\mathbf{E}$ is $4MN$ by $4MN$ and $\mathbf{D}(\mathbf{E}^T\mathbf{W}\mathbf{E})^{-1}\mathbf{D}^T$ is $2MN$ by $2MN$. For example, for $M = N = 5$ these matrices are 100 by 100 and 50 by 50, respectively. Calculation of such large inverses with a general matrix inversion method, such as that of Gauss-Jordan, may lead to loss of precision. However, both matrices are symmetric and positive definite. Therefore it is highly recommended that a Cholesky method (ref. 7) be used since it takes advantage of the characteristics of the matrices.

Example

The data chosen to illustrate the rational-spline surface smoother consist of terrain elevation measurements on a square area 1000 ft on each side. The measurements were taken at 100-ft intervals in both the x- and the y-direction. This yields a total of $m = 121$ measurements. Elevations are measured to an accuracy of 0.1 ft.

Figure 1 shows a surface plot of the terrain elevation data. The elevations range from 12.0 ft (at $x = 300$ ft, $y = 300$ ft) to 31.0 ft (at $x = 700$ ft, $y = 500$ ft). As shown in figure 1, the original data are relatively smooth. For the purposes of this example, the original elevation measurements are perturbed by addition of normally distributed error having a mean of zero and a standard deviation of 1 ft. A surface plot of the resulting measurements is shown in figure 2. It is to these measurements that the surface smoother is applied.

Knots are chosen to be located at the points 0 ft, 250 ft, 500 ft, 750 ft, and 1000 ft along each axis; this choice gives a 5 by 5 ($N = M = 5$) grid of knot locations. As a result, $4MN = 100$ unknown function and derivative values must be estimated. In the case of zero-tension values ($p_i = 0$ for $i = 1,2,3,4$ and $q_j = 0$ for $j = 1,2,3,4$), the bivariate rational-spline smoother reduces to a bicubic-spline smoother. Figure 3 shows a surface plot of the bicubic-spline smoothed values of the measurements including error. The values plotted in this figure are generated through evaluation of the bicubic spline at 50-ft intervals along each axis in order to show the behavior of the surface at intermediate points. Although the bicubic-spline surface is smooth, it contains oscillations which are not present in the original data. It is these artificial oscillations which can be removed or reduced by application of tension.

Figure 4 shows the rational-spline surface when all 8 tension parameters have a value of 10. Comparison with figure 3 shows that the nonzero tension gives the surface a less "rounded" appearance because the magnitude of the cubic terms is smaller when the tension has a value of 10 than when it has a value of 0. This tendency for the rational spline to be more linear for a tension value of 10 is clearly shown along the $y = 0$ edge for large x. This edge in figure 4 also has shallower undulations than the same edge in figure 3. This reduction in undulations is the purpose for applying tension.

As an example of the application of tension to subregions of a surface, only the tension parameters p_4 and q_4 are set to a value of 10 and all the others are set to 0. Figure 5 shows the surface resulting from these tension parameters. Comparison of figure 5 with figures 3 and 4 shows that the surface for the two regions (750 ft $\leq x \leq$ 1000 ft, 0 ft $\leq y \leq$ 1000 ft) and (0 ft $\leq x \leq$ 1000 ft, 750 ft $\leq y \leq$ 1000 ft) in figure 5 closely resembles the surface for the same regions in figure 4. In contrast the remaining surface of figure 5 more closely resembles the corresponding surface in figure 3. Hence, the surface can be smoothed with rational splines in which the tension selectively varies from region to region.

Concluding Remarks

An algorithm for surface smoothing with bivariate rational-spline functions on a rectangular grid has been presented. The rational-spline function combines the advantages of a cubic function having continuous first and second derivatives over the entire grid with the advantage of a function having variable tension. Adjustment of the tension parameters in the rational spline allows the user to reduce unwanted oscillations to any desired level across the surface. The multiple parameters of the rational spline provide adjustable parameters for each rectangular subregion and thus give the user control over local behavior of the surface.

The terrain elevation example presented illustrates the reduction in undesirable oscillations that is possible with a bivariate rational spline which smooths the measurements. The example also illustrates that the bivariate rational spline provides the capability of controlling local oscillations in the surface caused by trend variations in the data.

There is one major disadvantage to using the bivariate rational spline. Several computer runs with different tension values may be necessary to find a smoothing surface satisfactory to the user. It is recommended that for the initial run the user apply zero tension in order to ascertain regions of excessive oscillations. The user can then determine appropriate tension values by examining the surface plots from a few additional computer runs.

NASA Langley Research Center
Hampton, VA 23665-5225
March 30, 1987

Appendix

Derivation of Boundary Constraints

In this appendix equations (11) to (14), which express the constraints along the boundaries, are derived. The derivations of all four sets of equations are similar. A detailed derivation of equation (11) is given.

With the definition of the rational spline in equation (2), consider the interval $\bar{x}_i \leq x \leq \bar{x}_{i+1}$, fix j, and evaluate $f_{ij}(x,y)$ at $\bar{y}_j$. Thus, $h_{j1}(\bar{y}_j) = h_{j3}(\bar{y}_j) = 1$, $h_{j2}(\bar{y}_j) = h_{j4}(\bar{y}_j) = 0$, and

$$f_{ij}(x,\bar{y}_j) = \sum_{k=1}^{4} g_{ik}(x)(a_{ijk1} + a_{ijk3}) \qquad \text{(A1)}$$

For simplicity, define $b_{ijk} = a_{ijk1} + a_{ijk3}$. Then, from equation (A1) and the definitions of $g_{ik}(x)$,

$$\left.\begin{aligned} f_{ij}(\bar{y}_i,\bar{y}_j) &= b_{ij1} + b_{ij3} = \overline{F}_{ij} \\ f_{ij}(\bar{x}_{i+1},\bar{y}_j) &= b_{ij2} + b_{ij4} = \overline{F}_{(i+1)j} \end{aligned}\right\} \qquad \text{(A2)}$$

where $\overline{F}_{ij}$ and $\overline{F}_{(i+1)j}$ are unknown function values at the grid points.

The first derivatives of $g_{ik}(x)$ are

$$\left.\begin{aligned} g'_{i1}(x) &= -\frac{1}{dx_i} \qquad g'_{i3}(x) = \frac{-3u^2(p_it + 1) - u^3p_i}{dx_i(p_it + 1)^2} \\ g'_{i2}(x) &= \frac{1}{dx_i} \qquad g'_{i4}(x) = \frac{3t^2(p_iu + 1) + t^3p_i}{dx_i(p_iu + 1)^2} \end{aligned}\right\} \qquad \text{(A3)}$$

Differentiating equation (A1) with respect to x, applying equations (A3), and evaluating at $\bar{x}_i$ and $\bar{x}_{i+1}$ yields

$$\left.\begin{aligned} \frac{\partial f_{ij}}{\partial x}(\bar{x}_i,\bar{y}_j) &= -\frac{b_{ij1}}{dx_i} + \frac{b_{ij2}}{dx_i} - \frac{3 + p_i}{dx_i}b_{ij3} = \overline{FX}_{ij} \\ \frac{\partial f_{ij}}{\partial x}(\bar{x}_{i+1},\bar{y}_j) &= -\frac{b_{ij1}}{dx_1} + \frac{b_{ij2}}{dx_i} + \frac{3 + p_i}{dx_i}b_{ij4} \\ &= \overline{FX}_{(i+1)j} \end{aligned}\right\} \qquad \text{(A4)}$$

where $\overline{FX}_{ij}$ and $\overline{FX}_{(i+1)j}$ are the unknown first derivatives with respect to x at interior grid points.

The second derivatives of $g_{ik}(x)$ are

$$\left.\begin{aligned} g''_{i1}(x) &= g''_{i2}(x) = 0 \\ g''_{i3}(x) &= \frac{6u(p_it + 1)^2 + 6u^2p_i(p_it + 1) + 2u^3p_i^2}{dx_i^2(p_it + 1)^3} \\ g''_{i4}(x) &= \frac{6t(p_iu + 1)^2 + 6t^2p_i(p_iu + 1) + 2t^3p_i^2}{dx_i^2(p_iu + 1)^3} \end{aligned}\right\} \qquad \text{(A5)}$$

Differentiating equation (A1) twice with respect to x, applying equations (A5), and evaluating the results at $\bar{x}_i$ and $\bar{x}_{i+1}$ leads to

$$\left.\begin{aligned} \frac{\partial^2 f_{ij}}{\partial x^2}(\bar{x}_i,\bar{y}_j) &= \frac{2p_i^2 + 6p_i + 6}{dx_i^2}b_{ij3} \\ \frac{\partial^2 f_{ij}}{\partial x^2}(\bar{x}_{i+1},\bar{y}_j) &= \frac{2p_i^2 + 6p_i + 6}{dx_i^2}b_{ij4} \end{aligned}\right\} \qquad \text{(A6)}$$

Solve equations (A2) for b_{ij1} and b_{ij2}, substitute into equations (A4), and rearrange the results to obtain

$$-(2 + p_i)b_{ij3} - b_{ij4} = dx_i\overline{FX}_{ij} + \overline{F}_{ij} - \overline{F}_{(i+1)j} \qquad \text{(A7)}$$

$$b_{ij3} + (2 + p_i)b_{ij4} = dx_i\overline{FX}_{(i+1)j} + \overline{F}_{ij} - \overline{F}_{(i+1)j} \qquad \text{(A8)}$$

Define $\Delta_x\overline{F}_{ij} = \overline{F}_{(i+1)j} - \overline{F}_{ij}$ and solve equation (A7) for b_{ij4} to get

$$b_{ij4} = -(2 + p_i)b_{ij3} - dx_i\overline{FX}_{ij} + \Delta_x\overline{F}_{ij} \qquad \text{(A9)}$$

Substituting equation (A9) into equation (A8) and solving for b_{ij3} yields

$$b_{ij3} = \frac{(3 + p_i)\Delta_x\overline{F}_{ij} - dx_i\overline{FX}_{(i+1)j} - (2 + p_i)dx_i\overline{FX}_{ij}}{(2 + p_i)^2 - 1} \qquad \text{(A10)}$$

Substituting equation (A10) into equation (A9) gives

$$b_{ij4} = \frac{-(3 + p_i)\Delta_x\overline{F}_{ij} + (2 + p_i)dx_i\overline{FX}_{(i+1)j} + dx_i\overline{FX}_{ij}}{(2 + p_i)^2 - 1} \qquad \text{(A11)}$$

In order to derive equation (11), evaluate the first of equations (A6) at $i = 1$, substitute for b_{ij3} from equation (A10), and set the result to zero:

$$\frac{\partial^2 f_{1j}(x_1, y_j)}{\partial x^2} = 0$$

or

$$\frac{2p_1^2 + 6p_1 + 6}{dx_1^2}$$

$$\left[\frac{(3+p_1)\Delta_x \overline{F}_{1j} - dx_1 \overline{FX}_{2j} - (2+p_1) dx_1 \overline{FX}_{1j}}{(2+p_1)^2 - 1}\right] = 0$$

Substituting for $\Delta_x \overline{F}_{1j}$ and dividing both sides by the left factor, which is nonzero, yields

$$\frac{(3 + p_1)}{dx_1}\overline{F}_{1j} - \frac{(3 + p_1)}{dx_1}\overline{F}_{2j} + (2 + p_1)\,\overline{FX}_{1j} + \overline{FX}_{2j} = 0 \tag{A12}$$

Equation (A12) is identical to equation (11) and holds for $j = 1, 2, ..., N$.

For the derivation of equation (12), set to zero the second derivative with respect to x along the boundary $x = \bar{x}_M$ and substitute the second of equations (A6) with $i = M - 1$ and the definition of b_{ij4} in equation (A11).

Equations (13) and (14) are derived analogously by differentiation of equation (2) twice with respect to y when $x = \bar{x}_i$ for fixed i. The result is evaluated separately at the boundaries $y = \bar{y}_1$ and $y = \bar{y}_N$, and appropriate substitutions are made.

References

1. Böhm, Wolfgang; Farin, Gerald; and Kahmann, Jürgen: A Survey of Curve and Surface Methods in CAGD. *Comput. Aided Geom. Design,* vol. 1, no. 1, July 1984, pp. 1–60.
2. Späth, Helmuth (W. D. Hoskins and H. W. Sager, transl.): *Spline Algorithms for Curves and Surfaces.* Utilitas Mathematica Publ. Inc., 1974.
3. Frost, Charles E.; and Kinzel, Gary L.: An Automatic Adjustment Procedure for Rational Splines. *Comput. & Graphics*, vol. 6, no. 4, 1982, pp. 171–176.
4. Schiess, James R.; and Kerr, Patricia A.: *Rational-Spline Approximation With Automatic Tension Adjustment.* NASA TP-2366, 1984.
5. Schiess, James R.: *Two Algorithms for Rational Spline Interpolation of Surfaces.* NASA TP-2536, 1986.
6. Smith, Robert E., Jr.; Price, Joseph M.; and Howser, Lona M.: *A Smoothing Algorithm Using Cubic Spline Functions.* NASA TN D-7397, 1974.
7. Isaacson, Eugene; and Keller, Herbert Bishop: *Analysis of Numerical Methods.* John Wiley & Sons, Inc., c.1966.

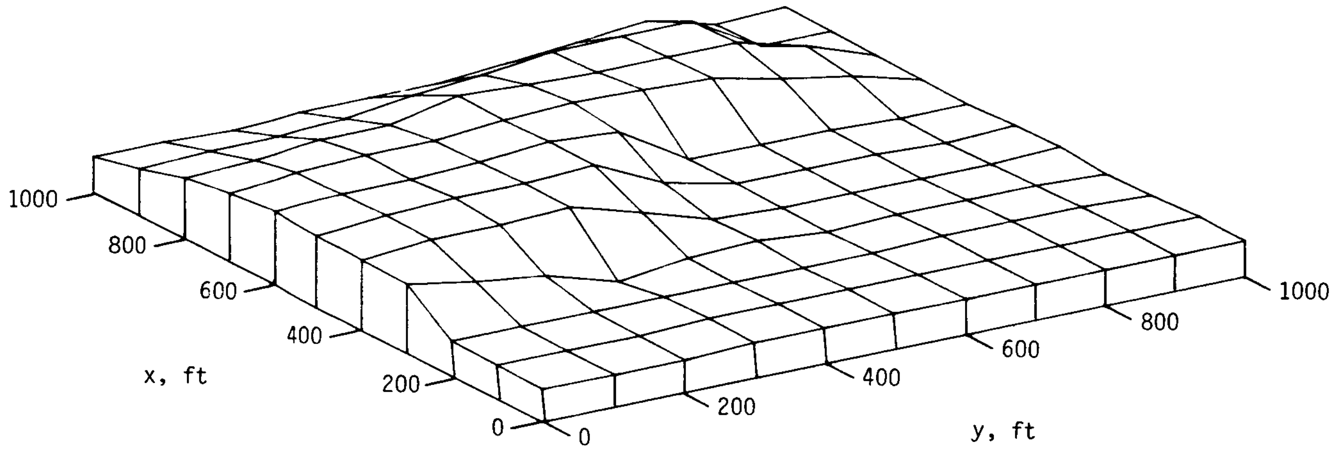


Figure 1. Terrain elevation data.

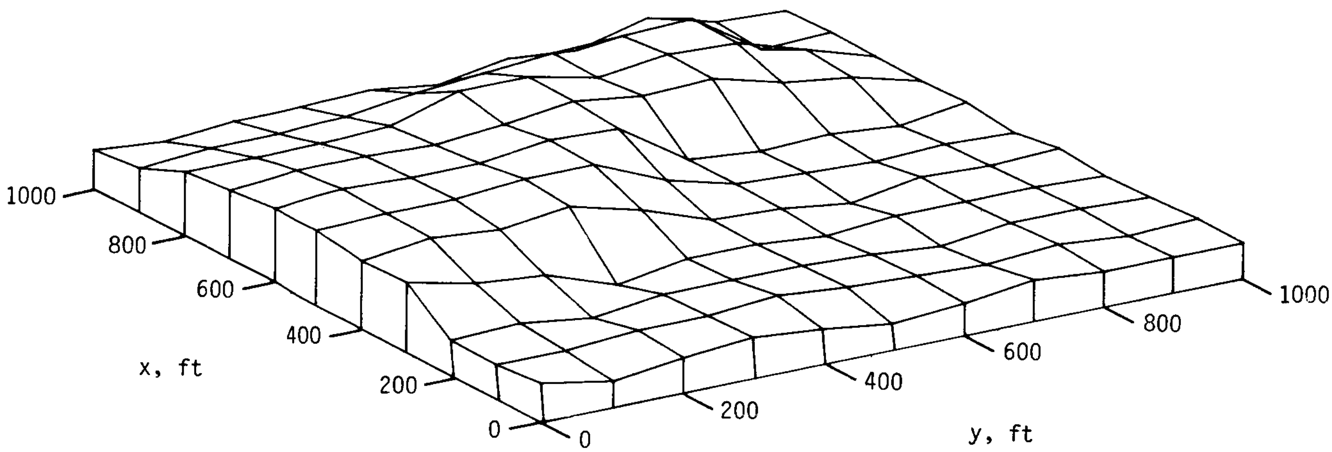


Figure 2. Terrain elevation data with random error (standard deviation of 1.0 ft) added.

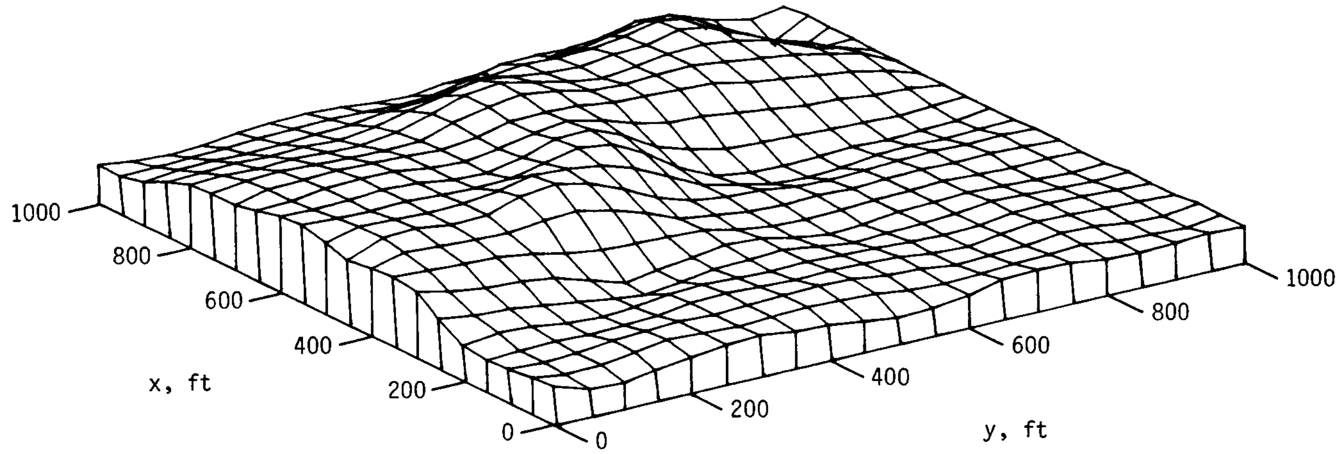


Figure 3. Bicubic-spline smoothed terrain elevation surface.

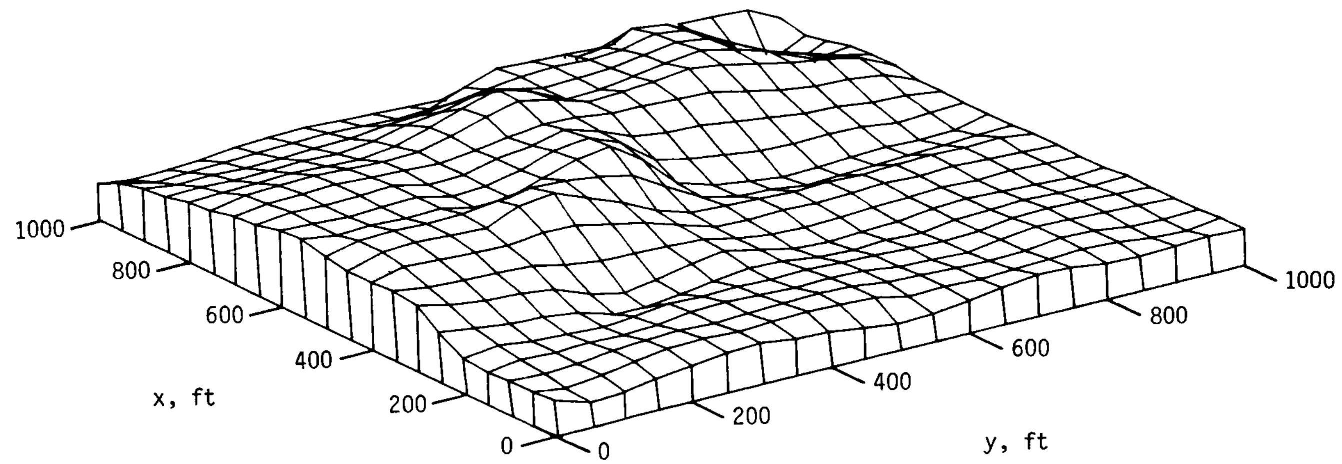


Figure 4. Rational-spline smoothed terrain elevation surface for $p_i = q_j = 10$.

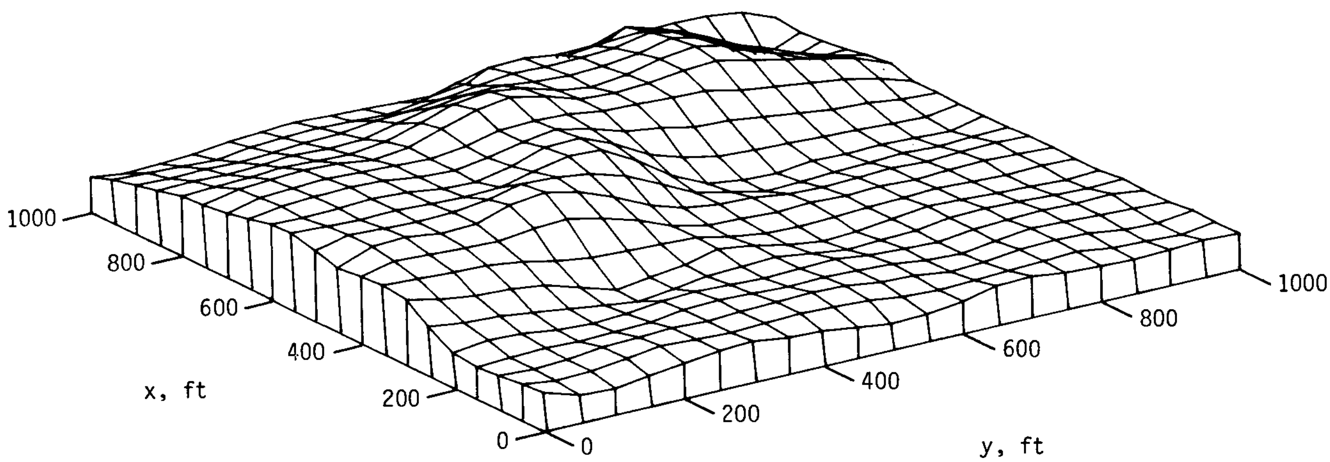


Figure 5. Rational-spline smoothed terrain elevation surface for $p_4 = q_4 = 10$.

National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TP-2708		

4. Title and Subtitle	5. Report Date
An Algorithm for Surface Smoothing With Rational Splines	June 1987
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
James R. Schiess	L-16272
9. Performing Organization Name and Address	10. Work Unit No.
NASA Langley Research Center Hampton, VA 23665-5225	505-90-21-02
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Washington, DC 20546-0001	Technical Paper
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This paper discusses an algorithm for smoothing surfaces with spline functions containing tension parameters. The bivariate spline functions used are tensor products of univariate rational-spline functions. A distinct tension parameter corresponds to each rectangular strip defined by a pair of consecutive spline knots along either axis. Equations are derived for writing the bivariate rational spline in terms of functions and derivatives at the knots. Estimates of these values are obtained via weighted least squares subject to continuity constraints at the knots. The algorithm is illustrated on a set of terrain elevation data.

17. Key Words (Suggested by Authors(s))	18. Distribution Statement
Weighted least squares Tension parameter Bicubic rational spline	Unclassified—Unlimited Subject Category 64

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	15	A02

For sale by the National Technical Information Service, Springfield, Virginia 22161-2171