Local smoothing by polynomials in n dimensions

By A. J. Cole and A. J. T. Davie*

A method of local smoothing of noisy data by making a least squares fit to a suitably chosen polynomial is described. Various n-dimensional formulae are derived and their effects compared empirically.

(First received November 1967 and in revised form June 1968)

We consider the problem of smoothing data in n dimensions by making a least squares fit of the data to a polynomial. This will be a local rather than a global fit in the sense that corrections will be made point by point using at each stage only a few selected points about the one currently under consideration. The one dimensional case has been considered in detail by Lanczos (1957) and Hildebrand (1956).

Impose a mesh of unit step length on the domain under consideration. There is no loss of generality in choosing co-ordinates so that the point currently under consideration is the origin. It will be assumed at this stage that the origin is interior to the domain. Boundary and near boundary points will be considered later.

Let P be the approximating polynomial and let $X = \{x_0, x_1 \ldots x_k\}$ be a set of mesh points including x_0, the origin. The problem then is to minimise

$$F = \sum_{i=0}^{k} (P_i - f_i)^2 \qquad (1)$$

where $P_i = P(x_i)$ and f_i is the given data at point x_i. If $\bar{P}$ is the polynomial minimising F then the corrected value at x_0 is $\bar{P}(x_0)$. That is, the corrected value is the constant in $\bar{P}$.

We restrict ourselves to symmetric sets X such that if $x = (x_1, x_2, x_3 \ldots x_n) \epsilon X$ then so also are all points $(\pm x_i, \pm x_j, \ldots \pm x_q)$ where any combination of signs is taken and $(i, j, \ldots q)$ is any permutation of $(1, 2, \ldots n)$. With this condition it follows that only polynomials with all combinations of even powers of the x_j need be considered since, by symmetry, differentiation of (1) with respect to the coefficients of P will lead to a set of equations which contains an independent subset involving only the coefficients of products of even powers of the x_j and the constant term. Further, only polynomials of even degree $2m$ need be considered since a polynomial of degree $2m + 1$ would lead to the same set of equations to be solved for the constant term. In particular, a linear fit may be accomplished by considering the constant term of the linear polynomial only. The linear case can therefore be classed as trivial though some results are given below for completeness.

Notation

Before continuing, we introduce notation and the basic identities for differences in many dimensions.

We use

$$\delta_{a, b, \ldots g} \text{ for } \sum_{i,j,\ldots q} \delta^{a+b+\ldots+g}_{x_i x_i \ldots x_i x_j \ldots x_j \ldots x_q \ldots x_q} f_0$$

* *Computing Laboratory, The University, St. Andrews, Fife*

where there are a suffices x_i, b suffices x_j, and so on.

$$e_i = (0, 0, \ldots, 1, \ldots 0)$$

where the 1 appears in the ith position

$$e_{ki} = (0, 0, \ldots, k, \ldots, 0)$$

where the k appears in the ith position

$$e_{\pm i, \pm j} = (0, 0 \ldots, \pm 1, \ldots, \pm 1, \ldots 0)$$

and so on.

$$f_{\pm ki} = f(\pm e_{ki})$$

$$f_{\pm i, \pm j} = f(e_{\pm i, \pm j})$$

and so on.

$$\Sigma^*_{ki} = \sum_{i=1}^{n} [f_{ki} + f_{-ki}]$$

$$\Sigma^*_{i,j} = \sum_{i=2}^{n} \sum_{j=1}^{i-1} [f_{i,j} + f_{i,-j} + f_{-i,j} + f_{-i,-j}]$$

and so on.

$$\Sigma'_{ki} = f_{ki} + f_{-ki}$$

$$\Sigma'_{i,j} = \sum_{\substack{j \neq i \\ j=1}}^{n} f_{\pm i, \pm j}$$

$$\Sigma'_{i,2j} = \sum_{\substack{j \neq i \\ j=1}}^{n} f_{\pm i, \pm 2j}$$

$$\Sigma'_{i,j,k} = \sum_{\substack{j,k \neq i \\ j<k \\ j,k=1}}^{n} f_{\pm i, \pm j, \pm k}$$

where the $\pm$ notation is used to imply that all possible combinations of signs are taken on the suffices.

It follows that

$$\sum_{i=1}^{n} \Sigma'_{ki} = \Sigma^*_{ki}$$

$$\sum_{i=1}^{n} \Sigma'_{i,j} = 2\Sigma^*_{i,j}$$

$$\sum_{=1}^{n} \Sigma'_{2i,j} = \sum_{i=1}^{n} \Sigma'_{i,2j} = \Sigma^*_{i,2j} = \Sigma^*_{2i,j}$$

$$\sum_{i=1}^{n} \Sigma'_{i,j,k} = 3\Sigma^*_{i,j,k}$$

and so on.

Downloaded from https://academic.oup.com/comjnl/article/12/1/72/311660 by guest on 16 August 2022

The following identities may be verified.

$$\delta_2 = \Sigma_i^* - 2nf_0$$
$$\delta_4 = \Sigma_{2i}^* - 4\Sigma_i^* + 6nf_0$$
$$\delta_{2,2} = \Sigma_{i,j}^* - 2(n-1)\Sigma_i^* + 2n(n-1)f_0$$
$$\delta_6 = \Sigma_{3i}^* - 6\Sigma_{2i}^* + 15\Sigma_i^* - 20nf_0$$
$$\delta_{4,2} = \delta_{2,4} = \Sigma_{2i,j}^* - 2(n-1)\Sigma_{2i}^* - 8\Sigma_{i,j}^* + 14(n-1)\Sigma_i^* - 12n(n-1)f_0$$
$$\delta_{2,2,2} = \Sigma_{i,j,k}^* - 2(n-2)\Sigma_{i,j}^* + 2(n-1)(n-2)\Sigma_i^* - \tfrac{4}{3}n(n-2)f_0$$

and conversely

$$\Sigma_i^* = \delta_2 + 2nf_0$$
$$\Sigma_{2i}^* = \delta_4 + 4\delta_2 + 2nf_0$$
$$\Sigma_{i,j}^* = \delta_{2,2} + 2(n-1)\delta_2 + 2n(n-1)f_0$$
$$\Sigma_{3i}^* = \delta_6 + 6\delta_4 + 9\delta_2 + 2nf_0$$
$$\Sigma_{2i,j}^* = \delta_{2,4} + 2(n-1)\delta_4 + 8\delta_{2,2} + 10(n-1)\delta_2 + 4n(n-1)f_0$$
$$\Sigma_{i,j,k}^* = \delta_{2,2,2} + 2(n-2)\delta_{2,2} + 2(n-1)(n-2)\delta_2 + \tfrac{4}{3}n(n-1)(n-2)f_0.$$

The linear case

Although the general form is

$$P = \sum_{i=1}^{n} a_i x_i + b$$

it suffices, as indicated above, to consider the special case $P = b$. The function to be minimised is then

$$F = \sum_{j=0}^{m} (b - f(x_j))^2$$

where there are $m + 1$ points in X. For a minimum

$$\frac{1}{2}\frac{\partial F}{\partial b} = (m+1)b - \sum_{j=0}^{m} f(x_j) = 0.$$

Hence

$$(m+1)(b - f_0) = -mf_0 + \sum_{j=1}^{m} f(x_j).$$

The required correction is therefore

$$(b - f_0) = \frac{1}{m+1}\left\{-mf_0 + \sum_{j=1}^{m} f(x_j)\right\}. \quad (2)$$

Using the central difference identities listed above, it is now possible to derive corrections in terms of central differences to correspond to different choices of the set X. For example with X consisting of the origin and all points of type e_i, we have

$$(b - f_0) = \frac{1}{2n+1}\{-2nf_0 + \Sigma_i^*\} = \frac{1}{2n+1}\delta_2.$$

Similar results follow for different sets X, but there is little point in developing them since the correction as expressed in (2) is simple to apply.

Quadratic and cubic polynomials

As shown above, the corrections corresponding to $n = 2m$ and $n = 2m + 1$ are identical. Thus it suffices to consider the quadratic form with even powers only. Let

$$P = \sum_{i=1}^{n} a_i x_i^2 + b.$$

We first tabulate the contributions, C, made to F by the various kinds of points defined above.

(i) the origin
$C = (b - f_0)^2$ (One term only)

(ii) points of type ke_i ($k = 1, 2$ or 3)
$C = (b + k^2 a_i - f_{ki})^2$ ($2n$ terms)

(iii) points of type $e_{i,j}$
$C = (b + a_i + a_j - f_{i,j})^2$ ($2n(n-1)$ terms)

(iv) points of type $e_{i,2j}$
$C = (b + a_i + 4a_j - f_{i,2j})^2$ ($4n(n-1)$ terms)

(v) points of type $e_{i,j,k}$
$C = (b + a_i + a_j + a_k - f_{i,j,k})^2$
($\frac{4}{3}n(n-1)(n-2)$ terms)

where all possible combinations of signs are taken.

Corresponding to the five cases listed above, we can now compute the contributions to $\frac{1}{2}\frac{\partial F}{\partial b}$ made by *all* points of the corresponding types.

(i) $b - f_0$

(ii) $2nb + 2k^2 \sum_{i=1}^{n} a_i - \Sigma_{ki}^*$

(iii) $2n(n-1)b + 4(n-1)\sum_{i=1}^{n} a_i - \Sigma_{i,j}^*$

(iv) $4n(n-1)b + 20(n-1)\sum_{i=1}^{n} a_i - \Sigma_{i,2j}^*$

(v) $\frac{4}{3}n(n-1)(n-2)b + 4(n-1)(n-2)\sum_{i=1}^{n} a_i - \Sigma_{i,j,k}^*$.

Similarly, the contributions to $\frac{1}{2}\frac{\partial F}{\partial a_i}$ made by summing the contributions made by all terms are

(i) 0

(ii) $2k^2 b + 2k^4 a_i - k^2 \Sigma'_{ki}$

(iii) $4(n-1)b + 4(n-2)a_i + 4\sum_{j=1}^{n} a_j - \Sigma'_{i,j}$

(iv) $20(n-1)b + (68n - 100)a_i + 32\sum_{j=1}^{n} a_j - \Sigma'_{2i,j} - 4\Sigma'_{i,2j}$

(v) $4(n-1)(n-2)b + 4(n-3)(n-2)a_i + 8(n-2)\sum_{j=1}^{n} a_j - \Sigma'_{i,j,k}$.

Downloaded from https://academic.oup.com/comjnl/article/12/1/72/311660 by guest on 16 August 2022

Summing over i, contributions made to $\frac{1}{2}\sum_{j=1}^{n}\frac{\partial F}{\partial a_i}$ are

(i) 0

(ii) $2k^2nb + 2k^4\sum_{i=1}^{n} a_i - k^2\Sigma^*_{k\,i}$

(iii) $4n(n-1)b + 8(n-1)\sum_{i=1}^{n} a_i - 2\Sigma^*_{i,\,j}$

(iv) $20n(n-1)b + 100(n-1)\sum_{i=1}^{n} a_i - 5\Sigma^*_{2\,i,\,j}$

(v) $4n(n-1)(n-2)b + 12(n-1)(n-2)\sum_{i=1}^{n} a_i - 3\Sigma^*_{i,\,j,\,k}$.

We give now an example of the derivation of a formula taking the set X to consist of all points of type origin, e_i and $e_{i,\,j}$.

Summing the terms appropriate to $\frac{1}{2}\frac{\partial F}{\partial b} = 0$ we have

$$(b - f_0) + \left(2nb + 2\sum_{i=1}^{n} a_i - \Sigma_i^*\right) + \left(2n(n-1)b + 4(n-1)\sum_{i=1}^{n} a_i - \Sigma^*_{i,j}\right) = 0.$$

Hence

$$\begin{aligned}(2n^2+1)(b-f_0) + (4n-2)\sum_{i=1}^{n} a_i &= -2n^2 f_0 + \Sigma_i^* + \Sigma^*_{i,j}\\ &= -2n^2 f_0 + \delta_2 + 2nf_0 + \delta_{2,\,2} + 2(n-1)\delta_2 + 2n(n-1)f_0\\ &= \delta_{2,\,2} + (2n-1)\delta_2. \qquad (3)\end{aligned}$$

Similarly, corresponding to $\frac{1}{2}\sum_{i=1}^{n}\frac{\partial F}{\partial a_i} = 0$ we have

$$\left(2nb + 2\sum_{i=1}^{n} a_i - \Sigma_i^*\right) + \left(4n(n-1)b + 8(n-1)\sum_{i=1}^{n} a_i - 2\Sigma^*_{i,\,j}\right) = 0$$

or

$$\begin{aligned}(4n^2-2n)(b-f_0) + (8n-6)\sum_{i=1}^{n} a_i &= -(4n^2-2n)f_0 + \Sigma_i^* + 2\Sigma^*_{i,\,j}\\ &= -(4n^2-2n)f_0 + \delta_2 + 2nf_0 + 2\delta_{2,\,2} + 4(n-1)\delta_2 + 4n(n-1)f_0\\ &= 2\delta_{2,\,2} + (4n-3)\delta_2. \qquad (4)\end{aligned}$$

Eliminating $\sum_{i=1}^{n} a_i$ from (3) and (4) we obtain

$$[(2n^2+1)(4n-3) - 2n(2n-1)^2](b-f_0) = [(4n-3) - 2(2n-1)]\delta_{2,\,2}.$$

Hence the correction, $b - f_0$, is

$$-\frac{\delta_{2,\,2}}{2n^2+2n-3} \quad (n \geqslant 2).$$

The method of derivation for other groupings is similar and we merely give results.

(a) The origin, e_i and $2e_i$

$$b - f_0 = -\frac{3}{18n+17}\delta_4$$

(b) The origin, e_i and $e_{i,j}$ (the case considered above)

$$b - f_0 = -\frac{1}{2n^2+2n-3}\delta_{2,2}$$

(c) The origin, e_i, e_{2i} and $e_{i,\,j}$

$$b - f_0 = \frac{1}{10n^2+12n+13}(7\delta_{2,2} - (4n-1)\delta_4)$$

(d) The origin, e_i, e_{2i} and e_{3i}

$$b - f_0 = -\frac{1}{7(2n+1)}(2\delta_6 + 9\delta_4)$$

(e) The origin, e_i, e_{2i}, $e_{i,2j}$

$$b - f_0 = -\frac{1}{68n^2-33}(8\delta_{2,4} + 64\delta_{2,2} + 3(2n-1)\delta_4)$$

(f) The origin, e_i, e_{2i}, $e_{i,2j}$, e_{3i}

$$b - f_0 = \frac{1}{66n^2+57n+24}(14\delta_{2,4} - (20n-6)\delta_6 + 112\delta_{2,2} - (87n-24)\delta_4)$$

(g) The origin, e_i, $e_{i,\,j}$, $e_{i,\,j,\,k}$

$$b - f_0 = \frac{-3}{4n^4-4n^2-24n+27}(2n\delta_{2,\,2,\,2} + (2n^2-2n-3)\delta_{2,\,2}).$$

These, of course, do not represent all of the combinations of the seven types of points, but an exhaustive list would be tedious.

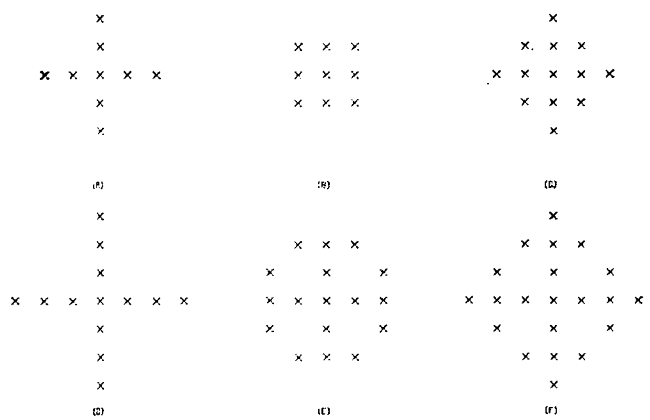

Fig. 1. Groupings of basic points used in smoothing

A diagrammatic form of the above types for $n = 2$ is given in **Fig. 1** showing which points are used in the evaluation. Case (g) is not included since it has no meaning for $n = 2$.

Boundary points

The formulae deduced for interior points depended heavily on symmetry for the simplicity of their derivation. When considering boundary points this symmetry is lost if only the basic set of points is considered and indeed the equivalence of formulae of orders $2m$ and $2m + 1$ no longer holds. Lanczos (1957) obtains the formula

$$\text{correction} = \frac{1}{5}\delta^3 + \frac{3}{35}\delta^4 \qquad (5)$$

Downloaded from https://academic.oup.com/comjnl/article/12/1/72/311660 by guest on 16 August 2022

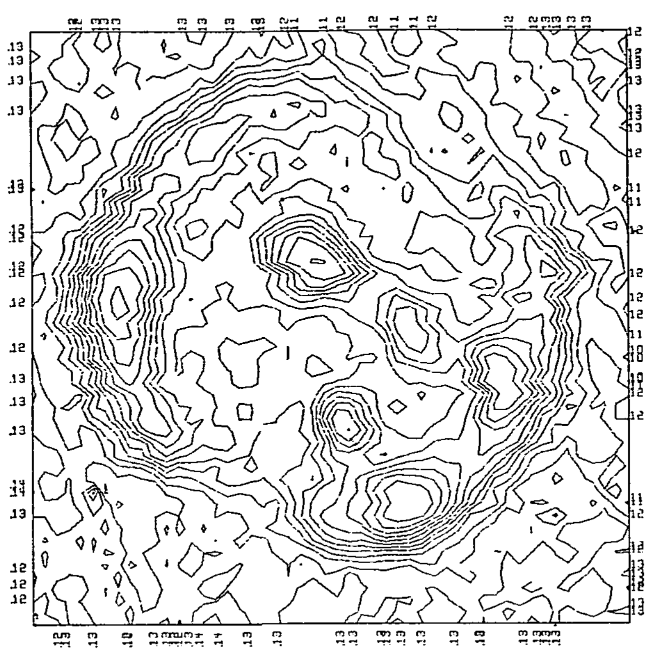

Fig. 2. Solar soft X-ray raw data

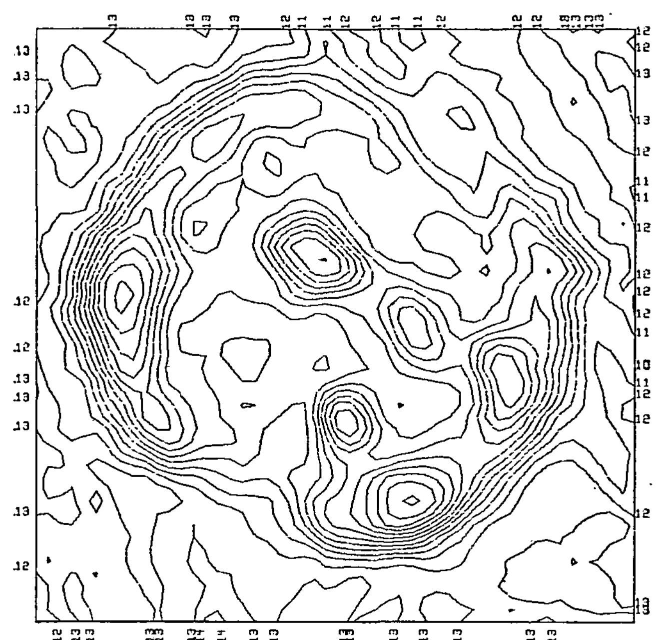

Fig. 3. Solar soft X-ray map correction from double linear smoothing

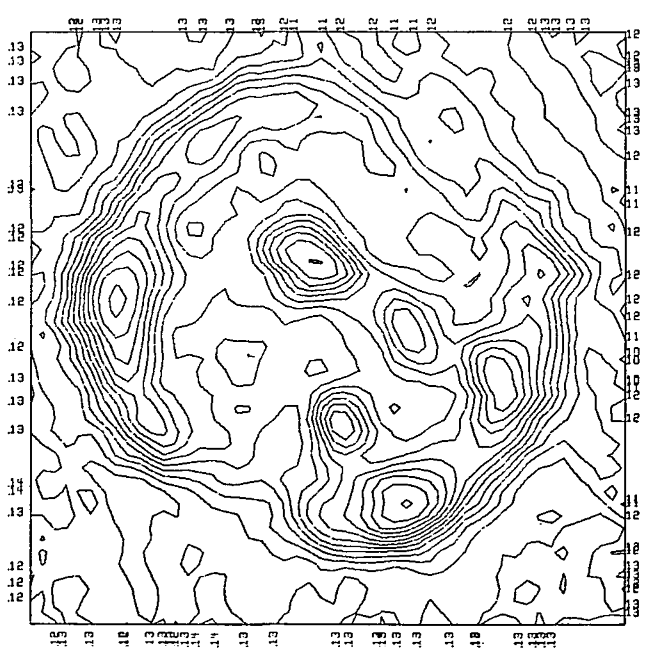

Fig. 4. Solar soft X-ray map correction from 9-point block

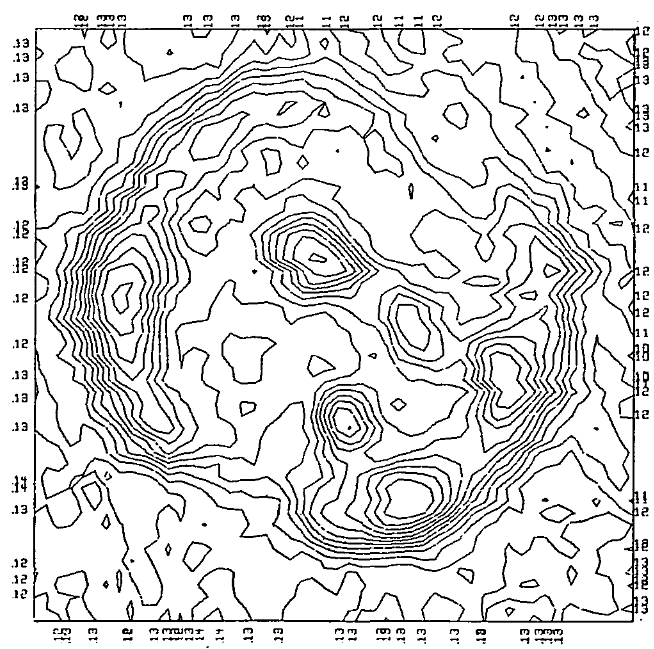

Fig. 5. Solar soft X-ray correction from 9-point cross

for the one dimensional quadratic case and Hildebrand (1956) obtains

$$\text{correction} = -\frac{1}{70}\,\delta^4 \qquad (6)$$

for the one dimensional cubic case, where δ^3 and δ^4 are the nearest available central differences.

Because of the asymmetry the methods of the preceding paragraphs become considerably more complex and do not lead to such concise formulae. In particular it is no longer sufficient to compute only the constant term in the approximating polynomial.

In the applications discussed in the next paragraph we see that the correction corresponding to case (*b*) above gives generally satisfactory results for interior points. Further, because of the compactness of the set of points used it can be applied to all points excepting those actually lying in the boundary planes of the rectangular parallelepiped of points under consideration. One method of smoothing in these planes is to drop down one dimension and smooth with the corresponding lower dimensional formula. In practice this method has given satisfactory results and in cases where adjacent interior points are significant a second smoothing can be carried

Downloaded from https://academic.oup.com/comjnl/article/12/1/72/311660 by guest on 16 August 2022

out by using a one dimensional formula applied normally to the bounding plane.

If this approach is not satisfactory then individual asymmetric formulae can be derived. For example, for the case $n = 2$ and using a quadratic approximating polynomial

$$ax^2 + bxy + cy^2 + dx + ey + k$$

through the points origin, $e_i, e_{i,j}$, then the smoothed values at (1, 0) and (1, 1) will be

$$a + d + k$$

and

$$a + b + c + d + e + k$$

respectively. Applying a conventional least squares method the coefficients

$$\frac{1}{9}\begin{pmatrix} -1 & -1 & 2 \\ 2 & 2 & -4^* \\ -1 & -1 & 2 \end{pmatrix} \text{ and } \frac{1}{36}\begin{pmatrix} -1 & 8 & -7^* \\ -4 & -4 & 8 \\ 5 & -4 & -1 \end{pmatrix}$$

are obtained as the linear sum multipliers of the corresponding function values

$$\begin{matrix} f(-1, 1) & f(0, 1) & f(1, 1) \\ f(-1, 0) & f(0, 0) & f(1, 0) \\ f(-1, -1) & f(0, -1) & f(1, -1) \end{matrix}$$

to give the corrections at the points indicated by the asterisks.

Some practical results

Two methods of comparing the efficiency of the above formulae were used. In the first certain functions were tabulated and then modified by adding a random error of a given order at each point. The resulting functions were then smoothed and the results compared by computing the sum of squares of deviations of the smoothed values from the known true values. **Table 1** gives results for the functions (i) $\sin(x + 2y)$ and (ii) $\log(x + y) + sqrt((2x^3 + y^3)/(x + y))$ both tabulated for $x = 0{\cdot}5\,(0{\cdot}05)0{\cdot}7$, $y = 0{\cdot}5(0{\cdot}05)0{\cdot}7$ and subjected to random errors of order 10^{-3}. For comparative purposes the second column contains the results for smoothing linearly first in the x direction and then in the y direction (the order of linear smoothing is immaterial as is easily proved). The formula used are labelled as above. In every case tried the nine point block of formula (b) gave the best results.

Table 1

Sums of squares of deviations

All results are multiplied by 10^4

FUNCTION	UN-SMOOTHED	DOUBLE LINEAR	*a*	*b*	*c*
(i)	2·8018	1·8001	1·9880	1·5779	1·6761
(ii)	2·7058	1·7985	1·9350	1·5783	1·6756

The second method used to test the results was to smooth some data from photographs of soft x-rays emitted from the sun and taken from a rocket fired from the Woomera range. The data was subject to background noise of various types. **Fig. 2(a)** shows an attempt to contour the unsmoothed data. There are several places where the contouring program failed resulting in breaks in the contours and also places, particularly in the corners, where background noise has obviously caused distortion. All of the methods applied smoothed the data sufficiently to remove breaks in contours. **Fig. 2(b)** shows the result of a double linear correction which has resulted in over smoothing to some extent. **Fig. 2(c)** corresponds to the nine-point block of formula (b) and is probably the best in that it has removed most of the noise without undue distortion. **Fig. 2(d)** corresponds to the nine-point cross of formula (a) and again oversmooths the data but removes the background noise satisfactorily.

We are indebted to Dr. Peter Russell of Leicester University for supplying us with this data (Russell, 1965a and b).

References

LANCZOS, C. (1957). *Applied Analysis*, Pitman.
HILDEBRAND, F. B. (1956). *Introduction to numerical analysis*, McGraw-Hill.
RUSSELL, P. C. (1965a). *Nature*, Vol. 205, pp. 684–685.
RUSSELL, P. C. (1965b). *Nature*, Vol. 206, pp. 281–282.

Downloaded from https://academic.oup.com/comjnl/article/12/1/72/311660 by guest on 16 August 2022