

Smoothing and interpolation with radial basis functions

Donald E. Myers

Department of Mathematics, University of Arizona, Tucson AZ, 85721, U.S.A.

E-Mail: myers@math.arizona.edu

1 INTRODUCTION

Smoothing and interpolation are two aspects of the problem of "fitting" a function to data. In some instances one has only smoothing, in others only interpolation and in others both smoothing and interpolation. We shall assume that interpolation means that the function values at a data location match the data values, this property is variously called perfect or exact. In many applications the function and perhaps even its form are unknown, i.e., there are no state equations and hence one must make some form of assumptions in order for the problem to be well-posed. Although we begin with an example where the data locations are in 1-space, we shall mostly consider the case of k -dimensional Euclidean space. There are at least two different objectives in fitting a function to data (as we shall see these are really equivalent but lead to different approaches to the problem), one is to obtain a function in analytic form and secondly, to estimate or compute the values of the function at non-data points. At the end we shall also consider a form of smoothing that is somewhat different and more akin to filtering.

2 AN EXAMPLE

Let $(x_1, y_1), \dots, (x_n, y_n)$ be points in 2-space and we visualize them on a scatter plot. The vertical axis is to represent the values of the function and the horizontal the independent variable (the data locations). Depending on the data, to interpolate may result in a very irregular curve. Depending on the size of n , the choice of the form of the interpolating function may

mean there is no solution. For example, if the interpolating function is a polynomial in x , then the degree must be at least n . Smoothing would correspond to assuming a model of the form

$$y = F(x) + \varepsilon \quad (1)$$

where $F(x)$ is the smoothing function and ε represents the noise or error. Smoothing then corresponds to removing the noise or error from the data. The most obvious way to do this is to use Least Squares, i.e. to minimize

$$\sum_{i=1}^n [y_i - F^*(x_i)]^2 \quad (2)$$

Each term in the sum could also have a weighting coefficient, the problem then is to decide the weights. As is well-known, if $F^*(x)$ is a polynomial of degree less than n then minimizing the expression in eq(2) leads to a system of linear equations (for the unknown coefficients). The choice of the degree of the polynomial corresponds to a choice about the degree of smoothing. The extreme case is a polynomial of degree one, i.e., a line is fitted to the data.

Although the above description of the problem is essentially deterministic and requires no theoretical assumptions, it is equivalent to a statistical version. Suppose that the model is of the form

$$Y = A + BX + \varepsilon \quad (3)$$

where Y is a random variable with mean $A + BX$, the ε 's are assumed to be independent normal with mean zero and constant variance σ^2 . Then the Maximum Likelihood estimates of A, B are the same as those obtained by Least Squares. The statistical formulation is useful not only in this case but also later on. At this point it is useful because it suggests an additional aspect of the problem.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n [y_i - \bar{y}]^2 \quad (4)$$

is a measure of the variation in the y -values, where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (5)$$

Obviously one smoothing of the data is to take $F^*(x)$ as a constant, $F^*(x) \equiv \bar{y}$ is the optimal choice for the constant. We may then ask how much better it is to use a model such as in eq(3), this is quantified by

$$\frac{1}{s^2} \left[s^2 - \sum_{i=1}^n [y_i - a - bx_i]^2 \right] \quad (6)$$

This quantity is commonly known as the percent variance explained (by the model). In the statistical model this quantity has an F distribution.

Finally we give a brief illustration of the additional form of smoothing alluded to in the Introduction. Let $[c, d]$ be an interval such that each x_i is in the interval and let m be an integer smaller than n (usually much smaller). Partition the interval $[c, d]$ into m disjoint, equal sized sub-intervals. For each sub-interval, compute the average of the y_i 's for which the corresponding x_i 's are in the sub-interval. Let this average be denoted by u_j and the mid-point of the sub-interval by w_j . We now replace the original data set $(x_1, y_1), \dots, (x_n, y_n)$ by the data set $(u_1, w_1), \dots, (u_m, w_m)$. The data has been "smoothed" but without obtaining an analytic form for the function. An interpolating function might then be fitted to the smoothed data. An obvious question to ask is whether interpolating the smoothed data is equivalent to smoothing the interpolated data. Furthermore, the interpolating function is in some sense an estimator for the "true" but unknown function and it is reasonable to ask how smoothing the "true" function relates to interpolating the smoothed data or smoothing the interpolating function.

3 POSITIVE DEFINITENESS

Because positive definiteness and conditional positive definiteness are critical to the unique determination of the coefficients in a radial basis function interpolator, for completeness basic properties are given.

Let $g(x, y)$ be a real valued function defined on $R_k \times R_k$. $g(x, y)$ is positive definite if for any points x_1, \dots, x_n and any coefficients $\lambda_1, \dots, \lambda_n$ the quadratic form

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j g(x_i, x_j) \quad (7)$$

is positive (except when all the coefficients are zero). Of course then $g(x_i, x_j)$ generates a positive definite matrix which is invertible. Let $f_0(x), \dots, f_p(x)$ be $p+1$ linearly independent functions defined on R_k , then $g(x, y)$ is conditionally positive definite with respect to these functions if the quadratic form, (7), is positive (except when all the coefficients are zero) for all coefficients satisfying

$$\sum_{i=1}^n \lambda_i f_j(x_j) = 0 \text{ for all } j = 0, \dots, p \quad (8)$$

In the case of a conditionally positive definite function, $g(x_i, x_j)$ does not generate a positive definite matrix and in general this matrix is not invertible. This matrix will have both positive and negative eigenvalues. However the following matrix is invertible

$$\begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix} \quad (9)$$

where the submatrix G has entries $g(x_i, x_j)$ and the submatrix F has entries $f_j(x_i)$. In the case that the $f_j(x)$'s are monomials in the coordinates of x , Micchelli (1986) has shown that this matrix is invertible. The proof is easily extendable to the more general case, see Myers (1988).

4 FITTING RADIAL BASIS FUNCTIONS

Consider data of the form $(x_1, y_1), \dots, (x_n, y_n)$ except that now we allow the x_i 's to be points in k -dimensional space. A radial basis function interpolator can be written in the form

$$F^*(x) = \sum_{i=1}^n b_i g(x, x_i) + \sum_{j=0}^p a_j f_j(x) \quad (10)$$

where $g(x, t)$ is a kernel function with properties to be determined later and the $f_j(x)$'s are linearly independent functions, e.g., monomials in the coordinates of x of total degree j . By analogy with the example in the previous section consider

$$\varphi(b_1, \dots, b_n, a_0, \dots, a_p) = \sum_{i=1}^n [y_i - F^*(x_i)] \quad (11)$$

Since $\varphi(b_1, \dots, b_n, a_0, \dots, a_p)$ is linear in the unknown coefficients, the minimum can be found by setting the partial derivatives equal to zero. At this point there are no constraints on the coefficients. One obvious solution is simply

$$\begin{aligned} y_1 &= F^*(x_1) \\ &\vdots \\ y_n &= F^*(x_n) \end{aligned} \quad (12)$$

Unfortunately there are too many unknowns for this system to have a unique solution unless all of a_j 's are zero. In that case a sufficient condition for invertibility of the coefficient matrix is that $g(x, t)$ be positive definite. Micchelli (1986) considered the case of a_0 not zero and the kernel function $|x-t|$. This kernel is not positive definite but rather is conditionally positive definite and one must add an additional constraint, namely

$$b_1 + \dots + b_n = 0 \quad (13)$$

In that case the system has a unique solution and F^* is an interpolating function. If not all of the $[y_i - F^*(x_i)]$'s are zero then the coefficients are the solution of the system

$$\begin{bmatrix} G \\ F^T \end{bmatrix} \begin{bmatrix} G & F \end{bmatrix} \begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} G \\ F^T \end{bmatrix} \begin{bmatrix} Y \\ 0 \end{bmatrix} \quad (14)$$

where G and F are as above, B is the vector of b_i 's, A is the vector of a_j 's and Y is the vector of y_i 's. Unfortunately

$$\begin{bmatrix} G & F \end{bmatrix}$$

is not square (hence not invertible) and

$$\begin{bmatrix} G \\ F^T \end{bmatrix}$$

is not square and hence not invertible. However if $F^T B = 0$ then the above system has the same solution as

$$\begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} Y \\ 0 \end{bmatrix} \quad (15)$$

Hence if $g(x, y)$ is conditionally positive definite with respect to the $f_j(x)$'s then

$$\begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix}$$

is invertible and hence the solution is

$$\begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} Y \\ 0 \end{bmatrix} \quad (16)$$

This neatly shows why the additional constraint equations are needed for the radial basis function interpolator and what they are. Suppose that $f_0(x)$ is identically 1 (zero degree polynomial) and $g(x, y)$ is at least asymptotic to a constant as $x - y$ tends to infinity. Then it is easy to see that for points sufficiently far away from the data locations, the first term on the right in eq (10) is zero or nearly so, hence the "extrapolated" is given only by the second term. The second term is analogous to a trend surface. Hence the two terms in eq(10) might be thought of representing the local variation and the regional variation. An extreme version of this is obtained when $g(x, y)$ is taken to be the Dirac delta function, i.e., is 1 for $x = y$ and zero otherwise. It is easy to see that the radial basis function estimator coincides with the data values at data locations but otherwise are given by the trend surface, David and Marcotte and David (1988). Clearly the radial basis function interpolator is not smoothed in that case even though the defining condition, eq (11), seems to suggest only smoothing.

5 THE DUAL FORM

Using the solution given in eq(16) the radial basis function interpolator can be written in a different form

$$F^*(x) = [G_0 \quad F_0] [B \quad A]^T \quad (17)$$

where G_0 is the vector with entries $g(x, x_i)$ and F_0 is the vector with entries $f_j(x)$ ($i = 1, \dots, n, j = 0, \dots, p$). Substituting the solution and noting that $F^*(x)$ is scalar valued

$$F^*(x) = [\lambda \quad \mu] [Y \quad 0]^T = \sum_{i=1}^n \lambda_i(x) y_i \quad (18)$$

where $[\lambda \quad \mu]^T$ is the solution of the system

$$\begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} G_0 \\ F_0 \end{bmatrix} \quad (19)$$

Of course the entries in λ, μ depend on x . Recall that positive definite functions are essentially (auto) covariances and the quadratic form given in eq(7) is the same as the variance of the linear combination

$$\sum_{i=1}^n \lambda_i W_i \quad (20)$$

where the W_i 's are random variables. Note that the W_i 's each have finite variance. Similarly a conditionally positive definite function is the negative of a generalized covariance, i.e., it can be used to compute the variance of a linear combination of random variables where the sum of the coefficients is zero and the random variables need not have finite variance. This stochastic formulation suggests a way to deal with smoothing and interpolation at the same time. In deriving the system of equations for the coefficients of the radial basis function interpolator it was assumed that the data were values of the function to be interpolated. Now we assume that the data are the values of the function to be interpolated plus an error term. That is, we use the model given in eq (1) and impose statistical conditions on it. We treat $F(x)$, the unknown function, as a random function with covariance function $g(x, y)$. The error term is assumed spatially uncorrelated and uncorrelated with $F(x)$. The covariance of the error term is the Dirac delta function. The covariance function of $Y(x)$ is given by

$$g(x, y) + \sigma^2 \text{ if } x = y$$

and by $g(x, y)$ if $x \neq y$. The system in eq (19) is slightly modified to become

$$\begin{bmatrix} G + \sigma^2 I & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} G_0 \\ F_0 \end{bmatrix} \quad (21)$$

Reversing the steps for proceeding from eq (10) to eq (18) we obtain the radial basis function interpolator with smoothing. The parameter σ^2 is analogous to the smoothing parameter in the formulation of the smoothing spline. The smoothing spline is a special case of this formulation but with a particular generalized covariance.

6 INTEGRATION AND SMOOTHING

Let L be a grid of points superimposed on the region of interest and suppose that u_{st} are the centers of the cells determined by the grid, $s = 1, \dots, m, t = 1, \dots, q$. Then one way to smooth the values at the centers of the cells is to replace the value at each center by the average over its cell. There are two problems to be resolved, (1) it is likely that the centers are not data locations, i.e., the values at the centers are not (all) known, (2) the values at other points in the cell (values to be averaged) are also likely not known. However the formulation given in eq(18) allows estimation of these averages using only the data locations that are given (which may or may not be within a cell of interest). Let V_{st} be the cell centered at u_{st} , then the average is given by

$$A_{st} = \frac{1}{V_{st}} \int_{V_{st}} F(x) dx \quad (22)$$

$$A_{st}^* = \sum_{i=1}^n \lambda_i(V_{st}) y_i \quad (23)$$

and the coefficients are obtained as the solution of a system almost the same as in eq(1), except that the entries on the right hand side, in G_0 , must be replaced by integrals

$$g(x_i, V_{st}) = \frac{1}{V_{st}} \int_{V_{st}} g(x_i, w) dw \quad (24)$$

Note that

$$A_{st}^* = \frac{1}{V_{st}} \int_{V_{st}} F^*(x) dx \quad (25)$$

Now we note that it is not necessary to choose a grid but rather only to choose a cell size and orientation, while it may not be practical to consider computing an average such as in eq (25) for each point in a region the properties of the smoothed function are still of interest.

7 MEASURES OF SMOOTHING

Suppose that the region of interest V is the union of congruent, disjoint (except for common edges) rectangles, V_{st} (See the description of the grid above). Let

$$A_V = \frac{1}{V} \int_V F(x) dx \quad (26)$$

A_V then is the average value of $F(x)$ over the region V . Obviously

$$A_V = \frac{1}{V} \sum_{s=1}^m \sum_{t=1}^q V_{st} A_{st} \quad (27)$$

A measure of the smoothness of the original (but unknown) function is given by

$$\begin{aligned} S^2 &= \frac{1}{V} \int_V [F(y) - A_V]^2 dy \\ &= \sum_{s=1}^m \sum_{t=1}^q \frac{1}{V} \int_{V_{st}} [F(y) - A_V]^2 dy \end{aligned} \quad (28)$$

S^2 is analogous to a regional variance (spatial rather than ensemble). Similarly let

$$S_{st}^2 = \frac{1}{V_{st}} \int_{V_{st}} [F(y) - A_{st}]^2 dy \quad (29)$$

then S_{st}^2 is variance within the rectangle V_{st} . Substituting from (27) we obtain

$$\begin{aligned} \int_{V_{st}} [F(y) - A_{st}]^2 dy &= \int_{V_{st}} [F(y) - A_{st} + A_{st} - A_V]^2 dy \\ &= \int_{V_{st}} [F(y) - A_{st}]^2 dy + V_{st} [A_{st} - A_V]^2 \end{aligned} \quad (30)$$

The cross-product terms are all zero. Let v denote a generic rectangle, i.e., congruent to all the V_{st} but centered at the origin. Then the volume of v = volume of V_{st} and $V = mqv$ then (29) becomes

$$\begin{aligned} &\frac{1}{V} \int_{V_{st}} [F(y) - A_V]^2 dy \\ &= \frac{1}{mq} \sum_{j=1}^m \sum_{k=1}^q S_{jk}^2 + \frac{1}{mq} \sum_{j=1}^m \sum_{k=1}^q [A_{jk} - A_V]^2 \end{aligned} \quad (31)$$

Hence the total variance is the average of the local variances plus the variance of the average values, since the total variance is fixed as one of the two terms on the right of (31) increases the other must decrease. As shown in Myers (1997), by applying expected values to (31) we obtain

$$g(V, V) = g(0, v) + g(v, V) \quad (32)$$

where

$$g(V, V) = \frac{1}{V^2} \int_V \int_V g(x, y) dx dy \quad (33)$$

$$g(0, v) = \frac{1}{v^2} \int_v \int_v g(x, y) dx dy \quad (34)$$

and

$$g(v, V) = \frac{1}{vV} \int_v \int_V g(x, y) dx dy \quad (35)$$

If instead of simply associating an average over a cell with its center we use a moving cell, i.e., for each point x in V we let

$$A_{v(x)} = \frac{1}{v} \int_{v(x)} F(x) dx \quad (36)$$

where $v(x)$ is a cell centered at x . Then the smoothness of the averaged function is

$$S_v^2 = \frac{1}{V} \int_V [A_{v(x)} - A_V]^2 dx \quad (37)$$

The smoothness of the original function is given by (28). The degree of smoothing is quantified nby

$$S_{vF}^2 = \frac{1}{V} \int_V [A_{v(x)} - F(x)]^2 dx \quad (38)$$

By using the Cauchy-Schwartz in equality it is easy to show that

$$[S_v - S]^2 \leq S_{vF}^2 \leq [S_v + S]^2 \quad (39)$$

As the volume of $v(x)$ tends to zero S_v^2 tends to S^2 and similarly as the volume tends V , S^2 tends to zero. Note that in practice where V is a bounded region we have to consider the intersections of the $v(x)$ and V which near the boundaries of V may be proper subsets of $v(x)$, i.e., there will likely be less smoothing near the boundary.

References

- [1] Cressie, N., Reply to Wahba's letter, *American Statistician*, **44**, pp. 256-258, 1990
- [2] Marcotte, D. & David, M., Trend surface analysis as a special case of IRF-k kriging, *Math. Geology*, **20**, pp. 821-824, 1988
- [3] Matheron, G., The intrinsic random functions and their applications, *Advances in Applied Probability*, **5**, pp. 439-468, 1973
- [4] Matheron, G., Splines and kriging: Their formal equivalence, In Syracuse University Geology Contribution 8, D. Merriam (ed), Dept. Of Geology, Syracuse University, 1981
- [5] Micchelli, C., Interpolation of scattered data: distance matrices and conditionally positive definite functions, *Constructive Approximation*, **2**, pp. 11-22, 1986
- [6] Myers, D.E., Statistical models for multiple scaled analysis, In *Scaling of Remote Sensing Data*, D.A. Quattrochi and M. Goodchild (eds), Lewis Publishers, pp. 273-294, 1997

- [7] Myers, D.E., Statistical methods for interpolation of spatial data, *J. Applied Science and Computations*, **1**, pp. 283- 318, 1994
- [8] Myers, D.E., Spatial interpolation: an overview, *Geoderma*, **62**, pp. 17-28, 1994
- [9] Myers, D.E., Cokriging, radial basis functions and the role of positive definiteness, *Computers Math. Applications*, **24**, pp. 139-148, 1992
- [10] Myers, D.E., Selection of a radial basis function for data interpolation. in *Advances in Computer Methods for Partial Differential Eq. VII*, R. Vichnevetsky, D. Knight and G.Richter (eds), IMACS, pp. 553-558, 1992
- [11] Myers, D.E., On variogram estimation, in *Proceedings of the First Inter. Conf. Stat. Comp., Cesme, Turkey, 30 Mar.-2 April 1987*, Vol II, American Sciences Press, pp. 261-281, 1991
- [12] Myers, D.E., Interpolation and estimation with spatially located data, *Chemometrics and Intelligent Laboratory Systems*, **11**, pp. 209-228, 1991
- [13] Myers, D.E., Interpolation with Positive Definite Functions. *Sciences de la Terre*, **28**, 251-265, 1988
- [14] Powell, M.J.D, Radial basis functions for multivariate interpolation: a review. In *Algorithms for the Approximation of Functions and Data*, M.G. Cox and J.C. Mason (eds), Oxford University Press, 1985
- [15] Wahba, G., Smoothing noisy data with spline functions, *Numer. Math.*, **24**, pp. 383-393, 1975
- [16] Zhang, R., Warrick, A. & Myers, D.E., Variance as a function of sample support size, *Math. Geology*, **22**, pp. 107-121, 1990.