SIAM J. SCI. STAT. COMPUT.
Vol. 12, No. 6, pp. 1304–1313, November 1991

004

COMPUTATION OF THIN-PLATE SPLINES*

ROBIN SIBSON† AND G. STONE†‡

Abstract. Thin-plate splines are an attractive method for interpolating and smoothing arbitrarily spaced points in the plane. A major problem in the application of thin-plate splines is that their computation involves the solution of a linear system that is ill-conditioned for large data sets. A technique for preconditioning the system is proposed that greatly extends the usefulness of thin-plate splines.

Key words. bivariate splines, thin-plate splines, spatial regression, nonparametric regression, ill-conditioned matrices, preconditioning, Dirichlet tessellation

AMS(MOS) subject classifications. 65U05, 65D07

1. Introduction.

1.1. Background. Splines have been used in a wide variety of applications for both interpolation and smoothing. Data analysts have found cubic splines to be a flexible and efficient alternative to parametric regression functions. Silverman (1985) gives an account of nonparametric regression based on cubic splines, and full details and many references may be found in his paper.

The success of the cubic spline approach in the one-dimensional case suggests that higher-dimensional extensions of it should be sought. In this paper we concentrate on a particular variational generalisation, one that provides a regression method independent of choice of origin and coordinate axes. The technique is therefore suitable for application to spatial regression problems. There are many papers on this topic in the literature; major references on the general problem are Duchon (1976a), (1976b) and Meinguet (1979a), (1979b). The extension to two dimensions has been carried through in a fairly direct way, to produce the method of *thin-plate splines*, discussed in a data-analytic context, including a treatment of smoothing, by Wahba (1979a); followed up by Wahba (1979b), (1980a), (1980b), (1981), (1983); Dyn and Wahba (1979); Wahba and Wendelberger (1980); Wendelberger (1981), (1982); and Bates and Wahba (1983).

There are three major differences between one-dimensional cubic splines and their two-dimensional counterparts, and one of these makes the computation much more laborious in two dimensions. It is the improvement of the computational procedures involved that concerns us in this paper.

1.2. Notation and mathematical formulation. We consider initially the generalisation to K dimensions. We shall use the term *data sites* for the points in K-space at which observations have been made, to emphasise that it may not be possible to exercise complete control over what these positions actually are. The data sites will be denoted as column coordinate vectors in some particular coordinate system by u_n for $n = 1, \cdots, N$. The observed values at the data sites will be called *data values*, and will be denoted by Z_n. Denote by z_n the fitted values at the data sites; then the usual

* Received by the editors December 4, 1989; accepted for publication (in revised form) September 7, 1990. This research was partially supported by a research grant from the United Kingdom Science and Engineering Research Council.

† School of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom.

‡ The work of this author was done while he was supported by a Science and Engineering Research Council Research Studentship.

residual-sum-of-squares error penalty is

$$E = \sum_{n=1}^{N} (Z_n - z_n)^2, \tag{1.2.1}$$

which arises in the same form and for the same reasons as in many other contexts, and might similarly be generalised.

A typical function on K-space will be denoted by z, and if the fitted values arise from such a function then z_n is simply $z(u_n)$. In one dimension, following the usual discussion of the cubic spline, the roughness of such a function is the squared second derivative, integrated over the whole line. This situation corresponds to a well-known physical model for the deformation of a thin stiff wire. In higher dimensions, a direct generalisation is obtained as follows. Write

$$R(u) = \sum_{i=1}^{K} \sum_{j=1}^{K} \left[\frac{\partial^2 z}{\partial u^i \, \partial u^j} (u) \right]^2 \tag{1.2.2}$$

and call this the *local roughness* at u, and then define the *roughness* to be

$$R = \int_{R^K} R(u) \, du, \tag{1.2.3}$$

where the notation is to be understood as implying a K-dimensional multiple integral.

Note that, although the second derivative (Hessian) matrix is symmetric, off-diagonal terms have to be counted both above and below the diagonal to achieve rotational invariance for the local roughness. The roughness then has the essential property of invariance under the choice of a coordinate system. In two dimensions this penalty again arises as the strain energy in a physical model, in this case for the infinitesimal deformation of a thin stiff lamina. It is from this model that the name "thin-plate splines" derives.

The interpolation problem is the problem of minimising the roughness penalty R subject to the exact interpolation conditions $z(u_n) = Z_n$. For the smoothing problem, a total penalty $E + \omega R$ (where $\omega > 0$ is the *smoothing parameter*) has to be minimised without constraints.

1.3. General form of the solution. It can be shown that the appropriate solution to the above variational problem has the form

$$z(u) = a + b^T u + \sum_{n=1}^{N} c_n \varphi(u - u_n), \tag{1.3.1}$$

where φ is a fundamental solution, given (see Meinguet (1979b)) by

$$\varphi(u) = \begin{cases} (12)^{-1} |u|^3 & \text{in one dimension,} \\ (16\pi)^{-1} \|u\|^2 \log \|u\|^2 & \text{in two dimensions,} \\ (-8\pi)^{-1} \|u\| & \text{in three dimensions.} \end{cases} \tag{1.3.2}$$

In each case, the definition of the fundamental solution is for the whole space excluding the origin, but may be extended by continuity to the whole space. As a basis for nonparametric regression, the variational problem from which the one-dimensional cubic spline arises accordingly generalises satisfactorily to two dimensions, but with (as the first major difference we identify) a drop from twice to once continuous differentiability in the solution. In three dimensions the solution is no better than continuous, and is not differentiable at the data sites. The generalisation to three

dimensions is accordingly unacceptable as the basis for a data-analytic smoothing or interpolation method. In four or more dimensions the generalisation collapses completely, because the fundamental solution cannot be extended to the origin by continuity.

The only suggestion known to us for making the higher-dimensional extensions of the variational problem well-posed is to modify the roughness penalty by replacing the second derivatives by higher-order derivatives. Wahba and Wendelberger (1980) give an example of the practical use of such functions in a meteorological problem. The difficulty with this approach is that the class of penalty-free functions is extended to include polynomials of degree greater than one, and it seems to us to be unsatisfactory not to penalise such functions as possessing some degree of roughness. Possibly a generalisation involving second- and higher-order derivatives together in the roughness penalty might provide what is wanted, but as far as we are aware this approach is unexplored.

The two-dimensional case accordingly has a special status as the only immediately acceptable generalisation to higher dimensions of the variational problem yielding the cubic spline in one dimension, and the remainder of this paper is concerned exclusively with bivariate splines.

2. Thin-plate splines over the whole plane.

2.1. The linear system. In the bivariate case the expression (1.3.1) has the form

$$z(u) = a + b^T u + \sum_{n=1}^{N} c_n (16\pi)^{-1} \|u - u_n\|^2 \log \|u - u_n\|^2 \tag{2.1.1}$$

with $(N+3)$ parameters. Where convenient, we can write $u = [x\, y]^T$, $u_n = [x_n, y_n]^T$. For interpolation, the parameters in (2.1.1) are determined by the interpolation conditions

$$z(u_m) = Z_m \quad \text{for } m = 1, \cdots, N \tag{2.1.2}$$

and by the three additional equations

$$\sum c_n = 0, \quad \sum c_n x_n = 0, \quad \sum c_n y_n = 0, \tag{2.1.3}$$

which arise from boundary conditions. Equations (2.1.2) and (2.1.3) may be written in the form

$$\begin{bmatrix} K & T \\ T^T & 0 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} Z \\ 0 \end{bmatrix}, \tag{2.1.4}$$

where K is an $N \times N$ matrix whose (m, n)th entry is $(16\pi)^{-1} \|u_m - u_n\|^2 \log \|u_m - u_n\|^2$, and that is accordingly symmetric, where T is an $N \times 3$ matrix whose nth row is $[1\ x_n\ y_n]$, and where $d = [a\, b^T]^T$. The entire system is thus symmetric. The system of equations arising for the smoothing problem is similar:

$$\begin{bmatrix} K + \rho I_N & T \\ T^T & 0 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} Z \\ 0 \end{bmatrix}, \tag{2.1.5}$$

where I_N is the $N \times N$ unit matrix, and $\rho = N\omega$. The scale on which the smoothing parameter is measured is seldom of interest, so a fairly cavalier attitude towards the choice of scale is harmless.

2.2. The use of B-splines. The analogous system in one dimension is of course of size $(N+2)$, and the boundary conditions are equivalent to a requirement of first-degree tail-out at either end of the range of the data sites. The structure of the system, in the interpolation case in both one and two dimensions, is that the diagonal entries are zero, and all off-diagonal entries in the main $N \times N$ block are in general nonzero

provided that data sites are distinct. In one dimension the solution procedure always takes advantage of some transformation which avoids this problem of nonsparsity, for example a change of basis from the original fundamental solutions to a basis of B-splines. The solution procedure, as described by Silverman (1985), depends only linearly on the number of data sites.

The terminology of "B-splines" has been applied to a number of constructions of functions of compact support over the plane. These appear to be either tensor products of univariate B-splines (de Boor (1978), Hu and Schumaker (1986)), or piecewise polynomial functions (Dahmen (1980), (1981)); they are certainly not special thin-plate splines in the way that univariate B-splines are special cubic splines.

In fact, thin-plate B-splines in the true sense do not exist; that is, a thin-plate spline cannot have compact support unless it is identically zero. This result, which is an immediate consequence of familiar theorems of analysis, applies to all types of "radial basis function" as reviewed generally by Powell (1985) and discussed specifically by Hardy (1971), Duchon (1976a), and Dyn and Levin (1983). This is the second major difference which we identify between cubic splines and thin-plate splines. It is not essential to have basis functions of compact support to achieve a sparse linear system, but in their absence it is necessary to build basis functions which are zero at all but a few data sites. This is possible, but is equivalent to solving the interpolation problem with a selection of specially chosen sets of data values, and thus represents a reformulation of the problem rather than a solution of it.

The nonexistence of B-splines, and the implication that there will be no effective way to reformulate the linear system (2.1.4) or (2.1.5) in sparse form, is the second major difference encountered in the move from one to two dimensions. It fundamentally alters the computational character of the problem. The third major difference arises from the imposition of a general "window" Ω in place of the whole of the line or plane as the domain of integration for the local roughness (Dyn and Levin (1982)). It is well known that in one dimension Ω may be taken as any interval containing all the data sites without altering the problem. For reasons similar to those preventing the construction of B-splines, in two dimensions the choice of window always affects the problem. We shall discuss the computational and practical implications of this in a further paper, but in this paper we limit discussion solely to the case where Ω is the whole plane.

2.3. Solution procedure for the linear system. The interpolation system (2.1.4) is the special case of (2.1.5) with $\rho = 0$. Wahba (1979a) deals with the more general problem which arises in an arbitrary number of dimensions and with a higher order of differentiation; in such a case the matrix corresponding to T has in general more than three columns. Wahba suggests the following procedure. First, construct the projection matrix $(I_N - T(T^TT)^{-1}T^T)$ and find its spectral decomposition. The eigenvectors corresponding to the unit eigenvalues form an orthonormal basis for the orthogonal complement of the space spanned by the columns of T.

In the specific case of interest here, that of the true thin-plate spline in two dimensions where T has three columns, a much more efficient procedure is available, as follows. First find a large triangle of data sites to serve as a reference triangle. It need not actually have maximal area, and a good enough algorithm is to start with any triangle of data sites and successively replace each vertex by that data site which maximises the area of the triangle which it makes with the other two, until no further improvement can be made. Without loss of generality, these data sites may be renumbered as 1, 2, 3. Next, construct vectors corresponding to the data sites $4, \cdots, N$ which

are not vertices of the reference triangle. The vector arising from data site n has entry -1 in position n and its entries in positions 1, 2, and 3 are the homogeneous coordinates obtained when data site n is expressed as an affine combination of the three data sites selected to form the reference triangle. All its remaining entries are zero.

All such vectors automatically lie in the orthogonal complement of the subspace spanned by the columns of T, and are independent, and there are $(N-3)$ of them, so they form a basis for it. The usual Gram-Schmidt procedure can be applied to orthonormalise this basis very efficiently. The relevant inner products are sums of four terms only, except at the normalisation step. The column vectors so determined form an $N\times(N-3)$ matrix which we shall denote by Q. Wendelberger (1981) suggested a rather different modification of Wahba's original procedure which has a similar objective to ours, of avoiding a spectral decomposition step at this stage, but which is designed to cope with the general case rather than to be very efficient for the true thin-plate case.

At this point we rejoin Wahba's procedure. Since $T^Tc=0$, we can write $c=Qe$ for some (unique) $(N-3)$-vector e, and substituting this in the equation $(K+\rho I_N)c+Td=Z$ and premultiplying by Q^T (which annihilates the second term) yields

$$Q^T(K+\rho I_N)Qe=Q^TZ \tag{2.3.1}$$

and on rearranging,

$$(Q^TKQ+\rho I_{(N-3)})e=Q^TZ, \tag{2.3.2}$$

which is a symmetric linear system of size $(N-3)$. Once this system has been solved for e, then c can be obtained directly from it, and d is then found by back substitution and solution of a system which reduces to 3×3. The importance of making the columns of Q orthonormal is now apparent. It ensures that the second term in the brackets is a multiple of the unit matrix, so all that is now needed to solve the system *for any* ρ, is to form the spectral decomposition of the matrix $B=Q^TKQ$. This is the matrix that arises in any case in the interpolation problem. Assuming all the data sites are distinct, and that they are not all collinear, B is a positive-definite symmetric matrix.

2.4. The role of the condition number. The limitations on the scale of the problem that can be solved are imposed both by time and storage considerations, and also by conditioning. In the smoothing problem, a multiple of the unit matrix is added to B, so we need be concerned only with the conditioning of B, which is the worst case. In the present context of a positive-definite symmetric matrix an appropriate measure of conditioning is the ratio of the largest to the smallest eigenvalue. Wahba (1979a), probably writing in the context of some form of single-precision arithmetic, suggests that the largest acceptable condition number in this sense may be about 10^6 or 10^7, and this view has not to our knowledge been seriously challenged. We would suggest that this may be overly optimistic for IEEE single-precision arithmetic, where we believe that a value of 10^5 may be a prudent bound; whereas for IEEE double-precision arithmetic very much larger values, perhaps up to about 10^{12}, should be acceptable.

There are two factors contributing to ill-conditioning. The first is the steady increase in condition number as the number of data sites increases, however benign their positioning. A useful quantitative picture of this situation is given by the condition number for a square grid of data sites, for grids of various sizes. For a 7×7 grid (49 data sites), it is about 250; for a 10×10 grid (100), about 1100; for a 14×14 grid (196), about 4200; for a 20×20 grid (400), about 18,000. This is in line with the theoretical result that the condition number for this particular layout of data sites increases approximately in proportion to the square of the number of data sites (Wahba (1979b),

Cox (1984), Utreras (1988)). On this basis a 30×30 grid would give a condition number around our suggested safe limit of 10^5 for single-precision arithmetic. Double-precision arithmetic should be able to handle any scale of problem on a square grid for which storage and time are available.

The second is the occurrence of close pairs of data sites. In the limit as two data sites move towards coincidence, the corresponding rows and columns of the matrix K tend to equality and the matrix B tends to singularity through cases with a small eigenvalue and consequently a large condition number. When randomly scattered data sites are generated for test purposes, the condition number obtained can fluctuate wildly. In a small number of replications we obtained one configuration of 100 points with a condition number below 5000 and another with a condition number exceeding 500,000. In such circumstances there is no "typical" condition number, and the conditioning of each problem has to be checked individually for acceptability, even if double-precision arithmetic is to be used. No guarantees can be given purely on the basis of the number of data sites. Dyn, Levin, and Rippa (1986) mention a limit of about 150 data sites "due to computational difficulties," which by implication are difficulties arising from ill-conditioning; their examples are all on highly structured evenly spaced systems of data sites, so they are probably not considering the effect of close points.

Strictly speaking, ill-conditioning could also arise from a tendency for the data sites all to approach collinearity; this is not a problem in practice and we shall ignore it.

2.5. Scaling the linear system. Changing scale isotropically in the plane, and separately changing the scale on which the data values are measured, has the same effect as varying the parameters α and β in the more general form

$$\varphi(u) = \alpha \|u\|^2 \log \beta \|u\|^2 \tag{2.5.1}$$

of the fundamental solution. Apart from an overall choice of the scale of B, and associated possible reinterpretation of the meaning of actual numerical values of the smoothing parameter, all of these choices leave B unaffected, except that changing Q for another matrix with the same properties will pre- and postmultiply B by an orthogonal matrix, which does not affect its spectrum, and hence does not alter its condition number. This is immediately obvious except for the effects of varying β, where the property depends on the orthogonality of the columns of Q to the columns of T.

These remarks are not true of the matrix on the left-hand side of (2.1.4) or (2.1.5), which is not even a positive-definite matrix; the actual spectral structure of this matrix is affected by rescaling, and by the way the columns of T are chosen to span the space of first-degree functions evaluated at the data sites. It is accordingly not very meaningful to quote values for the condition number of the matrix occurring in (2.1.4), as Dyn, Levin, and Rippa (1986) appear to have done.

The arithmetic performed on values of the fundamental solution in the calculation of B is most stable when its values do not have very large magnitude within the region occupied by the data sites, neither are they uniformly very small across it. In applications where the natural unit of scale leads to a coordinate range over the axes which is not of the order of magnitude of unity, for example some cartographic applications, then it may well be worth choosing a new scale for the coordinates which does achieve this, with a consequent change of scale for the smoothing parameter, whose interpretation does of course depend on these choices of scale.

2.6. Solution procedures in the ill-conditioned case. We have seen that ill-conditioning is likely to cause difficulties primarily where there are pairs of data sites close together. It can be argued that in such a case, exact interpolation is sensible only if the values at the two close data sites are themselves close, so the eigenvector corresponding to the small eigenvalue will make little contribution to the solution and it may be safer to ignore it completely at the cost of slightly failing to satisfy the interpolation conditions. A similar effect can be achieved by replacing the interpolation problem by a smoothing problem with a smoothing parameter which is small, but substantially larger than the small eigenvalue(s). Because what is being done is more readily interpretable, this may well be a preferable approach. Various other forms of approximation technique, oriented both to reducing the overall burden of computation as well as to the problem of conditioning, have been suggested by Bates and Wahba (1983) and Dyn, Levin, and Rippa (1986).

If the problem is actually an ill-conditioned smoothing problem, then it is only at small values of the smoothing parameter that any difficulty over condition number will become apparent. If there is a positive lower bound ρ_0 to the range of values of the smoothing parameter ρ that are of interest, then $B+\rho I$ can be written as $(B+\rho_0 I)+(\rho-\rho_0)I$ and the numerical linear algebra can be carried through with $B+\rho_0 I$ replacing B and $\rho-\rho_0$ replacing ρ, thereby reducing the condition number.

If it is really necessary to solve an ill-conditioned interpolation problem, then it is useful to recall that the only reason for insisting on orthonormality of the columns of Q is so as to keep a multiple of the unit matrix as the term arising from smoothing. If that term is absent, all that is necessary is to ensure that the columns of Q are independent, and are orthogonal to those of T, and they may be chosen in an attempt to improve the conditioning of B. Some of the numerical difficulties removed from the spectral decomposition problem by such preconditioning may reappear elsewhere in the arithmetic, but will generally be less harmful there.

2.7. A preconditioning method. We have developed a preconditioning method of the above type, where the matrix Q is replaced by a matrix R whose $(N-3)$ columns are merely required to be orthogonal to the columns of T, and independent. R is constructed as follows.

First a working "window" is placed round the configuration of points. This is required simply to provide a bounded convex polygonal region of the plane in which to work; the technique proposed is not sensitive to its choice. It may be conveniently taken to be a rectangular region whose sides are aligned with the coordinate axes and which has a reasonable clearance from the data sites, say, 10 percent of their spread.

The Dirichlet tessellation of the data sites within this window is then calculated. This is the subdivision of the window into convex polygonal *tiles*, one for each data site, such that the tile of a data site comprises that part of the plane nearer to that data site than to any other. An efficient computational procedure for finding Dirichlet tessellations has been proposed by Green and Sibson (1978). The *neighbours* of a data site in the tessellation are those other data sites whose tiles share a common boundary with its own. The strength of this neighbour relationship can be measured by assigning a weight to each pair which is zero for nonneighbours and is the length of the common boundary of their tiles divided by the distance apart for neighbours. These boundary-over-distance weights are symmetric, and are large for close neighbours because of the inverse-distance factor. They also have the property that each data site is the weighted centroid of its neighbours. This *local balance* property is trivial to prove for the Dirichlet tessellation in the plane; it was proved in the K-dimensional case by

Christ, Friedberg, and Lee (1982) from the divergence theorem, and can alternatively be proved from a similar result of Sibson (1980). Boundary-over-distance weights appear to have been first used for statistical purposes by Cliff and Ord (1969), but that application made no mention or use of local balance.

The columns of R are then constructed for the $(N-3)$ data sites which are not vertices of the reference triangle selected as in § 2.3. Each column is first zeroed, and then for each neighbour of the current data site, the boundary-over-distance weight is added to the entry for that neighbour and subtracted from the entry for the current data site itself. Orthogonality to the columns of T is guaranteed by the choice of weight for the current data site, and by local balance.

For data sites which are contiguous to the window edge, a special procedure has to be followed. Each such contiguity is treated as a neighbour relationship with a virtual point which is the reflexion of the current data site in that straight line segment of the window. The boundary-over-distance weight for the virtual point is then split between the data sites which are the vertices of the reference triangle in proportion to its homogeneous coordinates with respect to them. This retains the orthogonality property to the columns of T.

In either case, the column is finally given an appropriate normalisation by division by the tile area. Apart from this normalisation step, the procedure differs from the original construction set out in § 2.3 by expressing each data site as a convex combination of its neighbours rather than as an affine combination of the vertices of the reference triangle, these being used only if they happen to occur as neighbours, or if it is necessary to reexpress a virtual point (when negative weights may arise).

This procedure sounds complicated, but is in fact computationally very efficient. The motivation for it is that it generates linear combinations of the original fundamental solution which, in a rather crude sense, correspond to the application of the Laplacian operator. Large weights arise for close pairs, and this has precisely the appropriate effect to remove the ill-conditioning arising from such pairs, any numerical difficulty being transferred to the process of forming the linear combinations. In fact this "discrete Laplacian" is effectively applied twice over, because the matrix K is premultiplied by R^T as well as postmultiplied by R. If the Laplacian operator itself were applied twice to the fundamental solution, then a delta-function would result; the double application of its discrete analogue has the effect of producing a matrix in more nearly diagonal form, and the results of the preconditioning procedure are accordingly beneficial even in the case of regularly spaced data sites. This idea, of applying a discrete Laplacian to the original basis functions, was proposed for gridded data sites by Dyn and Levin (1983) and generalised for scattered data sites by Dyn, Levin, and Rippa (1986); the latter paper even used the Delaunay triangulation, the dual of the Dirichlet tessellation, but did not use boundary-over-distance weights and their properties.

This procedure has a dramatic effect on the condition number of B. For the regular square grid, it becomes 52 for 7×7, 72 for 10×10, 110 for 14×14, and 170 for 20×20. For 100 scattered points, it ranges from 35 to 95 on cases which include those for which the raw condition number is below 5000 and is above 500,000. Thus the condition number is reduced greatly, and grows more slowly with the number of data sites, even in the regular case; in the scattered case the bad properties arising from close pairs are eliminated; and the regular case now appears as typical within the range of scattered cases.

2.8. Limits on problem size. Even with single-precision arithmetic, it is clear that the limitation on the scale of the problem that can be tackled is imposed by storage

and time limitations rather than conditioning once this preconditioning procedure has been applied, and this is true as much for scattered data sites as for regularly spaced ones.

Two matrices, of size approximately $N \times N$, must be stored, and this is the dominant component of the storage requirement; in single-precision arithmetic this requires $8N^2$ bytes, so for example 8 Mbytes would be required for 1000 data sites; at the time of this writing, this is well within scientific workstation capabilities, but a problem on 10,000 data sites would tax the capabilities of a large mainframe.

The dominant component of the workload is the computation of the spectral decomposition. The time taken by the procedure we use increases as N^3 for large N. We estimate that it would take about four hours on a 1000×1000 matrix on the computer used for our tests (Sun-4/260); whilst no doubt absolute timing will become out of date as quickly as usual, the rapid rate of growth of both the storage and the time workload with N suggests that the practicable limit of scale is going to be below 10,000 data sites for some time to come, even allowing for the possible introduction of alternative algorithms taking advantage of parallel processing technologies.

Methods of solution which avoid the use of the spectral decomposition may well be worth exploring. Solution of (2.3.2) is possible by a variety of methods both direct and iterative, and where only a single vector Z of observed values, and a single value of the smoothing parameter are of interest, then a considerable gain in efficiency should be possible. Dyn, Levin, and Rippa (1986) discuss a variety of such methods, and conclude that the conjugate gradient approach is superior to the alternatives considered in their tests. But such methods still rapidly become laborious for large problems.

Practical problems often arise with many more than 10,000 data sites; for example, in aeromagnetic survey work it is common to have 50,000 to 100,000 observations in a single data set. We believe that such problems will indefinitely remain beyond the scope of thin-plate splines. We plan to report work on the development of splinelike methods for large-scale problems (and for the rapid solution of medium-scale problems) in future papers.

3. Conclusions. Thin-plate splines have been shown to form, in a suitable mathematical sense, the natural multidimensional generalisation of cubic smoothing splines. We have demonstrated why this generalisation breaks down without further modification beyond two dimensions. Focusing our attention on the two-dimensional case, we have identified three major differences between the behaviour of cubic splines and thin-plate splines. The lack of a sparse reformulation of the linear system characterising the solution is fundamental to the nature of the two-dimensional problem. It completely changes its computational character compared with that of the cubic spline problem. We have identified the sources of ill-conditioning in the linear system, and proposed a preconditioning procedure which eliminates ill-conditioning as a source of computational difficulty in the interpolation problem, which is the worst case. The use of this procedure, together with other appropriate numerical methods that we identify, extends the scale on which thin-plate splines can be applied without undue difficulty to data sets with 1000 observations.

REFERENCES

D. M. BATES AND G. WAHBA (1983), *A truncated singular value decomposition and other methods for generalized cross-validation*, Tech. Report 715, Department of Statistics, University of Wisconsin, Madison, WI.

N. H. CHRIST, R. FRIEDBERG, AND T. D. LEE (1982), *Weights of links and plaquettes in a random lattice*, Nuclear Phys. B, Field Theory Statist. Sys., 210, pp. 337–346.

A. D. CLIFF AND J. K. ORD (1969), *The problem of spatial autocorrelation*, in Studies in Regional Science, A. J. Scott, ed., Pron., London, pp. 25-55.

D. D. COX (1984), *Multivariate smoothing spline functions*, SIAM J. Numer. Anal., 21, pp. 789-813.

W. DAHMEN (1980), *On multivariate* B*-splines*, SIAM J. Numer. Anal., 17, pp. 179-191.

——— (1981), *Approximation by linear combinations of multivariate* B*-splines*, J. Approx. Theory, 31, pp. 299-324.

C. DE BOOR (1978), *A Practical Guide to Splines*, Springer-Verlag, Berlin, New York.

J. DUCHON (1976a), *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, RAIRO Anal. Numér., 10, pp. 5-12.

——— (1976b), *Fonctions spline du type "Plaque mince" en dimension* 2, RR231, Mathématiques Appliquées, Université Scientifique et Médicale, Grenoble.

N. DYN AND D. LEVIN (1982), *Construction of surface spline interpolants of scattered data over finite domains*, RAIRO Anal. Numér., 16, pp. 201-209.

——— (1983), *Iterative solution of systems originating from integral equations and surface interpolation*, SIAM J. Numer. Anal., 20, pp. 377-390.

N. DYN, D. LEVIN, AND S. RIPPA (1986), *Numerical procedures for surface fitting of scattered data by radial functions*, SIAM J. Sci. Statist. Comput., 7, pp. 639-659.

N. DYN AND G. WAHBA (1979), *On the estimation of functions of several variables from aggregated data*, Tech. Report 1974, Mathematics Research Centre, University of Wisconsin, Madison, WI.

P. J. GREEN AND R. SIBSON (1978), *Computing Dirichlet tessellations in the plane*, The Computer J., 21, pp. 168-173.

R. L. HARDY (1971), *Multiquadric equations of topography and other irregular surfaces*, J. Geophysical Res., 76, pp. 1905-1915.

C. L. HU AND L. L. SCHUMAKER (1986), *Complete spline smoothing*, Numer. Math., 49, pp. 1-10.

J. MEINGUET (1979a), *An intrinsic approach to multivariate spline interpolation at arbitrary points*, in Polynomial and Spline Approximation, B. N. Sahney, ed., D. Reidel, New York.

——— (1979b), *Multivariate interpolation at arbitrary points made simple*, Z. Angew. Math. Phys., 30, pp. 292-304.

M. J. D. POWELL (1985), *Radial basis functions for multivariable interpolation: a review*, presented at the IMA Conference on Algorithms for the Approximation of Functions and Data, RMCS, Shrivenham.

R. SIBSON (1980), *A vector identity for the Dirichlet tessellation*, Math. Proc. Cambridge Philos. Soc., 87, pp. 151-155.

B. W. SILVERMAN (1985), *Some aspects of the spline smoothing approach to non-parametric regression curve fitting* (with discussion), J. Roy. Statist. Soc., Ser. B, 47, pp. 1-52.

F. I. UTRERAS (1988), *Convergence rates for multivariate smoothing spline functions*, J. Approx. Theory, 52, pp. 1-27.

G. WAHBA (1979a), *How to smooth curves and surfaces with splines and cross-validation*, in Proc. 24th Design of Experiments Conference, F. Dressel, ed., Academic Press, New York.

——— (1979b), *Convergence rates of "thin plate" smoothing splines when the data are noisy*, in Smoothing Techniques for Curve Estimation, T. Gasser and M. Rosenblatt, eds., Lecture Notes in Mathematics, Vol. 757, Springer-Verlag, Berlin, New York, pp. 232-245; Tech. Report 557, Department of Statistics, University of Wisconsin, Madison, WI, 1979.

——— (1980a), *Ill posed problems: Numerical and statistical methods for mildly, moderately, and severely ill posed problems with noisy data*, Tech Report 595, Department of Statistics, University of Wisconsin, Madison, WI.

——— (1980b), *Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data*, Tech. Report 595, Department of Statistics, University of Wisconsin, Madison, WI.

——— (1981), *Spline interpolation and smoothing on the sphere*, SIAM. J. Sci. Statist. Comput., 2, pp. 5-16; Tech. Report 584, Department of Statistics, University of Wisconsin, Madison, WI, 1979.

——— (1983), *Bayesian "confidence intervals" for the cross validated smoothing spline*, J. Roy. Statist. Soc., Ser. B, 45, pp. 133-150; Tech. Report 645, Department of Statistics, University of Wisconsin, Madison, WI, 1981.

G. WAHBA AND J. WENDELBERGER (1980), *Some new mathematical methods for variational objective analysis using splines and cross-validation*, Monthly Weather Rev., 108, pp. 1122-1143; Tech. Report 578, Department of Statistics, University of Wisconsin, Madison, WI, 1979.

J. G. WENDELBERGER (1981), *The computation of Laplacian smoothing splines with examples*, Tech. Report 648, Department of Statistics, University of Wisconsin, Madison, WI.

——— (1982), *Smoothing noisy data with multidimensional splines and generalized cross validation*, Ph.D. thesis, Department of Statistics, University of Wisconsin, Madison, WI.