Monografías del Seminario Matemático García de Galdeano **33**, 83–90 (2006)

FITTING SCATTERED DATA IN 2D BY RBF TECHNIQUES

M. Bozzini and L. Lenarduzzi

Abstract. In this note we shall present some problems about fitting scattered data in $2d$ from regular functions, and that, in our opinion, are to be studied more in depth.

Keywords: Multiquadric, scattered data, adaptivity, interpolation.

AMS classification: 65D10, 41A05.

§1. Introduction

The purpose of this paper is to investigate the problems that may arise when we fit a set of two dimensional scattered data by the very popular tool of radial basis functions (RBFs). In particular we want to discuss the most recent techniques of the literature which aim to provide a RBF approximation with two desirable properties: the good quality of reproduction and the stability. These two aspects are strongly related to the data locations, to the data size N, that is commonly said moderate (N of order of tens), standard (N of order of hundreds), large (N of order of thousands), and to the regularity of the underlying function f.

In fact it is known that when we interpolate data coming from a function regular enough, the approximations feature convergence as the points get more dense. On the other hand, the spectral condition of the interpolation matrix (see [16]) is bounded from above by a function of the minimal distance among the point locations, which increases as N gets large.

Moreover, experimentally, one observes that the condition of the interpolation matrix, depends on the configuration of the data points too.

Here we discuss the case of samples with moderate (§2) and standard size (§3), leaving out the case of large size.

In our opinion, the techniques presented here provide partial solutions only to the aforesaid problems and some questions are still open.

In the following we call $X = \{x_i\}$, $i = 1,\dots,N$, the set of separated points in $\Omega \subset \mathbb{R}^2$ of size $N := |X|$ at which the functional values $\{f_i = f(x_i)\}$ are given. We assume that the unknown function f belongs to $C^\alpha(\Omega)$, $\alpha \geq 1$.

Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ be a radial basis function (RBF), that is to say a translation invariant kernel on two arguments $x, y \in \mathbb{R}^2$; in particular we shall consider the scaled multiquadric $\Phi_\delta(x) := -\sqrt{\delta^2 + \|x\|^2}$. We shall denote $A_{X,X}$ the matrix of interpolation associated to the current kernel Φ_δ.

§2. Sample of moderate size

The case of samples of moderate size arises for example in mathematical models in which both a good accuracy and a low computational cost is searched for. An important and relevant case is the solution of elliptical problems when the differential equation is part of an iterative cycle. Problems of this kind occur in the recent modelling within nanotechnologies.

Another important case is connected to real world problems where it is impossible or too expensive to get a large sample. Typical examples can be found in the medical field (heart potential mapping).

We want to remark that in these cases we haven't problem with the conditioning because the points are generally well separated. In what follows we quote two papers. The first is related to mathematical modelling and the second to recovering a function.

The paper [11] deals with elliptical problems. The authors make use of a collocation method with scattered data to approximate the solution by shifts of the multiquadric function.

It is well known that multiquadric functions are radial functions very suitable in the case of regular functions. In a recent paper, [10] (we shall return to this paper in §3), an algorithm is presented to construct a stable multiquadric interpolant even by setting δ very large.

The usage of such a procedure for the solution of a collocation method on the unitary circle, can be used, in the case of functions smooth enough, to obtain accurate approximations. The authors in their paper use this technique with only 50 collocation points for the solution of a Poisson problem with Dirichlet conditions.

By choosing the optimal value of δ as estimated by Rippa's method [15], they get the impressive accuracy of $10\,e(-12)$.

Concerning the recovering of a function f, in [7] it is shown how it is possible to make up, by the modest but sufficient information about the phenomenon, an accurate solution to the problem.

Having few functional evaluations at disposal, the goodness of reconstruction depends on the choice of the basis functions that must be as close as possible to f. If a function of the basis were proportional to f, just one data would be enough to reconstruct f.

So, we consider a set of N linearly independent functions, each one depending on a parameter:

$$D_{\alpha_i} := \{\Psi_i(x,\alpha_i),\ \ x\in\Omega,\ \ \alpha_i\in\mathbb{R}^s,\ \ s\geq 1,\ \ i=1\ldots,N\}.$$

When we prescribe a set of values $\{\hat{\alpha}_i\}$ we determine the basis $\{B_\Psi(\hat{\alpha}_i)\}$ and the corresponding linear space. Then the problem is shifted to the choice of the basis more adherent to the phenomenon.

Let us fix a positive operator $\mathscr{K}$ that provides the measure of adherence to f as $\alpha_i\in D$ varies, where D is the subdomain of $\mathbb{R}^s$ in which the unicity of the interpolating function is guaranteed. Some examples of $\mathscr{K}$ are: the entropy, the risk, the energy and the norm ℓ^p.

For example we consider the multiquadrics with variable shape: at the generic location x_i of our sample set and with the Frenet frame (n,t)

$$\Psi_i(n,t) = \{b_i(n-n_i)^2+(t-t_i)^2+\delta_i^2\}^{1/2},\ \ \alpha_i=(b_i,\delta_i^2)\in D,$$

where

$$b_i = b_i(\kappa) = (1+\kappa\mu_i\nu_i)^{-1},\ \ i=1,\ldots,N;$$

where μ_i is function of the gradient, ν_i is function of the radius of curvature; μ_i and ν_i estimated by the sample. In addition,

- for uniform locations: $\delta_i =$ median (neighbouring distances from x_i);
- with empties of data:

$$\delta_i = \text{mean (neighbouring distances)} + 0.5 * \big(\text{median (neighbouring distances)} - \text{mean (neighbouring distances)}\big).$$

As operator $\mathscr{K}$ we choose the ℓ_2 norm and this will allow to determine the optimal κ . Sufficient theoretical conditions for the unique solvability of such interpolation processes are discussed in [5].

Example 1. We take $N = 67$ data points from the sygmoidal function on $[0,1]^2$. The errors of the tabulated reconstruction are $e_2 = 0.004$ and $e_\infty = 0.026$. On the contrary with the classical Φ_δ, with the suitable value $\delta = 0.1$, the errors are $e_2 = 0.013$ and $e_\infty = 0.067$.

§3. Sample of standard size

The case of N standard is the most considered in the literature.

In the case of scattered data, it is well known that RBFs are a useful tool. In the recent paper [17], the authors show why and how RBFs are useful in various fields of Numerical Analysis including approximation, interpolation and meshless methods for solving partial differential equations.

Of particular interest are the multiquadrics because it was proved that they have exponential convergence when f belongs to the reproducing kernel Hilbert space (RKHS) associated to the multiquadric, see [13] and [16]. Estimates are provided in [14] for the case of f outside of the RKHS.

The main drawback of the multiquadric is that it may lead to bad conditioning of the linear system to be solved. Schaback proved in [16] that that the spectral condition value $\mathscr{K}_2(A_{X,X})$ of the interpolation matrix $A_{X,X}$, is bounded from above by a function of $q(X)/\delta$, where $q(X)$ is the minimal distance among the X locations; such a function increases exponentially as $q(X)/\delta$ diminishes.

There is a tradeoff between accuracy and stability for each radial basis. Many authors have faced the problem by providing different solutions. Here we cite the most recent ones at our knowledge. The already cited paper [10] provides a stable method to calculate (here $\delta = 1/\varepsilon$):

$$(1 + (\varepsilon \|x\|)^2)^{1/2},$$

for each value of ε and in particular for small values of it, case in which the multiquadric interpolant gives high accuracy generally but severe bad conditioning (see [13]). They calculate the multiquadric interpolant on a circle of the plane $\varepsilon-$complex where the matrix $A(\varepsilon)$ is well conditioned and they evaluate it at equidistant points; then they apply the inverse FFT

and they get the coefficients of the Taylor expansion for a given $\underline{x}$ in case that there are no poles within the circle:

$$s(\underline{x},\varepsilon) = s_0(\underline{x}) + \varepsilon^2 s_1(\underline{x}) + \varepsilon^4 s_2(\underline{x}) + \cdots$$

This method allows to get solutions stable and highly accurate where the function is very flat.

In [6] basis functions with strong decay at infinity and good stability are presented. They are denoted kernel B-splines.

For the space of the polynomials up to order k (whose dimension we denote $q(k)$), the $N \times q(k)$ matrix

$$P_{X,k} := \left(p_j(x_i)\right)_{x_i \in X, 1 \le j \le q(k)}$$

and the space

$$V_{X,k} = \left\{\alpha \in \mathbb{R}^N \,:\, P_{X,k}^T \alpha = 0\right\}$$

are defined.

It is assumed $k > m$, m minimal order of conditional positivity $N \ge q(k) > q(m)$, and $\operatorname{rank} P_{X,k} = q(k)$.

The kernel B-spline is defined as

$$u_{X,k,\alpha}(x) := \sum_{x_j \in X} \alpha_j \Phi(x - x_j)$$

with $\alpha \in V_{X,k}$.

By denoting $B = \operatorname{null}(P_{X,k}^T)$, the matrix of the linear system to be solved is $B^T A_{X,X} B$.

In the case of f fitted by a polynomial of order k without too many oscillations, the scheme with order k provides reconstructions very accurate, besides making the condition $\mathscr{K}_2(B^T A_{X,X} B)$ lower than $\mathscr{K}_2(A_{X,X})$ (lower of several orders for k large enough).

On 100 scattered points and using the multiquadric with $\delta = 0.01$, we get $\mathscr{K}_2(A_{X,X}) = 4.22\, e(6)$ and $\mathscr{K}_2(B^T A_{X,X} B) = 1.20\, e(6)$ for $k = 1$; we get $\mathscr{K}_2(B^T A_{X,X} B) = 3.52\, e(4)$ for $k = 6$, and for $k = 13$ we get $\mathscr{K}_2(B^T A_{X,X} B) = 7.72$. For $\delta = 1$: we get $\mathscr{K}_2(A_{X,X}) = 3.83\, e(18)$, $\mathscr{K}_2(B^T A_{X,X} B) = 8.23\, e(16)$ with $k = 1$; we get $\mathscr{K}_2(B^T A_{X,X} B) = 3.27\, e(12)$ with $k = 6$, and we get $\mathscr{K}_2(B^T A_{X,X} B) = 2.62\, e(4)$ with $k = 13$.

We note that when choosing the polynomial order less or equal than the order m of the conditionally positive definite radial basis, the l^2- conditioning is not modified essentially.

In [6] there is an example with 121 data from the function $f(x,y) = (\sqrt{x^2+y^2} - 0.6)_+^4$ within $\Omega = [0,1]^2$. By taking $k = 10$, e_2, the root mean squared error on a uniform grid, is of the order of $10\, e(-5)$ and the lowering of the condition is of five orders.

Whenever f is not well fitted by a polynomial of order k large, we can get accurate reconstructions of f also by taking order k, provided that there are not regions too empty of data and that there are several data along the boundary.

Example 2. We take $N = 148$ data from humps and dips, see the locations in Figure 1, rather scarce towards the boundary. The results for $k = 4$ and $\delta = 0.35$ are $e_2 = 2\, e(-3)$ and $e_\infty = 3.9\, e(-2)$, with a loss of accuracy for e_2 of the 10% respect to the result with $k = 1$; it is $\mathscr{K}_2(B^T A_{X,X} B) = 4.6\, e(11)$, while $\mathscr{K}_2(A_{X,X}) = 1.6\, e(14)$. The graphical output is shown in Figure 2.

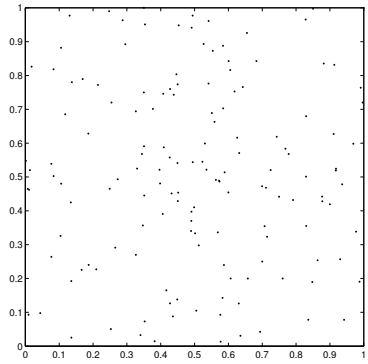

Figure 1: Locations of the data points.

Both modalities said above not always are fully satisfying for practical reasons; for example the kernel B-splines cannot be applied to PDE problems because the resulting RBF-PDE coefficient matrix does not enjoy the property of being strictly positive.

Therefore the researchers to fix the problem of ill conditioning have studied methods of approximation (no more of interpolation) that provide approximating functions with errors less or equal than those of interpolation.

In [12], Lagrangian basis functions are constructed according to the principle of the least squares. Namely , let A be the $N \times N$ collocation matrix. For the generic point x_i, we consider a local neighbourhood of $m < N$ centers, and we denote S_i its index set.

Let B_i be the $m \times N$ submatrix formed by selecting m rows of A from S_i, the least squares problem

$$\| B_i w_i - e_i \|_2, \quad i = 1, \ldots N,$$

is solved, where w_i is the $i-th$ row of the matrix W such that $WA \simeq I$ from which

$$A\alpha = b \implies \alpha = Wb.$$

Another way, described in [8], and suitable for non uniformely scattered data is that one of constructing a multiquadric approximant according to the ℓ_2 norm.

Precisely the purpose is that one of obtaining a stable approximating function close to the unstable interpolant of a smooth function.

The condition of the matrix of interpolation depends on $q(X)/\delta$, as said above, but experimentally one observes that the condition of $A_{X,X}$ depends on the configuration of the data too and that a bad local configuration implies a bad global condition. In [8] we propose an algorithm that here we summarize briefly.

A thresholding value to bound the condition number of the local systems of interpolation is computed.

The local thinning of the points in order not to overcome the threshold is done on the points as sorted by the procedure by Floater and Iske in [9] (such a sorting increases the minimal distance between points).

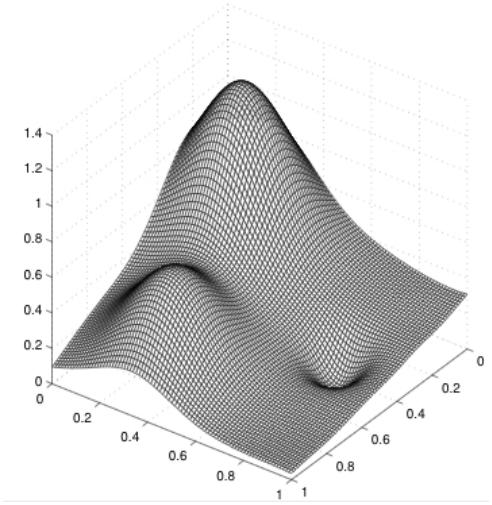

Figure 2: Graphical output.

After this step that has made the local regions more homogeneous for what condition is concerned, a global least squares global approximant is constructed respect to all the data, by MQ basis functions centered at those data points not discarded during the step of controlling the local regions (we denote T this set of knots).

Example 3. We take $N = 149$ data on $\Omega = [0,1]^2$ for the function $2\cos(10\,x)\,\sin(10\,y) + \sin(10\,xy)$. We take the multiquadric with $\delta = 0.35$. The algorithm selects the T subset of size 148. It is $\mathscr{K}_2(A_{X,T}) = 6.6\,e(10)$; $e_2(X,T) = 6.5\,e(-3)$; $e_\infty(X,T) = 1.2\,e(-1)$. See the locations of X dotted and T circled in Figure 3 and the graphical output in Figure 4; notice that there are regions rather empty of data, also at the boundary. Compare with $\mathscr{K}_2(A_{X,X}) = 2.1\,e(19)$, (so large because of a couple of points at distance 10^{-9} from each other), with warning of not full rank from MatLab, and with $e_2(X,X) = 6.5\,e(-3)$ and $e_\infty(X,X) = 1.2\,e(-2)$.

We recall that a way of avoiding bad conditioning is also that one of selecting a subset of significant data from (X, f_X) and of interpolating this subset. In [4] a scheme is proposed to choose a small subset. Then this subset is interpolated by using hexagonal multiquadric kernel B-splines, that are proposed in [6], but with an adaptive construction of different shapes and scales.

References

[1] BEATSON, R., CHERRIE, J. B., AND MOUAT, C. T. Fast fitting of radial basis functions: Methods based on preconditioned GMRES iteration. *Adv. Comput. Math. 11* (1999), 253–270.

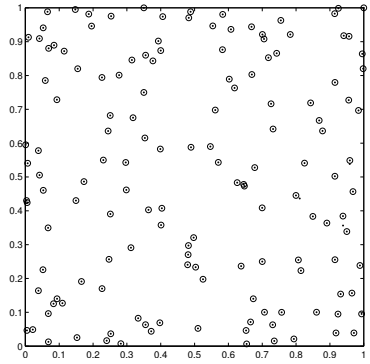

Figure 3: Locations of X and T.

[2] BEATSON, R., AND LEVESLEY, J. Good point/bad point iterations for solving the thin-plate spline interpolation equations. In *Approximation Theory X: Wavelets, Splines and Applications*, Chui, Schumaker and Stöckler, Eds. 2002, pp. 17–25.

[3] BEATSON, R., AND NEWSAM, G. N. Fast evaluation of radial basis functions: I. *Computers Math. Appl. 24* (1992), 7–19.

[4] BOZZINI, M., AND L. LENARDUZZI, L. Bivariate knot selection procedure and interpolation perturbed in scale and shape. In *Curve and Surface Design*, Lyche, Mazure and Schumaker, Eds. Vanderbilt University Press, Nashville, 2003, pp. 1–10.

[5] BOZZINI, M., LENARDUZZI, L., ROSSINI, M., AND SCHABACK, R. Interpolation by basis functions of different scales and shapes. *Calcolo 41* (2004), 77–87.

[6] BOZZINI, M., LENARDUZZI, L., AND SCHABACK, R. Kernel B-Splines and interpolation. *Numer. Alg.*, to appear.

[7] BOZZINI, M., AND LENARDUZZI, L. Reconstruction of surfaces from a not large data set by interpolation. *Rendiconti di Matematica, Serie VII, 25* (2005).

[8] BOZZINI, M., AND LENARDUZZI, L. Stable least squares MQ approximation by local thinning. *submitted*, 2005.

[9] FLOATER, M., AND ISKE, A. Thinning and approximation of large sets of scattered data. In *Advanced Topics in Multivariate Approximation*, F. Fontanella, K. Jetter, and P. J. Laurent, Eds. World Scientific, Singapore, 1996, pp. 87–96.

[10] FORNBERG, B., AND WRIGHT, G. Stable computation of multiquadric interpolants for all values of the shape parameter. *Computers Math. Applic. 48* (2004), 853–867.

[11] LARSSON, E., AND FORNBERG, B. A numerical study of some radial basis function based solution methods for elliptic PDEs. *Computers Math. Applic. 46* (2003), 891–902.

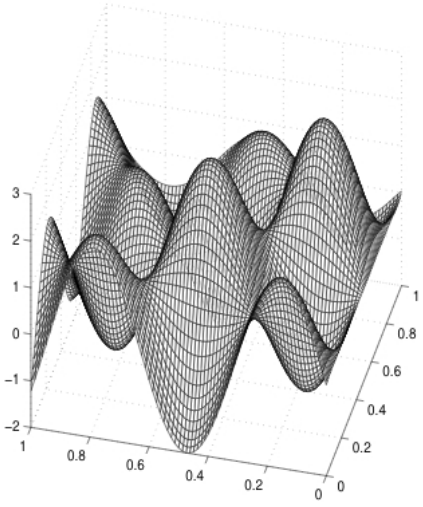

Figure 4: Graphical output.

[12] LING, L., AND KANSA, E. A least squares preconditioner for radial basis functions collocation methods. *Adv. Comput. Math.* 23 (2005), 31–54.

[13] MADYCH, W. R. Miscellaneous error bounds for multiquadric and related interpolators. *Computers Math. Applic. 24* (1992), 121–138.

[14] NARCOVICH, F. J. Recent developments in error estimates for scattered-data interpolation via radial basis functions. *Numer. Alg. 39* (2005), 307–315.

[15] RIPPA, S. An algorithm for selecting a good value for the parameter *c*. *Adv. Comput. Math. 11* (1999), 193–210.

[16] SCHABACK, R. Error estimates and condition numbers for radial basis function interpolation. *Adv. Comput. Math. 3* (1995), 251–264.

[17] SCHABACK, R., AND WENDLAND, H. Kernel techniques: from machine learning to meshless methods. *Acta Numerica.* (2006), 1–97.

Mira Bozzini
Università di Milano Bicocca
Dipartimento di Matematica ed Applicazioni
via Cozzi 53, 20125 Milano (Italy)
mira.bozzini@unimib.it

Licia Lenarduzzi
Istituto di Matematica Applicata e Tecnologie Informatiche del C. N. R. Sez. Milano
via Bassini 15, 20133 Milano (Italy)
licia@mi.imati.cnr.it