Reconstruction of surfaces from a not large
data set by interpolation

MIRA BOZZINI – LICIA LENARDUZZI

ABSTRACT: *Many papers discuss the problem of recovering a function when a set of data is known on a domain $\Omega \subset \mathcal{R}^2$. Most of these papers assume that the size of the data set is large. At our knowledge, little or practically nothing is said about the case in which the sample has a size that is moderate.*
In this paper we tackle this problem. We indicate a way for its solution and hence we give a concrete example. Some numerical experiments are shown.

1 – Introduction

When the data are assigned on a domain $\Omega \subset \mathcal{R}^2$ (in general $\Omega \equiv Q = [0,1] \times [0,1]$) and are accurate enough, an interpolating function $If(\underline{x})$ is constructed.

About this problem there is a broad literature, among which we recall the well known paper [5] by Franke, that is the most hexaustive note about the numerical validity of the different methods proposed in the literature for $2-$variate scattered data.

There it is stressed that the modified Shepard method and the interpolants obtained by linear combination of radial basis functions provide a good solution when the size n is large.

Now it is necessary to consider the adjectives *good* and *large*. The former deals with the quality of the approximation, that in the cited paper is referred

to both in term of error, evaluated as

$$(1) \qquad e_n^p = \left(\sum_{j=1}^{M} \mid If(\underline{\xi}_j) - f(\underline{\xi}_j) \mid^p /M \right)^{1/p} , \ 1 \leq p < \infty,$$

$$(2) \qquad e_n^\infty = \max_j \mid (If(\underline{\xi}_j) - f(\underline{\xi}_j) \mid$$

where $\{\underline{\xi}_j\}_{j=1}^{M}$ are the points of a regular thick grid on Ω, and of the requirement that there are no undue oscillations.

It is obvious that the latter term *large* depends on the current function f. In general, samples with size of the order of hundreds are considered *large* in the above mentioned paper by Franke for smooth functions, but at present, problems in which the size of the data set is huge up to millions are considered too, see [4] for example.

At this point a question arises: what shall we do when the size of the data set is moderate?

The question is important not only from the speculative point of view, but mainly from the applicative one. We recall in fact at least two situations relative to real problems in which one must consider samples of small size.

The former is relative to all those cases in which it is expensive or practically impossible to collect many data. (For example, we recall the problem of putting few leads on the chest of a patient to represent his/her heart potential in such a way that any pathology can be diagnosed correctly, see [1].) The current paper is relevant to this situation.

A latter case concerns the storage of data; a huge amount of data is at disposal, but it is necessary to select a subset of data to store a reduced set of information (for example, with the purpose of describing the evolution of a glacier): this is called a problem of *compression* of the information. Different solutions have been proposed by methods of approximation of the data whose algorithms are experimental [3], [7], [8] and [2].

We notice that in the compression problem we still deal with a large sample and this makes easier to construct a good approximant.

In Section 2 of this note we tackle the problem of constructing a function that interpolates the data and provides a good reconstruction of an unknown smooth function in the case that the sample size is moderate. In Section 3 we indicate a method that determinates the interpolating function according to the modality described at the previous section, as concrete example. For it an automatic algorithm is implemented in Section 4. At last in Section 5 we show some numerical examples to support the validity of our proposal.

2 – Modality to construct an interpolant

The problem of interpolating a function is generally tackled according to the following scheme.

For a function $f \in C(\Omega)$, $\Omega \subset \mathcal{R}^2$, we consider a linear subspace of $C(\Omega)$ and we individuate a suitable basis $\{\phi_i(\underline{x})\}_{i=1}^n$ by which we determine the function $g_n(\underline{x}) = \sum_{i=1}^n a_i(f)\phi_i(\underline{x})$ that interpolates the assigned data $\{\underline{x}_i, f_i\}_1^n$.

Besides we require that as the size n increases, the corresponding sequence of the approximating functions $\{g_n\}$ converges to f.

It is mostly useful to know the rate of convergence with respect to the supremum norm because, in this case, we know in advance the order of magnitude of the error as n is large enough.

In other words, the reliability with which we obtain a satisfying reconstruction of the unknown function f by the interpolating function g_n, is based on the asymptotic properties of g_n, when we are considering a **large** sample.

For the problem tackled in the current paper, it is peculiar that the sample size is moderate and this does not allow us to make use of the asymptotic properties to guarantee a good approximation.

Therefore it is necessary to individuate a different approach to the problem.

We reconsider the problem of interpolation described above.

If we analyze the modalities according to which the problem was solved in the literature carefully, we notice that, in general, the basis functions depend only on the assigned locations $\{\underline{x}_i\}_{i=1}^n$.

The consequence is that, almost surely, infinite informations are needed to reconstruct the unknown function exactly; that is

$$f(\underline{x}) = \sum_{i=1}^\infty a_i \phi_i(\underline{x}).$$

(For instance it is known that when we consider the shift invariant functions on $\mathcal{R}$, a function ϕ reproduces a polynomial of degree m if a sequence $\{a_i\}_{i \in \mathcal{Z}}$ exists such that

$$x^l = \sum_{i \in \mathcal{Z}} a_i \phi(x - x_i) \qquad l = 0, \dots, m.$$

That means that it is necessary to have a vector $a \in l_1(\mathcal{Z})$, whose components do not vanish, to reproduce a constant.)

On the contrary, if ideally we could choose, among the basis functions, a function $\phi_j(\underline{x})$ proportional to f, only one information would be needed to reconstruct f exactly:

$$f = a_j \phi_j(x), \quad a_i = 0, \ \forall i \neq j.$$

These considerations suggest that if we choose the basis functions conforming to f, fewer basis functions and hence fewer functional values are needed to obtain a good approximation to f.

Taking into account this point of view, the construction of a satisfying interpolating function has to be done by an adaptive technique.

Here we propose a technique by constructing the adaptive basis functions by the following scheme:

- we consider n linear independent functions, each of them depending on some parameters. We indicate such functions with

$$\{\phi_i(\underline{x}, \underline{\alpha}_i) \qquad \underline{x} \in \Omega, \qquad \underline{\alpha}_i \in \mathcal{R}^s \ \ s \geq 1, \ i = 1, \ldots n\}.$$

  (Some examples in $1d$ are: the B-splines with variable degree, the B-splines with variable knots, the multiquadric with variable parameter.)
  When we fix a vector set $\{\underline{\hat{\alpha}}_i\}_{i=1}^n$ the corresponding set of functions $\{\phi_i(\underline{x}, \underline{\hat{\alpha}}_i\}_{i=1}^n$ provides a basis $B_\phi(\{\underline{\alpha}_i\}_i)$ spanning a linear space.
- We consider the family $\mathcal{F}_\phi$ of all possible bases as $\underline{\alpha}_i$ varies

$$\mathcal{F}_\phi = \cup_{\underline{\alpha}_i \in \mathcal{R}^s} B_\phi(\{\underline{\alpha}_i\}_i)$$

  subject to the condition that each basis gives a unique interpolant.
- Having prescribed a positive functional operator K, we define: the best basis $B_\phi(\{\underline{\hat{\alpha}}_i\}_i) \in \mathcal{F}_\phi$ for the assigned interpolation problem and according to the operator K the one that minimizes the operator K as $\underline{\alpha}_i$ varies in the subset $D \subset \mathcal{R}^s$ in which the unicity of the interpolating function is guaranteed.

We point out that the operator K, used to construct the adaptive basis functions, provides in addition a quantitative estimate of the approximation obtained.

The last remark stresses the importance of choosing the operator K according to the phenomenon at hand, as the measure of the goodness of fit is provided by this operator. Classical operators presented in the literature are for instance: the measure of entropy, the measure of risk, a measure of energy, the measure according to the l_p norm $1 \leq p < \infty$.

3 – An example

We begin with the following remark.
Within radial functions, the multiquadric functions

$$\phi_0(\underline{x}, \underline{x}_j) = -\sqrt{(x - x_j)^2 + (y - y_j)^2 + c^2}, \quad j = 1, \ldots, n$$

received wide attention. In particular, many papers were devoted to the choice of the parameter c. For our problem, the paper [6] is interesting: there, for dimension one, it is conjectured that the scaling of the multiquadric should be

proportional to the local radius of curvature of the function to be interpolated. On one hand, this conjecture supports the fact that the basis functions must be constructed adaptively; on the other hand it allows us to notice that the adaptive process must include other parameters together with the scaling parameter.

In fact in $2d$, according to the conjecture by Hon and Kansa, when we consider the restriction of f along the tangent t to the contour at a location $\underline{x}_j$ of the data set, we should choose a value c_t of the scaling parameter depending on the local radius of curvature relevant to the restriction considered. On the other hand, if we consider the restriction of f along the gradient g, we should take a value c_g different from c_t in general. So we adapt the argument of the bidimensional multiquadric as follows:

$$a_j(x^g - x_j^g)^2 + b_j(y^t - y_j^t)^2 + c_j^2, \qquad \underline{\alpha}_j = (a_j, b_j, c_j) \in D \subset \mathcal{R}_+^3$$

to balance the distance term and the parameter differently in direction t and in direction g.

Following the scheme described in the previous section, we compute the interpolating function in two steps:

- construction of a locally adaptive basis.
  We consider the family of the linear independent functions

  $$\mathcal{F}_\phi = \cup_{\underline{\alpha}_j \in D} \{\phi(\underline{x}, \underline{x}_j, \underline{\alpha}_j)\}_1^n \qquad \underline{\alpha}_j \in D \subset \mathcal{R}_+^3, \qquad \underline{x}, \underline{x}_j \in \Omega,$$

  with $\{\phi(\underline{x}, \underline{x}_j, \underline{\alpha}_j) = \{a_j(x^g - x_j^g)^2 + b_j(y^t - y_j^t)^2 + c_j^2\}^{1/2}$.
  We select the average of the local errors in the l_2 norm as operator K to individuate the basis functions in agreement with f.
  In order to provide an efficient numerical method that fullfils what said above, we restrict the range of the parameters. We take $a_j = 1$ and $c_j = c$, $j = 1, \dots, n$, with c of the order of magnitude of the average distance among the data locations. The component b_j is made dependent on the main curvatures in this way:

  $$b_j = b_j(\kappa) = (1 + \kappa \mu_j \nu_j)^{-1}, \qquad j = 1, \dots, n$$

  where μ_j is directly proportional to the modulus of the gradient at $\underline{x}_j$ and ν_j is directly proportional to the radius of curvature of the contour at $\underline{x}_j$. Operating with the set $\{\underline{\alpha}_j(\kappa) = (1, b_j(\kappa), c)\}_1^n$, the best collection $B_\phi(\{\underline{\alpha}_j\}_1^n)$ is the one obtained by minimizing with respect to κ the functional

  $$\sum_{j=1}^n (I_{\text{loc}}^j f(\underline{x}_j; \{\underline{\alpha}_h(\kappa)\}_{\underline{x}_h \in U(\underline{x}_j)}) - f(\underline{x}_j))^2, \qquad \kappa \in \mathcal{R}_+$$

where $I_{\mathrm{loc}}^j f$ is the interpolant of the data belonging to a prescribed circular neighborhood $U(\underline{x}_j)$, excluding $\underline{x}_j$. We name $\{\hat{\underline{\alpha}}_j\}_{j=1}^n$ the optimal values of the $\{\underline{\alpha}_j\}_{j=1}^n$.

If the solution of the problem of minimum is $\kappa = 0$, we choose the values $\{c_j\}_1^n$ according to a decreasing function of the local density of the data points.

- Computation of the interpolating function with the basis obtained at the previous step:

$$If(\underline{x}) = \sum_{j=1}^n \hat{a}_j \phi(\underline{x}, \underline{x}_j, \hat{\underline{\alpha}}_j).$$

4 – Algorithm

In this section we sketch the algorithm implementing the proposed method and it fulfils two different requirements.

1) It determines the parameters for the local adaptation of the basis functions to f from a small sample size with scattered locations or at most with locations within regions where the experimenter feels that it is important to probe the current phenomenon (for example in the heart potential data case, the data are sampled more thickly in the front part of the chest).

2) The approximating function enjoys stability properties. That is to say that when the locations $\{\underline{x}_j\}$ are changed a little or the initial approximations of the $\{b_j\}_{j=1}^n$ are rough, the approximating function must maintain the order of accuracy.

STEP A. Determination of the best collection $B_\phi \{\underline{\alpha}_j\}_{j=1}^n$.

- Computation of c.
- Computation of the interpolant by the classical multiquadric basis $\{\phi_0(x, x_i)\}_1^n$

$$I^0 f(\underline{x}) = \sum_{i=1}^n \eta_i \phi_0(\underline{x}, \underline{x}_i).$$

- Computation of the $\{\mu_j\}$: we compute the quantities $\operatorname{grad} I^0 f(\underline{x}_j)$. We put

$$\tau_j = \frac{|\operatorname{grad} I^0 f(\underline{x}_j)|}{\operatorname{median}_h |\operatorname{grad} I f^0(\underline{x}_h)|} \quad \text{and}$$

$$\mu_j = \begin{cases} \tau_j & \text{if } \tau_j > 0.5 \\ 0 & \text{if } \tau_j < 0.5. \end{cases}$$

- Computation of the $\{\nu_j\}$: we consider the contours $\{C(\underline{x}_j)\}_{j=1}^n$ of $I^0 f$ related to the heights $\{I^0 f(\underline{x}_j)\}_{j=1}^n$; each contour $C(\underline{x}_j)$ is described as a set of l_j

sorted points $Q = \{Q_k\}_{k=1}^{l_j}$ which are intersection with a prefixed grid (specified in the numerical examples); we individuate the point $Q_r \in Q$ closest to $\underline{x}_j$, we consider the point Q_r^m that precedes Q_r of $i = [10 * l_j/(\max_j l_j)]$ locations and the point Q_r^p that follows Q_r of i locations along $C(\underline{x}_j)$.

We calculate ν_j as

$$\nu_j = \text{mean}(|\cos(\text{grad}(Q_r), \text{grad}(Q_r^m))|, |\cos(\text{grad}(Q_r), \text{grad}(Q_r^p))|).$$

- Estimate of κ by localized cross validation: for each $\underline{x}_j$, let us indicate $U(\underline{x}_j, R)$ the circle of radius R (in the numerical examples we specify the radius) and center at $\underline{x}_j$; we interpolate the data with locations within $U(\underline{x}_j, R)$ but leaving out $\underline{x}_j$ and by using the basis $\{\phi(\underline{x}; \underline{x}_h, \underline{\alpha}_h(\kappa))\}_{h \in U(\underline{x}_j, R), h \neq j}$ with κ chosen such that

$$\sum_j \left(f_j - \sum_{x_h \in U(\underline{x}_j, R), \ h \neq j} \eta_h \phi(\underline{x}_j; \underline{x}_h, \underline{\alpha}_h(\kappa)) \right)^2$$

is minimized.

For $\kappa = 0$ we put $c_j = \text{median}_{\underline{x}_h \in n(\underline{x}_j)} \text{dist}(\underline{x}_j, \underline{x}_h)$, $n(\underline{x}_j)$ neighborhood of $\underline{x}_j$ according to the Delaunay triangulation, in the case that the data points are uniformely distributed. In the case that there are regions empty of data, it is better to put $c_j = \text{mean}_{\underline{x}_h \in n(\underline{x}_j)} \text{dist}(\underline{x}_j, \underline{x}_h) + 0.5 * (\text{median}_{\underline{x}_h \in n(\underline{x}_j)} \text{dist}(\underline{x}_j, \underline{x}_h) - \text{mean}_{\underline{x}_h \in n(\underline{x}_j)} \text{dist}(\underline{x}_j, \underline{x}_h))$.

STEP B. Interpolation by using the collection $B_{\phi, \{\underline{\hat{\alpha}}\}_{i=1}^n}$ as determined with step A. Let denote If the final interpolant:

$$If(\underline{x}) = \sum_{i=1}^n \hat{a}_i \phi(\underline{x}, \underline{x}_i, \hat{\alpha}_i).$$

The cost of the procedure is of the order of $(2/3)n^3$ operations.

5 – Numerical examples

In this section we present two different functions to show the effectiveness of our proposal.

We have taken a thick regular grid 81×81 where to calculate $I^0 f$ to estimate gradient and contours well; to estimate the gradient at $\underline{x}_j$ we proceed in this way: we estimate grad $I^0 f$ at the grid points and then we compute grad $I^0 f(\underline{x}_j)$ by

interpolating with a local cubic the gradient at the four vertices of the square to which $\underline{x}_j$ belongs.

If the data are scattered we have taken $R = 0.25$ for the localized cross validation: the choice of R is such that $U(\underline{x}_j, R)$ includes about ten data points, which agrees with the choice of a local neighborhood for the local methods, see [5]; in the case that the data are not uniformly located, we have taken R variable to include ten points.

About the determination of κ, the golden section search method can be used but in these examples we simply performed a grid search.

We provide the errors e_n^2 and e_n^∞ defined in (1) and (2) with $M = 3721$, the graphical output and the comparison with the output of the classical multiquadric interpolant on the same data set.

Case A. The former example deals with the sigmoidal function which is relevant to many applicative problems and which is used as test function for shape preserving algorithms in the literature. The function is shown in fig. 1.
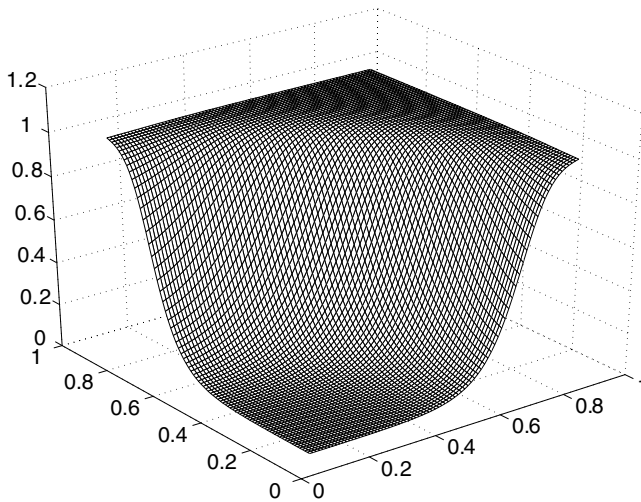
Fig. 1: True function.

Example 1. We consider a scattered data set, 40 points of which are located inside $[0.1, 0.9]^2$ and 13 points are placed at the boundary of the domain (see fig. 2). The initial approximation $I^0 f$, calculated with $c = 0.1$, suitable choice for the classical multiquadric, is shown in fig. 3; we have $e_n^2 = 0.011$ and $e_n^\infty = 0.051$. The graphical output with the locally adapted basis functions is shown in fig. 4; it results $e_n^2 = 0.004$ and $e_n^\infty = 0.023$.

Example 2. We consider the data locations presented in fig. 5, 40 points of which fall within the domain of the graphical output $[0.1, 0.9]^2$, which is framed
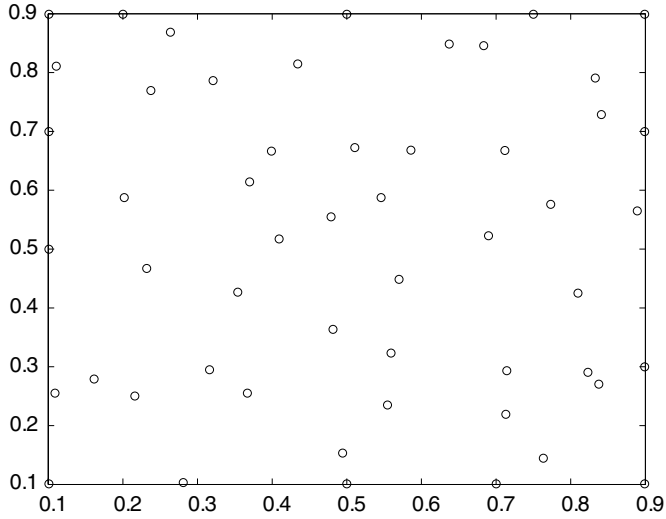
Fig. 2: Data set.
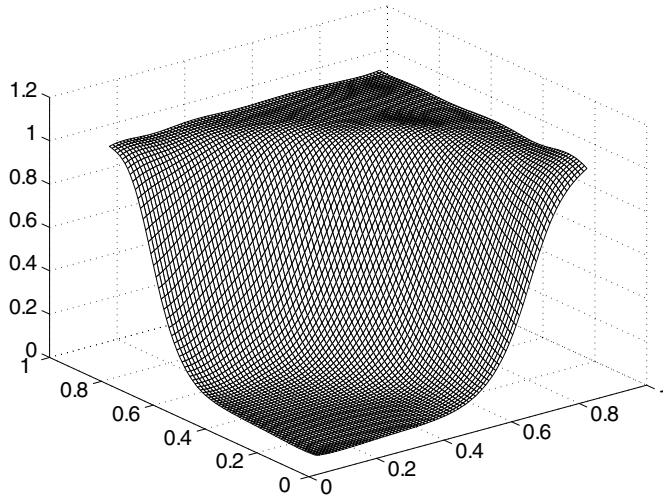
Fig. 3: Reconstruction with the classical ϕ_0 basis.

in the figure, and other data points are placed outside for a total amount of 67 data points.

- In fig. 6 it is shown the graphical output of the $I^0 f$ function ($c = 0.15$) and the errors are $e_n^2 = 0.013$ and $e_n^\infty = 0.067$, while in fig. 7 it is shown the
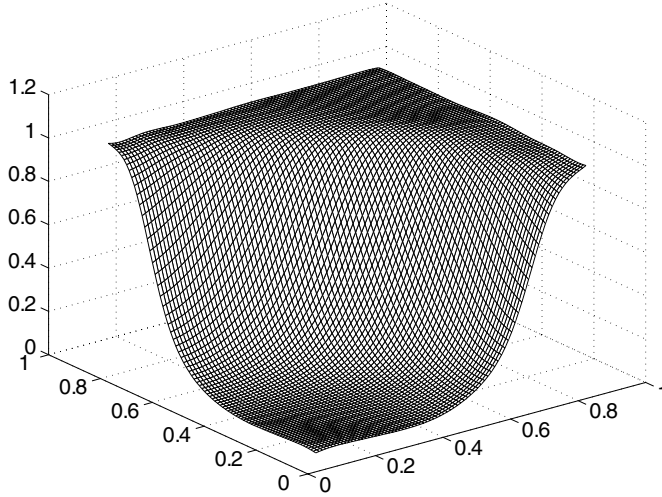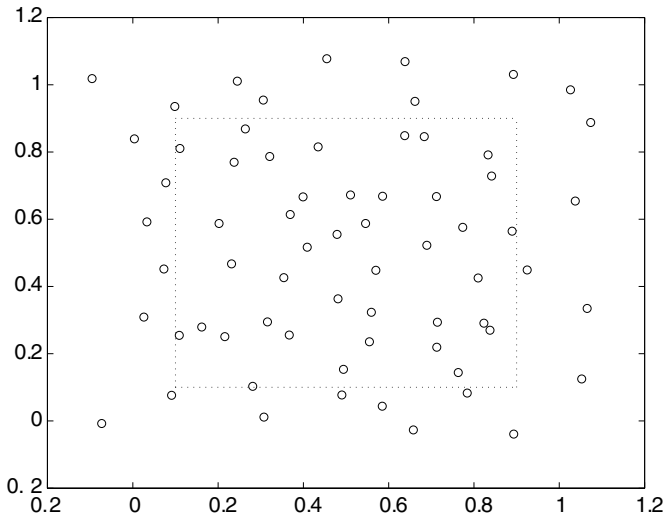
Fig. 4:   Reconstruction with the adapted ϕ basis.

Fig. 5:   Data locations.

graphic obtained with our method and in this case the errors are $e_n^2 = 0.004$
and $e_n^\infty = 0.026$.

• For this case we check the stability with respect to the data locations and
  with respect to the estimate of the derivatives.
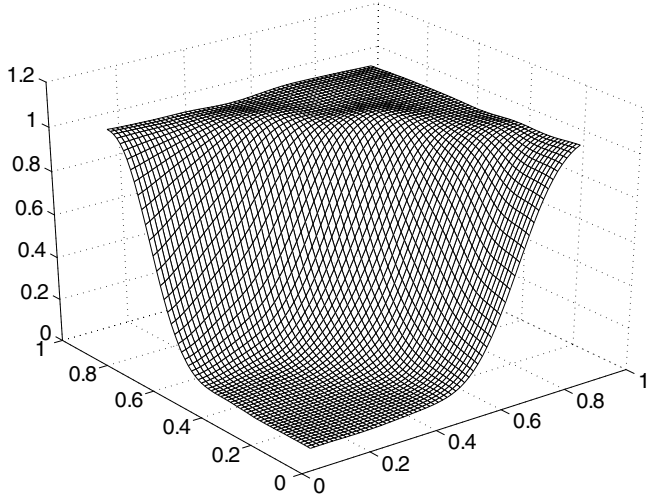
Fig. 6:   Reconstruction with the classical ϕ_0 basis.
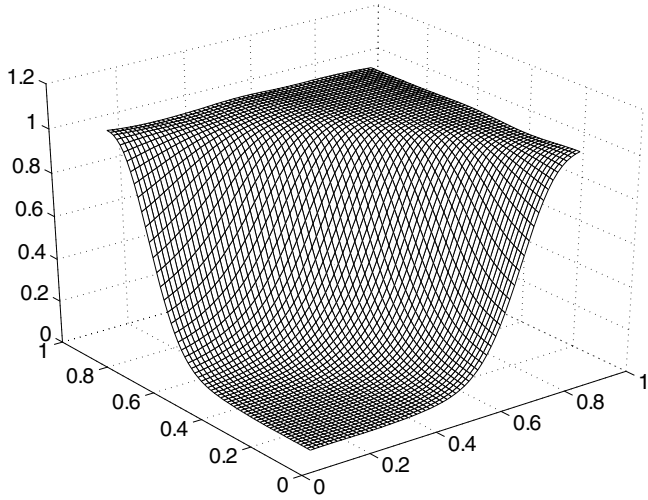
Fig. 7:   Reconstruction with the adapted ϕ basis.

– With respect to different locations of the data, we consider a new set obtained by the same routine for random numbers, but with a different value of the seed. The result is shown in fig. 8.
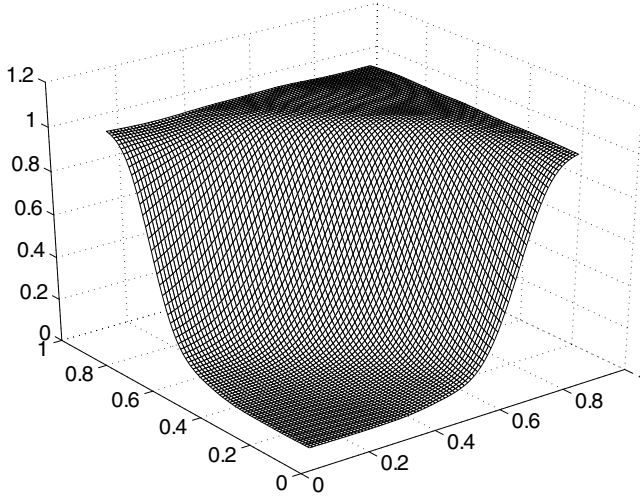
Fig. 8: Stability respect to the data locations. An other data set: reconstruction with the adapted ϕ basis.

– With respect to a different estimate of the derivatives, due to a different less good initial $I^0 f$ approximation (calculated with $c = 0.35$) and shown in fig. 9, the reconstruction with the adapted basis functions (calculated with $c = 0.15$) is shown in fig. 10.
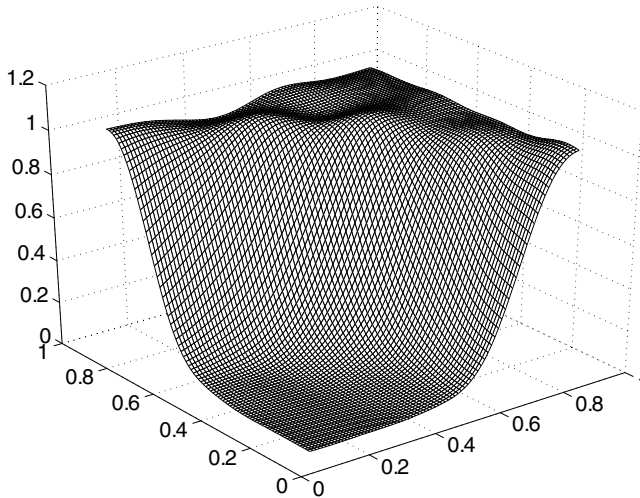
Fig. 9: Reconstruction with the classical ϕ_0 basis.
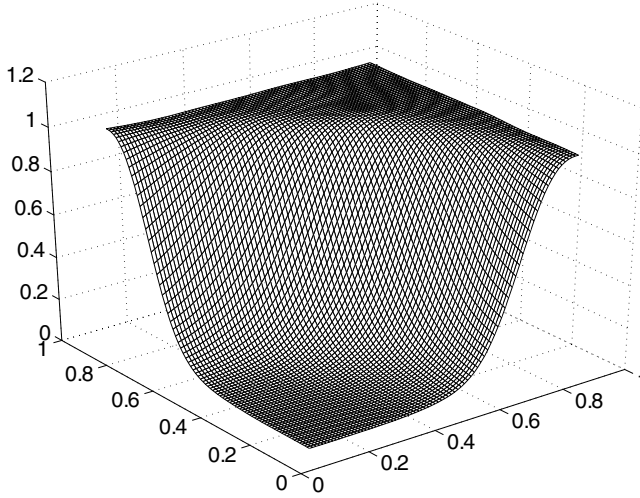
Fig. 10: Stability respect to the derivatives. Reconstruction with the adapted ϕ basis.

CASE B. The latter function is the classical *humps and dips* test function, shown in fig. 11. This function has a radial structure, as also confirmed when running the algorithm: the output is $\kappa = 0$.
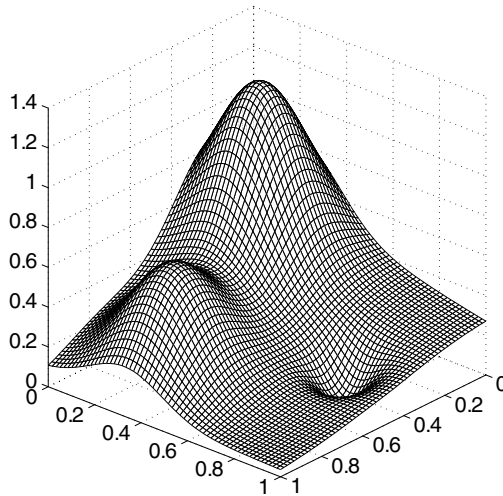
Fig. 11: True function.

EXAMPLE 3. We consider 30 data uniformely scattered in the domain, whose locations are shown in fig. 12. In this case the results are equivalent to
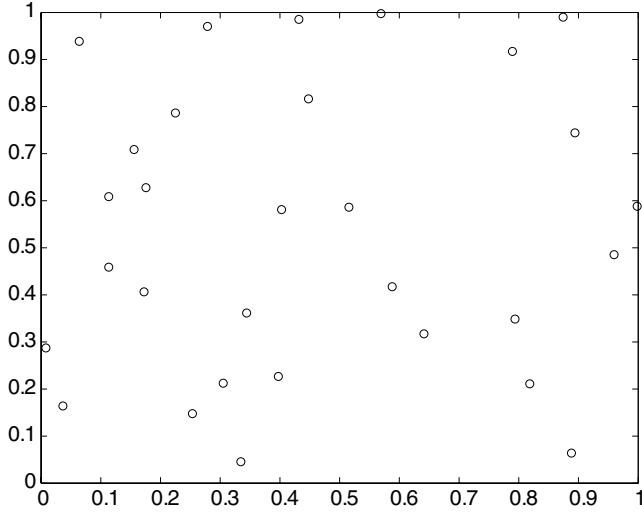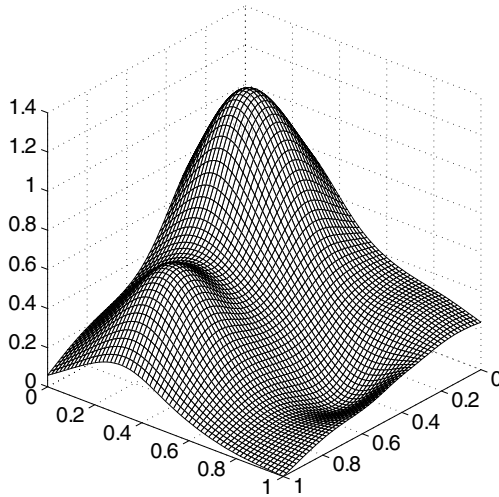
Fig. 12:   Data set.

those obtained by the non adaptive technique, even if there are small differences in the error and in the graphic.

We obtain $e_n^2 = 0.022$ and $e_n^\infty = 0.093$. The graphical output is shown in fig. 13.

Fig. 13:   Reconstruction with variable c_i.

When running with the basis ϕ_0 with constant $c = 0.277$, it is $e_n^2 = 0.024$ and $e_n^\infty = 0.109$. The graphical output is shown in fig. 14.
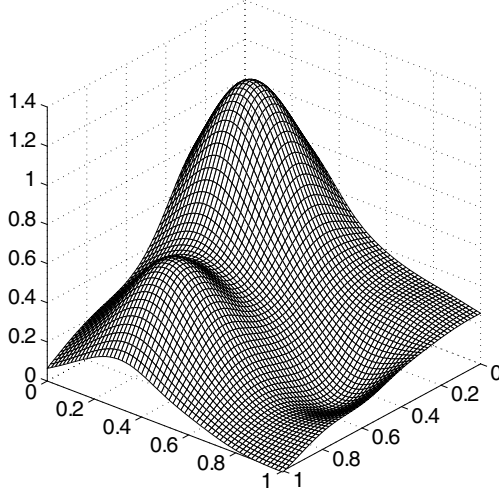

Fig. 14: Reconstruction with the classical basis.

EXAMPLE 4. The locations of the data stress the humps and the dip of the test function. The function is evaluated at 27 data points in the domain $[0,1] \times [0,1]$, 8 points of which staying at the boundary; their locations are shown in fig. 15.
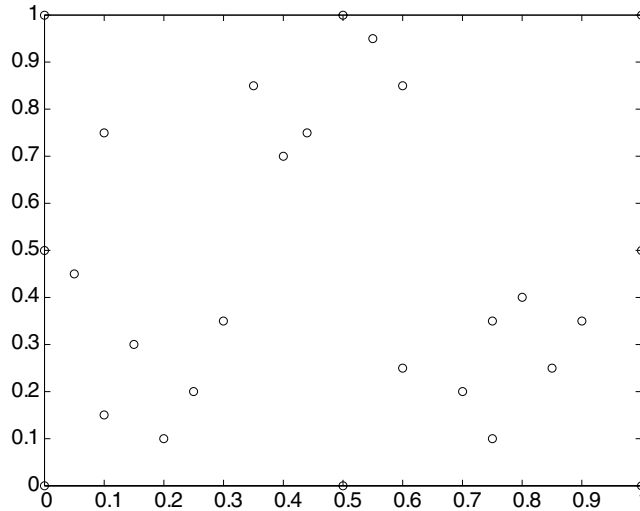

Fig. 15: Data locations.

We obtain $e_n^2 = 0.037$ and $e_n^\infty = 0.175$ and the condition number of the interpolation matrix is $\mathcal{K}_2(A) = 268$, the graphic is presented in fig. 16.
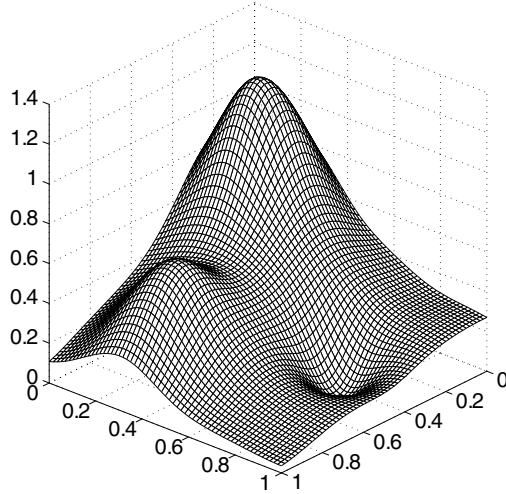

Fig. 16:   Reconstruction with variable c_i.

On the contrary, when running with the basis ϕ_0 with constant $c = \text{mean}_{i=1}^n c_i$ it is $e_n^2 = 0.048$ and $e_n^\infty = 0.157$ with condition number $\mathcal{K}_2(A) = 4.16 \times 10^9$. The graphical output is shown in fig. 17.
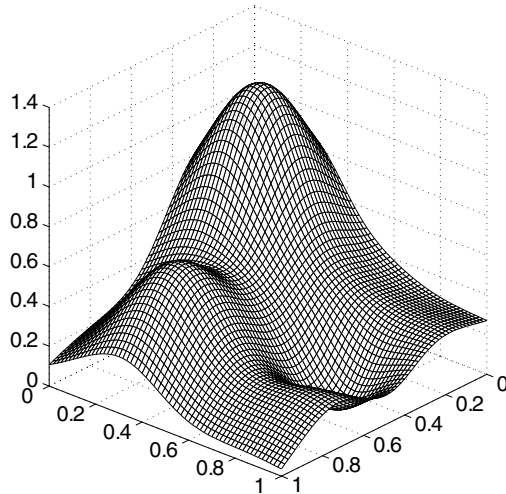

Fig. 17:   Reconstruction with the classical basis.

REFERENCES

[1] M. BOZZINI – F. DE TISI – L. LENARDUZZI: *Reduced set of leads for electro-cardiography maintaining significant information*, J. Comp. Appl. Math., **12**-**13** (1985), 195–206.

[2] M. BOZZINI – L. LENARDUZZI: *Bivariate knot selection procedure and interpolation perturbed in scale and shape*, in "Curve and Surface Design Saint-Malo 2002", T. Lyche – M. L. Mazure – L. Schumaker eds., 2003, Nashboro Press.

[3] M. BUHMANN – A. LE MÉHAUTÉ: *Knot removal with radial function interpolation*, C. R. Sci. Paris Ser. I, **320** (1995), 501–506.

[4] D. DAVYDOV – F. ZEILFELDER: *Scattered data fitting by direct extension of local polynomials with bivariate splines*, Adv. Comp. Math., **21** (2004), 223–271.

[5] R. FRANKE: *Scattered data interpolation: Tests of some methods*, Math. Comp., **38** (1982), 181–200.

[6] Y. HON – E. KANSA: *Circumventing the ill-conditioning problem with multi-quadric radial basis functions: applications to elliptic partial differential equations*, Comput. Math. Applic., **39** (2000), 123–127.

[7] R. MORANDI – A. SESTINI: *Data reduction in surface approximation*, in "Mathematical Methods for Curves and Surfaces: Oslo 2000", T. Lyche – L. L. Schumaker eds., 2001, Varderbilt University Press, 315–324.

[8] R. MORANDI – A. SESTINI: *Geometric knot selection for radial scattered data approximation*, in "Algorithms for Approximation IV", J. Levesley – I. J. Anderson – J. C. Mason (eds.), 2002, Huddersfield, University Press.

INDIRIZZO DEGLI AUTORI:

Mira Bozzini – Università di Milano Bicocca – Dipartimento di Matematica ed Applicazioni – via Cozzi 53 – 20125 Milano, Italy
E-mail: bozzini@matapp.unimib.it

Licia Lenarduzzi (*to be contacted for communication*) – Istituto di Matematica Applicata e Tecnologie Informatiche del C. N. R. – Sez. Milano – via Bassini 15 – 20133 Milano, Italy
E-mail: licia@mi.imati.cnr.it