MOX-Report No. 02/2016

Reconstruction of 3D scattered data via radial basis functions by efficient and robust techniques

Crivellaro, A.; Perotto, S.; Zonca, S.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Reconstruction of 3D scattered data via radial basis functions by efficient and robust techniques

Alberto Crivellaro[♯] Simona Perotto[†] Stefano Zonca[†]

December 24, 2015

[♯] CVLab, École Polytechnique Fédérale de Lausanne
Office BC 306, Station 14, CH-1015 Lausanne, Switzerland
alberto.crivellaro@epfl.ch

[†] MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
simona.perotto@polimi.it
stefano.zonca@polimi.it

Keywords: multivariate interpolation , least-squares approximation , adaptive algorithm , radial basis functions , 3D noisy and lacking scattered data

Abstract

We propose two algorithms to overcome separately two of the most constraining limitations of surface reconstruction methods in use. In particular, we focus on the large amount of data characterizing standard acquisitions by scanner and the noise intrinsically introduced by measurements. The first algorithm represents an adaptive variant of the multi-level interpolating approach proposed in [15], based on an implicit surface representation via radial basis functions. The second procedure is based on a least-squares approximation to filter noisy data. An extensive numerical validation is performed to check the performances of the proposed techniques.

1 Introduction

Surface reconstruction consists in retrieving a virtual model starting from a set of scattered three-dimensional data lying on the surface of an object of interest. This issue characterizes many present-day technological applications, from

Figure 1: The Stanford armadillo model: surface reconstruction (left) and corresponding polygonization (right).

medicine to facial recognition systems, or to the fine arts. The virtual model can be eventually used for visualization or diagnostic purposes, for industrial reverse engineering [4] or to generate a polygonal mesh aimed to scientific computing.

This paper provides two new algorithms which tackle, separately, two of the most constraining limitations of surface reconstruction, i.e., the size of the data sampling typical, for instance, of standard acquisitions by scanner and the noise intrinsically introduced by measurements. The proposed algorithms are both based on an implicit representation of the surface of interest via Radial Basis Functions (RBF) and provide, independently, an improvement of an algorithm previously proposed in [15]. In more detail, we modify the procedure proposed there, relying on a multi-level approach to merge local and global supported RBF. Several advantages are guaranteed by a multi-level reconstruction, we mention, in particular, the sparsity of the matrix, a highly accuracy of recovered surfaces with details at different scales or non-uniformly distributed data, a good quality polygonization of the surface (see Fig. 1 for an example, from the Stanford 3D scanning repository [1]). Nevertheless, the multi-level reconstruction still exhibits a local redundancy in the data and a marked sensitivity to noise. These drawbacks justify the proposal of the new algorithms in this paper.

The first algorithm is suited to efficiently deal with large data set, by avoiding any redundancy in the selection of the data used for surface reconstruction. In short, it may be considered as an adaptive version of the multi-level approach and provides an accurate approximation of a surface Σ resorting to a limited number of points. We complete the multi-level algorithm with an error indicator to refine the cloud of points where Σ exhibits the most complex behaviour.
The second variant is proposed to handle the reconstruction of noisy data. For this purpose, we advocate a least-squares technique which is properly regularized via a ridge regression approach [8].

The natural follow-up of the present work will be the appropriate merging of the two approaches finalized in an adaptive least-squares procedure able of

tackling large sparse and noisy data sets.

The paper is organized as follows. In Section 2, we furnish the multi-level interpolation method presented in [15], after a short introduction on RBF interpolation for scattered data. A detailed numerical check is performed to verify the actual advantages led by this approach. Section 3 details the first new procedure, i.e., the adaptive algorithm. The associated numerical investigation concerns standard data sets as well as data from medical measurements. This last check turns out to be very challenging due to complexity of the considered geometries. Finally, in Section 4, we introduce the least-squares counterpart of the multi-level algorithm and we numerically assess the robustness of such a scheme. Some conclusions are drawn in the last section and a possible follow-up of the work is provided.

2 Interpolation of scattered data via RBF

Let us consider a cloud $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ of N scattered points describing an unknown surface $\Sigma \subset \mathbb{R}^3$. We assume that points in X are endowed with unit inward normals, collected in $\Gamma = \{\mathbf{n}_1, \dots, \mathbf{n}_N\}$, thus defining an orientation on Σ. We aim at reconstructing Σ by exploiting the data in X and Γ. This issue can be generally formulated as the multivariate interpolation problem: given $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \subset \mathbb{R}^3$ and $S = \{f_1, f_2, \dots, f_N\} \subset \mathbb{R}$, with $f_j = f(\mathbf{x}_j)$ for $1 \leq j \leq N$, find a continuous function $F : \mathbb{R}^3 \to \mathbb{R}$ such that

$$F(\mathbf{x}_j) = f_j \quad \text{for } 1 \leq j \leq N. \tag{1}$$

The scalar function F represents the unknown we are interested in, where we assume Σ to coincide with the zero level set $\{\mathbf{x} \in \mathbb{R}^3 : F(\mathbf{x}) = 0\}$ of F, according to an implicit representation of the surface. This simply leads us to select in (1) $f_j = 0$, for $1 \leq j \leq N$.

The multivariate interpolation is a problem less standard with respect to the well-established univariate interpolation [18]. In more detail, according to the Mairhuber-Curtis theorem, it is not *a priori* guaranteed the existence of a multivariate polynomial interpolating an arbitrary set of data in $\mathbb{R}^3$ [24].

In the sequel we resort to a multivariate interpolation based on RBF. This type of interpolation is not covered by the Mairhuber-Curtis theorem. In such a case, problem (1) is well-posed under reasonable assumptions on the selected radial function and on the set X [24].

In the literature, we may distinguish between *globally* or *compactly* supported RBF (for an interesting comparison, we refer to [13]). Globally supported RBF provide good-quality reconstructions even for non uniformly distributed or incomplete data. At the same time, the global support of F associates a full interpolation matrix with conditions (1), thus making the evaluation of F at a generic point $\mathbf{x} \in \mathbb{R}^3$ computationally expensive, especially when the cardinality of X becomes large [4, 20, 10, 7, 21]. On the contrary, compactly supported RBF

are characterized by a sparse interpolation matrix. This significantly reduces the computational effort to evaluate $F(\mathbf{x})$, while ensuring that data perturbations have only a local effect on the reconstructed surface. Nevertheless, compactly supported RBF are not suited to deal with incomplete data. Moreover, the compact support of F, which is in general confined to a thin layer around Σ, may be problematic during the polygonization of the surface. A polygonization step smaller than the thickness of this layer is necessarily required.

In the sequel, we adopt the approach proposed in [15] which merges the computational efficiency of compactly supported RBF with the robustness of a globally supported RBF approximation. The interpolating function F in (1) is chosen as

$$F(\mathbf{x}) = \sum_{j=1}^{N} \big(g_j(\mathbf{x}) + \lambda_j\big)\phi_\sigma(||\mathbf{x} - \mathbf{x}_j||), \tag{2}$$

where $g_j : \mathbb{R}^3 \to \mathbb{R}$ and $\lambda_j \in \mathbb{R}$ are unknown functions and weights, respectively, $\phi : [0, 1) \to [0, +\infty)$ is the selected compactly supported RBF, such that $\phi_\sigma(\cdot) = \phi(\cdot/\sigma)$, with $\sigma \in \mathbb{R}$ a scaling factor estimated from the density of X (see [14] and the sections below for further details), and where $||\cdot||$ denotes the Euclidean norm in $\mathbb{R}^3$. Each term in (2) involves a radial function whose support coincides with a neighborhood of $\mathbf{x}_j$ of radius σ.
In [15] the authors provide a single-level and a multi-level strategy, both based on representation (2). In the next section, we itemize the main steps of the second approach since we will introduce two variants of such a scheme in Sections 3 and 4.

2.1 The multi-level interpolation algorithm

The basic idea of this algorithm is to replace a direct computation of a compactly supported interpolating RBF via an iterative algorithm which gradually adds the data to be interpolated. For this purpose, the procedure starts by generating a coarse-to-fine hierarchy of points from the cloud X, thus identifying the levels $X_1, X_2, \ldots, X_n$ such that $\gamma_1 \leq \gamma_2 \leq \ldots \leq \gamma_n$ with $\gamma_k = \text{card}(X_k)$ and $1 \leq k \leq n$. Each level is also endowed with a corresponding set of unit normals $\Gamma_k = \{\mathbf{n}_1^k, \mathbf{n}_2^k, \ldots, \mathbf{n}_{\gamma_k}^k\}$, and the number n of levels is arbitrarily fixed by the user. In more detail, the bounding box Ω_0 containing all the points in X is recursively subdivided into cells via an octree algorithm. Then, the levels X_k are constituted by the gravity centers $\mathbf{x}_i^k$ of the points of X contained in each cell, whereas the normals $\mathbf{n}_i^k$ are computed as an average of the corresponding normal vectors in Γ [15]. According to this algorithm, it is not necessarily guaranteed that $X_i \subset X_{i+1}$ as well as the points of the different levels do not coincide *a priori* with points in X.

Then, the function F identifying the unknown surface Σ is recursively defined via the rule

$$F_0(\mathbf{x}) = -1; \quad F_k(\mathbf{x}) = F_{k-1}(\mathbf{x}) + o_k(\mathbf{x}) \quad 1 \leq k \leq n, \tag{3}$$

$o_k(\mathbf{x})$ being an offset function taking into account the points of the current level, chosen such that $F_k(\mathbf{x}_i^k) = 0$ for any $\mathbf{x}_i^k \in X_k$. Thus, once the number n of levels has been fixed, function F can be computed simply as the sum of $n+1$ terms, being $F(\mathbf{x}) \equiv F_n(\mathbf{x})$. Notice that, if a rough approximation of the surface Σ is demanded, the recursive procedure (3) can be limited to a reduced number of levels (compare, for instance, the different level of detail in panels (c) and (d) of Figure 2, where n is set to 4 and 6, respectively).
Independently of the level, we assume for the offset the same form as in (2), i.e.,

$$o_k(\mathbf{x}) = \sum_{\mathbf{x}_i^k \in X_k} \big(g_i^k(\mathbf{x}) + \lambda_i^k\big)\phi_{\sigma_k}(||\mathbf{x} - \mathbf{x}_i^k||). \tag{4}$$

Notice that the scaling factor σ_k now depends on the considered level. Concerning the selection of the function ϕ, we employ the Wendland compactly supported RBF $\phi(r) = (1-r)_+^4(4r+1)$, which identifies a positive definite basis, with $(\cdot)_+^n$ the truncated power function with exponent n [23].

The unknown functions g_i^k in (4) are computed via a local quadratic least-squares approximation solved on a σ_k-neighborhood of $\mathbf{x}_i^k$. The weights λ_i^k are determined via a global interpolation, by imposing the interpolation constraints

$$F_k(\mathbf{x}_i^k) = F_{k-1}(\mathbf{x}_i^k) + o_k(\mathbf{x}_i^k) = 0, \quad \forall\ \mathbf{x}_i^k \in X_k \tag{5}$$

(for all the details of the procedure we refer to [15]). Concerning the selection of the scaling factor σ_k, we employ the same recursive formula as in [15], by picking

$$\sigma_1 = cL; \quad \sigma_{k+1} = \frac{\sigma_k}{2} \quad 1 \le k < n, \tag{6}$$

where L is the diagonal of the initial bounding box Ω_0, while c is a user-defined parameter.

Due to the algorithm employed for the level identification, we cannot ensure that the final approximation F_n necessarily interpolates the initial data set X.

2.2 Numerical assessment

We numerically check the performances of the multi-level interpolation approach. We have coded this algorithm in C++, by employing the linear algebra libraries `Lapack` [11] and `OpenNL` [17]. For the visualization of the implicit surface, we use the computer graphics algorithm `marching tetrahedra` that generates a triangulation of the surface by resorting to `marching cubes` [12]. In particular, we use the `Bloomenthal` algorithm which is a very efficient version of `marching tetrahedra` [2, 3]. To solve the linear system associated with the interpolation constraints (5), we use the conjugate gradient method combined with the Jacobi preconditioner [19]. In particular, we employ the normalized residual for the stopping criterion, by demanding an accuracy equal to 10^{-10}. Finally, for the sake of simplicity, we scale the clouds of data so that the diagonal L of the initial

level k	γ_k	$\mathtt{nz}_k$	σ_k	#iterations	$K(A_k)$
1	8	6	0.6027	8	2.727
2	47	7	0.3014	20	4.749
3	231	8	0.1507	62	168.525
4	1261	11	0.0753	100	101.457
5	8676	19	0.0377	88	74.572
6	34833	16	0.0188	56	34.167

Table 1: The Stanford bunny model: quantitative information about the multi-level interpolation approach.

bounding box Ω_0 has unitary length and a vertex coincides with the origin of the Cartesian coordinate system.

As reference data set we select the Stanford bunny model from the Stanford 3D scanning repository [1], consisting of $N = 34833$ points. We build a hierarchy with six levels and we set $c = 0.75$ in (6) as in [15]. In particular, the last level X_6 coincides with the whole data set X.

Table 1 provides some quantitative information related to the generation of the coarse-to-fine hierarchy of points and to the solution of system (5). In particular, for each level X_k, we furnish the corresponding cardinality γ_k, the average number $\mathtt{nz}_k$ of non-zero entries per row of the corresponding matrix A_k identified by the interpolation constraints (5), the radius σ_k in (4), the number of iterations demanded by the solver to reach the desired accuracy and the condition number of A_k. The number of points at each level grows exponentially. Despite this, the computational cost of the algorithm remains contained. This is due to the fact that the local support of the RBF decreases as the level increases, so that the number of non-zeros entries per row remains limited. At each level, the time demanded to solve the corresponding linear system is below the second due to the sparsity of the matrix A_k. Finally, the iterations required by the preconditioned conjugate gradient method to converge as well as the condition number of A_k preserve contained values.

The qualitative results of our investigation are gathered in Figures 2 and 3. With reference to Figure 2, panels (a)-(d) show the levels 1, 2, 4 and 6 of the multi-level interpolation algorithm. The reconstruction of the bunny gradually improves and eventually is very accurate. Panel (e) displays the level set associated with F. As expected, this function can be assimilated to a signed distance from the surface, assuming positive values inside Σ and negative values outside the surface. Panel (f) shows the robustness of the multi-level approach in dealing with incomplete data, being able to effectively repair two holes in the base of the bunny. Panel (h) highlights the capability of the approach in providing a good quality reconstruction even in the presence of data with a non-uniform spatial density. The initial data are shown in panel (g). The density of the points in the back part of the bunny is about one thirtieth of the distribution in the front part.

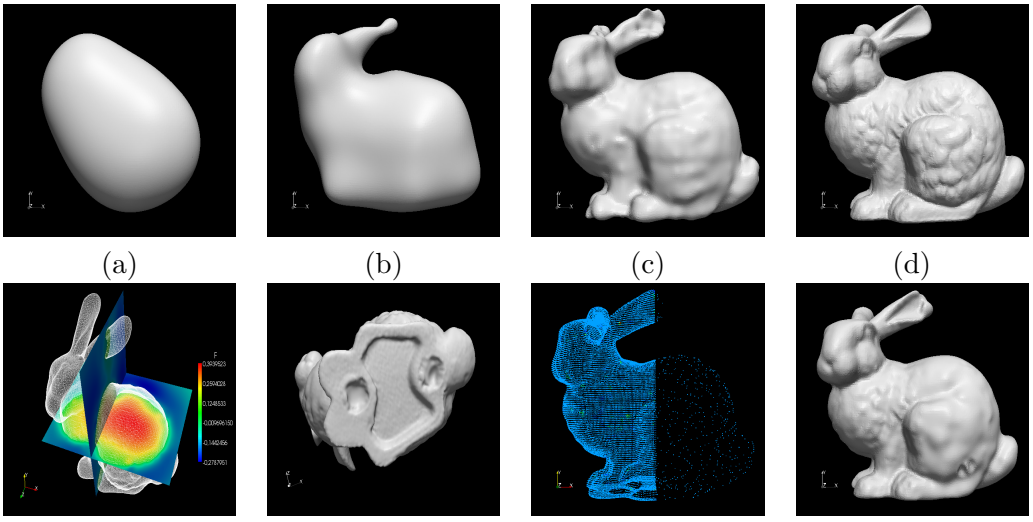
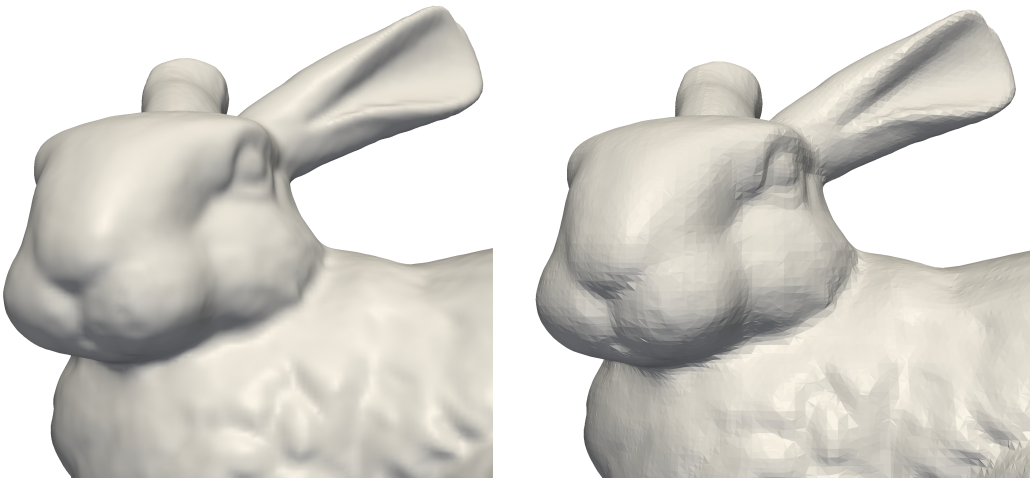


Figure 2: The Stanford bunny model: qualitative information about the multi-level interpolation approach.

Finally, Figure 3 (b) shows the polygonization of the surface detail in panel (a). Despite the compact support of F, the polygonal mesh discretizing the bunny surface is of good quality. Few seconds are, in general, demanded to generate the grid via the `Bloomenthal` algorithm. In the specific case, for instance, the time associated with a polygonization step of 0.01 amounts to 15.05 seconds.

To summarize, from this numerical check, we can state that the multi-level reconstruction preserves the main advantages of a standard compactly supported RBF approximation (in particular, the sparsity of the interpolation matrix; the accuracy in the reconstruction of surfaces with details at different scales), while overcoming some limits of such an approximation. In particular, it successfully interpolates data with a non-uniform spatial distribution, fills holes and cuts in the surface, provides an implicit function F with a (nearly) global support and allows to obtain a good quality polygonization of the surface. Nevertheless, the multi-level approach still suffers of some drawbacks. For instance, it would be advisable to limit the considerable growth of the number of points at each level. This often yields a final level containing either all or too few points with respect to the original set X. Moreover, during the creation of a new level, no check is performed on the accuracy already guaranteed by the previous levels. This may lead to a local redundancy in the data identifying the reconstructed surface. Finally, the multi-level reconstruction still exhibits a marked sensitivity to noise (see, for instance, Figure 11, left). The two next sections are meant to address these limitations by providing two new variants of the multi-level interpolation algorithm.



(a) (b)

Figure 3: The Stanford bunny model: polygonization (b) of the surface detail in (a).

3 An adaptive multi-level approach

The adaptive approach will lead to reconstructions which involve less points with respect to the standard multi-level procedure. Consequently, the computational effort for reconstructing Σ will be sensibly reduced.

3.1 Adaptive creation of the levels

Goal of the adaptive approach is to obtain, at the lowest levels, a rough reconstruction of the features of Σ in correspondence with the coarsest scales, and then, at the successive levels, to adaptively enrich such a reconstruction in the regions where details at small scales are present. A check on the local accuracy ensured by the reconstructed surface will drive the adaptive enrichment.

As in the original multi-level algorithm, we recursively subdivide the initial bounding box Ω_0 via an octree algorithm to produce a coarse-to-fine hierarchy of levels, $X_1, X_2, \ldots, X_n$. Nevertheless, we adopt a new criterion to select the points of the levels. In particular, the levels X_k are now constituted by the points $\mathbf{x}_i^k$ of X closest to the center of gravity of the points of X contained in each cell, while $\mathbf{n}_i^k$ exactly coincides with the normal in Γ associated with $\mathbf{x}_i^k$. Then, starting from the $\tilde{k}$th level, with $\tilde{k}$ a user-defined integer, we introduce a filter to discard some points at each level, with the aim of refining the reconstruction only where details at small scales are relevant. For this purpose, we exploit the fact that, as shown in the previous section, F behaves essentially like a signed distance from the surface, so that the farthest the point $\mathbf{x}$ the largest the value $|F(\mathbf{x})|$. As a consequence, the quantity $|F(\mathbf{x})|$ can be assumed as an estimator for the interpolation error.

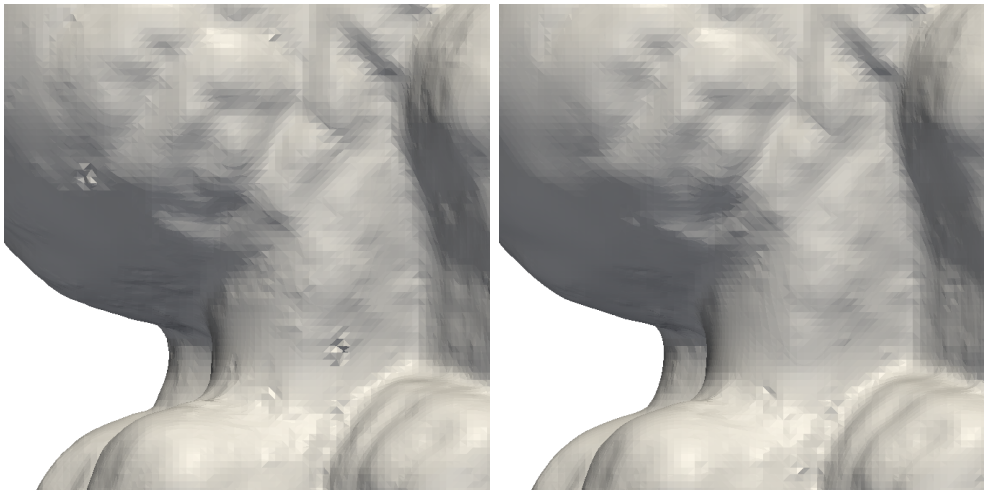


Figure 4: The Stanford bunny model: a detail of the reconstructed surface via the criterion based on a fixed rate (left) and on a fixed number (right) of the points at each level.

Thus, to create the kth level, with $k \geq \tilde{k}$, we first identify all the points $\mathbf{x}_i^k$ of the level X_k (together with the associated unit normals in Γ_k), and then we keep only the points leading to the highest error with respect to the previous surface reconstruction, i.e., such that $|F_{k-1}(\mathbf{x}_i^k)|$ is the largest possible value.

An adaptive choice of the points demands a sufficiently sharp reconstruction of the surface up to level $\tilde{k}-1$, in order to have a precise idea of the areas where a refinement of the sampling is required. This justifies the introduction of an adaptive selection only at the highest levels. The choice of $\tilde{k}$ essentially depends on the total number of points in X. In general, the smaller $\tilde{k}$, the higher the number of iterations required to achieve a certain accuracy. In our numerical experiments, we usually set $\tilde{k} = 5$.

We assessed two different criteria to select the worst points. In the first case, a fixed rate of the total number of points in X_k is picked. In the second case, a fixed number M of points is selected at each new level k, with $k \geq \tilde{k}$. The first criterion suffers of the same limits of the original multi-level approach, by exhibiting an exponential growth of the number of points, level by level. The second strategy avoids this exponential growth but, in general, creates a high number of levels with M points. In such a case, it may be advisable to keep the value of the radius σ sufficiently large, for instance, by setting $\sigma_k = \sigma_{\tilde{k}}$ for each $k \geq \tilde{k}$.

The numerical assessment suggests that the second criterion is more effective guaranteeing a better quality of reconstruction with respect to the number of points involved. Figure 4 shows a detail of the Stanford bunny model, reconstructed via the two adaptive strategies. The criterion based on a fixed number of points yields a slightly smoother surface.

The whole adaptive procedure is itemized in Algorithm 1.

Algorithm 1 : ADAPTIVE MULTI-LEVEL ALGORITHM

input : $X, \Gamma, n, \tilde{k}, M$
build the bounding box Ω_0 containing the points in X;
for $k = 1, \dots, n$ **do**
  divide Ω_{k-1} into 8^k cells, $\Omega_k^1, \dots, \Omega_k^{8^k}$, via an octree algorithm;
  for each cell Ω_k^i, with $i = 1, \dots, 8^k$, find the point $\mathbf{x}_i^k \in X$ closest to the center of gravity of the points $\mathbf{x} \in X \cap \Omega_k^i$;
  if $k < \tilde{k}$ **then**
    add $\mathbf{x}_i^k$ to X_k, for $i = 1, \dots, 8^k$;
    add the associated vectors $\mathbf{n}_i^k \in \Gamma$ to Γ_k, for $i = 1, \dots, 8^k$;
  else
    add the M points $\mathbf{x}_i^k$ with the greatest error $|F_{k-1}(\mathbf{x}_i^k)|$ to X_k;
    add the associated vectors $\mathbf{n}_i^k \in \Gamma$ to Γ_k;
  end if
end for
output : $\{X_k\}_{k=1}^n$, $\{\Gamma_k\}_{k=1}^n$

3.2 Numerical assessment

The numerical check of this section aims at consolidating the good properties of the adaptive multi-level algorithm. The adaptive algorithm has been implemented in the same environment of the standard procedure, and we preserve the solver as well.
As expected, it can be checked that the adaptive procedure yields an accurate, albeit parsimonious, approximation of the surface Σ that consequently requires a fast post-processing. Indeed, the adaptive algorithm avoids any redundancy since, when creating the kth level, it automatically discards the points $\mathbf{x}_i^k$ such that $F_{k-1}(\mathbf{x}_i^k) = 0$.

3.2.1 Standard vs adaptive multi-level algorithm

We focus on two benchmark problems. The first one represents a brain surface, the second one coincides with the Stanford bunny tackled in Section 2.2 [1].

The brain dataset consists of $N = 46650$ points (see Figure 5). In Table 2 we quantitatively compare the reconstruction provided by the multi-level algorithm in [15] and by the adaptive variant here proposed. The number n of levels is fixed to 6 in the first case and to 7 in the adaptive case. The adaptive selection of the points is activated at the fifth level ($\tilde{k} = 5$), and the number M of points preserved at each level $k \geq \tilde{k}$ is set to 8000. Finally, also in the adaptive case, we halve the radius σ according to (6).

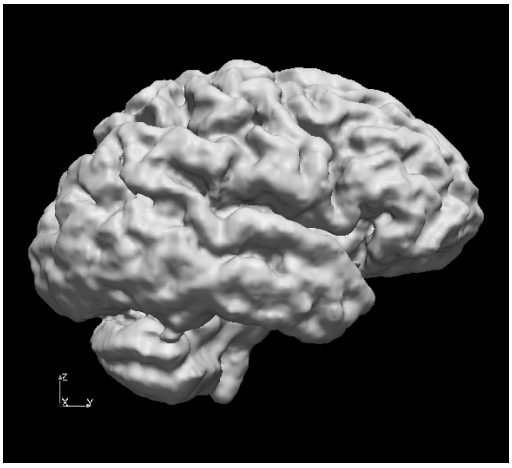


Figure 5: The brain data set.

A comparison between the two main panels of the table confirms the exponential growth of the number of points in the non-adaptive case, in contrast to the stagnating cardinality of the adaptive approach. The different criterion adopted to select the points $\mathbf{x}_i^k$ leads to a different accuracy already before starting the adaptive selection, with an error which is about the half in the adaptive case. At the fifth level, the two methods reach a similar accuracy even though the cardinality of X_5 in the adaptive case is about one fourth of the number of points employed by the standard approach. As a consequence, the computational cost is considerably damped in the adaptive case, dealing with a linear system of reduced order. In a dual way, the adaptive approach allows to considerably reduce the interpolation error after sampling about the same number of points. This is evident by comparing the highlighted values in the table. Finally, we remark that the small error provided at the sixth level of the standard algorithm is due to the employment essentially of all the available data.

	Non-adaptive			Adaptive		
level k	γ_k	σ	error	γ_k	σ	error
1	8	0.5563	4626.62	8	0.5563	1981.41
2	60	0.2781	2961.20	60	0.2781	1393.13
3	412	0.1390	1700.13	412	0.1390	937.038
4	2894	0.0695	804.102	2894	0.0695	444.684
5	25955	0.0348	180.089	8000	0.0348	186.200
6	46637	0.0173	0.06046	8000	0.0173	86.3005
7	-	-	-	8000	0.0086	39.7091

Table 2: The brain dataset: quantitative comparison between the standard multi-scale and the adaptive reconstruction.

level k	γ_k	$\mathtt{nz}_k$	σ_k	#iterations	$K(A_k)$
1	8	6	0.6027	8	2.727
2	47	7	0.3014	20	4.749
3	231	8	0.1507	62	168.525
4	1261	11	0.0753	100	101.457
5	4000	12	0.0377	73	58.351
6	4000	6	0.0188	41	21.811
7	4000	1	0.0094	18	3.991

Table 3: The Stanford bunny model: quantitative information about the adaptive multi-level interpolation approach.

Now we apply the adaptive multi-level procedure to the Stanford bunny model in Figure 2. In more detail, we replicate the quantitative analysis provided in Table 1 for the standard approach, by selecting $n = 7$, $\tilde{k} = 5$, $M = 4000$ and by gradually halving the radius σ as in (6). Table 3 summarizes the main results of such an analysis. Notice that the basic algorithm employs one level less since it already employs all the points in X at the sixth level. Despite the increased number of levels and the additional check on the error demanded by the adaptive procedure, the computational time required by the adaptive multi-level algorithm is lower, namely 66 seconds to be compared with 90 seconds of the standard multi-level method. Finally, in the adaptive case, we detect a higher sparsity of the matrix A_k as well as a reduction of the condition number $K(A_k)$.

Figure 6, top shows a qualitative comparison between the reconstructions provided by the two approaches. The two surfaces are fully comparable despite the lower sampling performed by the adaptive procedure (less than 14000 points to be compared with the full set X of data, consisting of 34833 points). The enlarged views in the bottom part of the figure highlight that the details lost by the adaptive multi-level approach as well as the little roughness are essentially irrelevant.

3.2.2 Reconstructions from medical data

We assess the performances of the adaptive multi-level approach on scattered data provided by medical measurements such as computed tomography (CT) scan or magnetic resonance imaging (MRI). The complexity of the surfaces to be reconstructed makes this a challenging task. Indeed, the involved surfaces often exhibit details at very different scales. In addition, the data are usually acquired on parallel sections (slices), thus exhibiting local variations in the spatial density. We compare the reconstructions provided by the proposed adaptive procedure with the output of the Vascular Modelling ToolKit (VMTK), a state-of-the-art collection of libraries and tools for 3D reconstruction, geometric analysis,

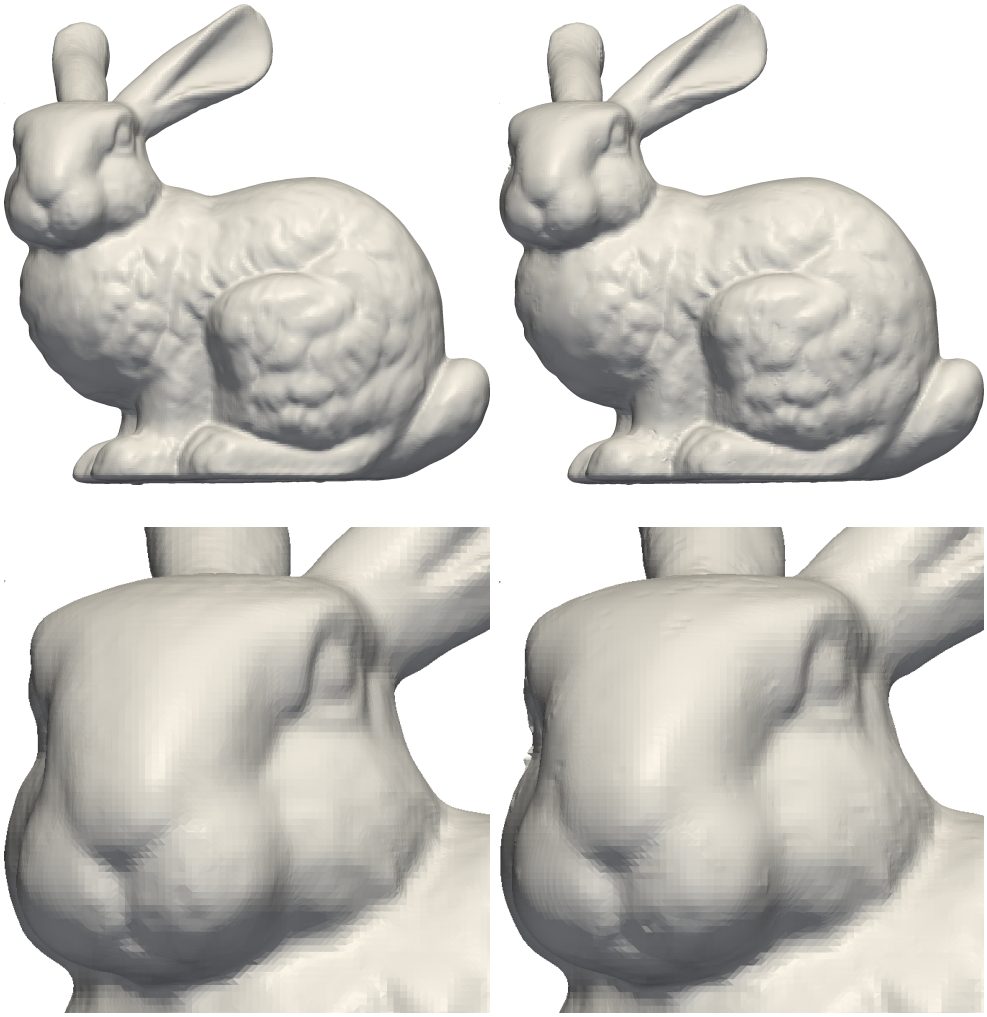


Figure 6: The Stanford bunny model: qualitative comparison between the standard (left) and the adaptive (right) multi-level reconstruction; the whole bunny (top) and a detail (bottom).

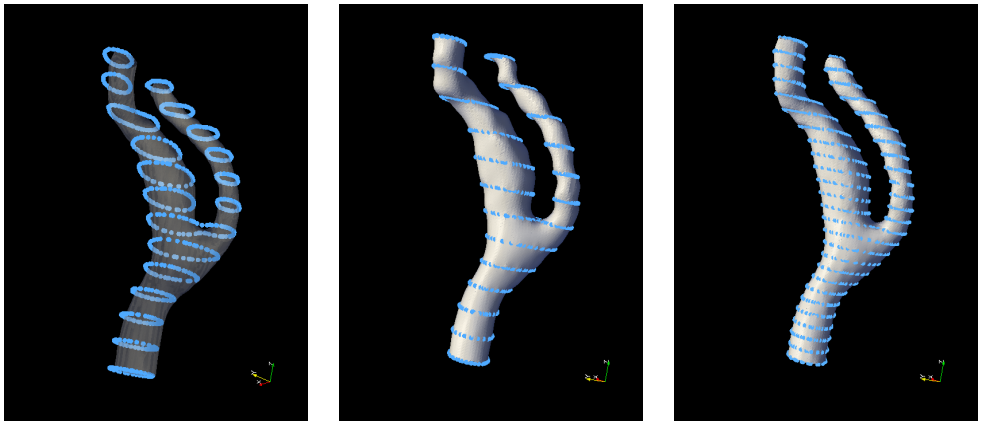
(a) (b) (c)

Figure 7: Reconstruction of an aortic arch provided by the adaptive multi-level method (a) and by VMTK, via bi-harmonic radial basis functions (b) and thin plate splines (c).

mesh generation and surface data analysis for image-based modeling of blood vessels [22].

Figure 7 gathers the results associated with the reconstruction of an aortic arch. We employ the VMTK interpolation routines based on globally supported biharmonic radial basis functions and on thin plate splines, respectively both contained in the VMTK script `vmtkrbfinterpolation`. Panels (b) and (c) show that, in the reconstruction performed by VMTK, the presence of details at different scales introduces spurious surfaces that are difficult to remove even with a manual post-processing. In particular, the spline reconstruction yields a second wide surface, external to the aortic arch. On the contrary, no artifacts are present in the reconstruction provided by the adaptive multi-level method, as displayed in panel (a).

In Figure 8, we assess the robustness of the adaptive algorithm when dealing with data acquired on parallel sections. We consider the reconstruction of a carotid bifurcation whose total length is 42mm (see panel (a)). The distance among the slices plays an important role. In panel (b), we perform a first reconstruction starting from data on equispaced slices at a reciprocal distance of 3mm. In such a case, the total number of points at our disposal is $N = 912$. The shape of the bifurcation is correctly retrieved even though the reconstructed surface exhibits some bumps, which are particularly evident along the two branches. Panel (c) shows the output of the adaptive procedure when the distance among the slices is diminished to 1.5mm and the number N essentially doubles ($N = 1814$). Any roughness of the surface is now erased. This confirms the capability of the adopted approach to provide good quality reconstructions even when data are not uniformly distributed. The total time T demanded by the multi-level procedure is very contained as shown in Table 4, which distinguishes between the time T_r for the surface reconstruction and the time T_p required by the polygonization



(a) (b) (c)

Figure 8: Adaptive reconstruction of a carotid bifurcation starting from data on parallel sections (a). Reconstruction with data acquired on slices at a distance of 3mm (b) and of 1.5mm (c).

of the surface. Polygonization takes much more time than reconstruction. The adaptive algorithm reduces the times, in particular the reconstruction time. All the times associated with the finer sampling are lower, likely in contrast to the expectation. This is justified by the choice made for the initial radius σ_1, set to 1.5 in the coarse case and to 1 for the fine sampling.

	$N = 912$			$N = 1814$		
	T_r	T_p	T	T_r	T_p	T
standard	1.015	8.145	9.168	0.788	5.480	6.278
adaptive	0.624	7.793	8.424	0.282	4.601	4.891

Table 4: Reconstruction of a carotid bifurcation: times (in seconds) demanded by the two multi-level procedures and for the two samplings of the surface.

4 A least-squares multi-level approach

This section provides a variant of the multi-level approach suited to deal with noisy data. As expected, the strong sensitivity to noise represents a limit also of the adaptive procedure (see Figure 9, left for an example). In more detail, we preserve the criterion adopted in the basic approach to generate levels X_k and to compute functions g_i^k in (4), while replacing the interpolation conditions (5) for the weights λ_i^k with a least-squares approach, which exploits a set of points larger than X_k.

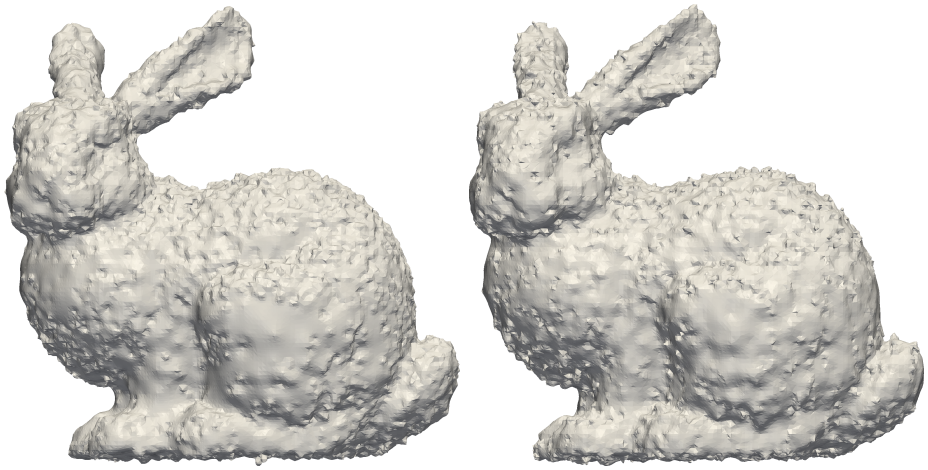


Figure 9: The Stanford bunny model, reconstruction from noisy data: adaptive multi-level approach (left); standard regression method (right).

To this aim, we introduce the global error associated with level k

$$E^k_{global}(\boldsymbol{\lambda}_k) = \frac{1}{L}\sqrt{\sum_{i=1}^{Q} F_k(\mathbf{y}_i)^2}, \tag{7}$$

where the points $\mathbf{y}_i$ belong to a set Y of cardinality $Q > \gamma_k$, with $\gamma_k = \text{card}(X_k)$, and L is the length of the diagonal of the initial bounding box. The quantity $E^k_{global}(\boldsymbol{\lambda}_k)$ coincides with an l^2-norm of the least-squares error. We have explicitly highlighted the dependence of the global error on the unknown coefficients λ^k_i gathered in the vector $\boldsymbol{\lambda}_k$. Different choices are possible for the set Y which, *a priori*, may coincide with any set of cardinality greater than γ_k, even with the whole set X (this is the default choice in the numerical validation of the next section). Finally, the scaling factor $1/L$ in (7) makes $E^k_{global}(\boldsymbol{\lambda}_k)$ independent of the data scale. The same scaling has been adopted in the interpolatory approach.

Thus, to identify the coefficients λ^k_i, we minimize the global error by computing

$$\underset{\boldsymbol{\mu}_k \in \mathbb{R}^{\gamma_k}}{\text{argmin}} \left[E^k_{global}(\boldsymbol{\mu}_k)\right]^2$$

where $\boldsymbol{\mu}_k$ denotes a generic vector of $\mathbb{R}^{\gamma_k}$. This is equivalent to impose

$$\frac{\partial \left[E^k_{global}(\boldsymbol{\mu}_k)\right]^2}{\partial \boldsymbol{\mu}_k}\Big|_{\boldsymbol{\mu}_k = \boldsymbol{\lambda}_k} = \mathbf{0}$$

i.e., to solve the normal equation system

$$\mathbf{A}^T_k \mathbf{A}_k \boldsymbol{\lambda}_k = \mathbf{A}^T_k \mathbf{b}_k, \tag{8}$$

where $\mathbf{A}_k \in \mathbb{R}^{Q\times\gamma_k}$ is the rectangular matrix of entries $[\mathbf{A}_k]_{ij} = \phi_{\sigma_k}(\|\mathbf{y}_i - \mathbf{x}_j^k\|)$, while the generic entry of the vector $\mathbf{b}_k \in \mathbb{R}^Q$ is provided by

$$b_i = -\left(F_{k-1}(\mathbf{y}_i) + \sum_{j=1}^{Q} A_{ij}G_{ij}\right)$$

with $G_{ij} = g_j^k(\mathbf{y}_i)$. For the radius σ_k we preserve the choice in (6).

In contrast to the expectations, this new approach does not lead to significant improvements when dealing with noisy data (see Figure 9, right for an example), especially when the number n of levels involved in the multi-level approach is large.

4.1 A ridge regression penalization

Following [16], to improve the performances of the least-squares reconstruction we modify the error functional (7) via a ridge regression, which essentially penalizes the size of the regression coefficients [8]. We consequently replace the global error associated with level k, $\left[E_{global}^k(\boldsymbol{\lambda}_k)\right]^2$, with the new quantity

$$\left[E_{reg}^k(\boldsymbol{\lambda}_k)\right]^2 = \left[E_{global}^k(\boldsymbol{\lambda}_k)\right]^2 + T_{reg}\|\boldsymbol{\lambda}_k\|_{\sigma_k}^2, \tag{9}$$

where $\|\boldsymbol{\lambda}_k\|_{\sigma_k}$ denotes the weighted norm

$$\|\boldsymbol{\lambda}_k\|_{\sigma_k} = \frac{1}{\sigma_k}\sqrt{\sum_{j=1}^{\gamma_k}\left(\lambda_j^k\right)^2}$$

and T_{reg} is a positive parameter to be properly set. Clearly, with $T_{reg} = 0$ we recover the original least-squares approach. We apply the ridge regularization at each level k of the multi-level approach in contrast to [16] where the regularization is adopted in a single-level strategy. Now, we are led to compute

$$\underset{\boldsymbol{\mu}_k\in\mathbb{R}^{\gamma_k}}{\operatorname{argmin}} \left[E_{reg}^k(\boldsymbol{\mu}_k)\right]^2$$

i.e., (8) changes into the new system

$$\left(\mathbf{A}_k^T\mathbf{A}_k + T_{reg}\mathbf{D}_k\right)\boldsymbol{\lambda}_k = \mathbf{A}_k^T\mathbf{b}_k, \tag{10}$$

with $\mathbf{D}_k \in \mathbb{R}^{\gamma_k\times\gamma_k}$ the diagonal matrix of entries $D_{ii} = (L/\sigma_k)^2$ for $i = 1,\dots,\gamma_k$, whose solution is $\boldsymbol{\lambda}_k = (\mathbf{A}_k^T\mathbf{A}_k + T_{reg}\mathbf{D}_k)^{-1}\mathbf{A}_k^T\mathbf{b}_k$.

The quadratic term $T_{reg}\|\boldsymbol{\lambda}_k\|_{\sigma_k}^2$ in (9) introduces a regularization. In particular, the invertibility of the matrix $\mathbf{A}_k^T\mathbf{A}_k + T_{reg}\mathbf{D}_k$ is guaranteed, thus ensuring the uniqueness of the solution to problem (10). This is not always the case for

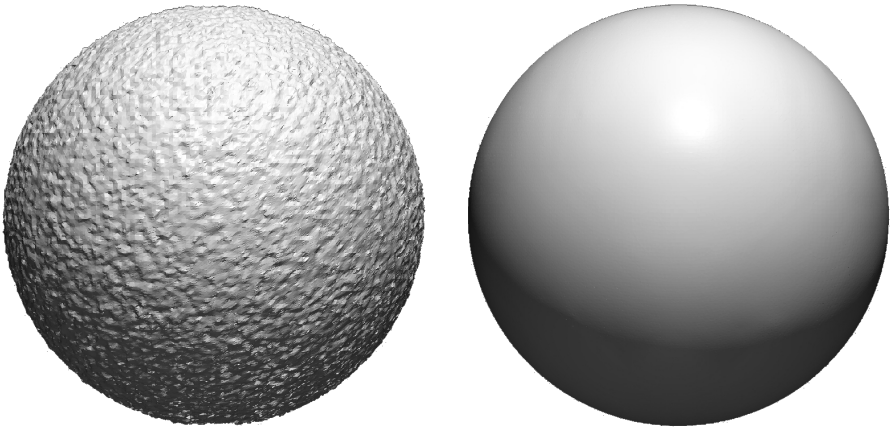


Figure 10: Reconstruction of a sphere from noisy data: standard multi-level approach (left); ridge regression method (right).

the standard least-squares system (8) which admits a unique solution only if matrix A_k is full rank.

The penalization acts by shrinking the value of the coefficients λ_j^k, while the parameter T_{reg} tunes such a contraction. In more detail, when T_{reg} increases, the norm of the vector $\boldsymbol{\lambda}_k$ decreases, i.e., $\|\boldsymbol{\lambda}_k\|_{\sigma_k} \to 0$ as $T_{reg} \to \infty$. Viceversa, when the value of T_{reg} reduces to zero, we have that $\|\boldsymbol{\lambda}_k\|_{\sigma_k} \to \|\boldsymbol{\lambda}_k^{OLS}\|_{\sigma_k}$, where we have denoted by $\boldsymbol{\lambda}_k^{OLS}$ the solution to the ordinary least-squares approximation. Hence, the ridge regression allows to impose an upper and a lower bound to the variability of the coefficients λ_j^k, which turns out to be a crucial feature when dealing with noisy data.

Remark 4.1 *In [8], it is established that a ridge regression regularization is equivalent to an ordinary least-squares approximation plus a constraint on the sum of the regression coefficients, which is demanded below a certain threshold $\mathcal{T}$. In particular, a precise one-to-one correspondence between $\mathcal{T}$ and T_{reg} is proved. In [9], the authors demonstrate that the ridge regression approach produces a better estimate for the regression coefficients with respect to the ordinary least-squares approximation in the case of data affected by noise and provided that T_{reg} is properly chosen. In particular, the superiority of the ridge regression form is proved in terms of mean squared error.*

4.2 Numerical assessment

To assess the effect of the ridge regression penalization on the multi-level least-squares approach, we generate synthetic data by adding noise to exact data and then we try to recover the original configuration. The `Matlab` built-in function `awgn` is employed to add the noise to the original coordinates of the points

Figure 11: The Stanford bunny model, reconstruction from noisy data: standard multi-level approach (left); ridge regression method (right).

in X. In particular, we consider a white Gaussian noise and we quantify the corresponding intensity with respect to the initial data via the signal-to-noise ratio (SNR). Moreover, we assume as compactly supported RBF function $\phi(r) = (1-r)_+^2$ [24].

We start from a very trivial set X of data that describes the surface of a sphere. The Gaussian noise added to these points is characterized by a SNR equal to $40dB$. We choose $Y \equiv X$ and $T_{reg} = 0.1$. Figure 10 compares the output provided by the standard multi-level approach and by the ridge regression variant. The reconstruction on the left exhibits many oscillations and confirms the poor performances of the interpolatory scheme. On the contrary, the regularized least-squares approach completely removes the noise and produces a very smooth surface.

As second test case, we consider the Stanford bunny model. We modify each point of the original data set by adding a Gaussian noise such that SNR is equal to $40dB$, and we set $Y \equiv X$ and $T_{reg} = 0.01$. Figure 11, left shows the very noisy reconstruction provided by the standard multi-level method. The interpolation step exactly reproduces the noise. The regularized least-squares approach delivers a reliable approximation of the original surface as displayed by the right panel in Figure 11.

We check the sensitivity of the regularized least-squares approximation with respect to the parameter T_{reg}. For this purpose, we analyze the trend of the error Θ as a function of T_{reg}, where Θ is computed as the average of the distances between the points on the reconstructed surface and the exact data distribution. Figure 12 shows the corresponding trend, together with the behaviour of Θ for $T_{reg} = 0$ (i.e., for the multi-level ordinary least-squares approximation). It is evident that, despite for a limited window of values for T_{reg}, the ridge regression approach always furnishes a better approximation. Moreover, as expected from

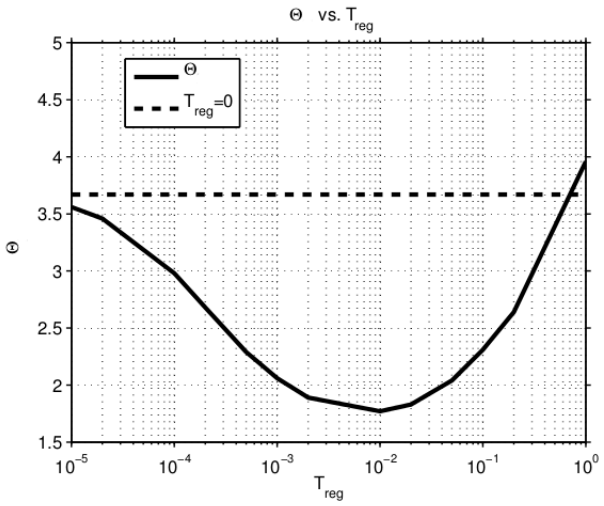


Figure 12: Reconstruction of the Stanford bunny model from noisy data: error trend as a function of T_{reg} for the ridge regression method (solid line) and for the standard multi-level approach (dashed line).

the theory, when T_{reg} tends to zero, the error approaches the one obtained via the multi-level ordinary least-squares method.

Finally, we have numerically investigated the dependence of the reconstruction on the selected set Y. Nevertheless, both the test cases exhibit a low sensitivity to this choice.

5 Conclusions

The numerical assessment tackled in the previous sections confirms the good performances of both new algorithms. They provide the expected improvements on the multi-level interpolating approach proposed in [15].

The validation in Section 3.2.1 shows that the adaptive procedure furnishes an accurate and parsimonious representation of surface Σ. In more detail, on the brain model, the adaptive approach is able to guarantee a certain accuracy by employing about one fourth of the points demanded by the standard algorithm, and, viceversa, the accuracy of the surface reconstructed via the adaptive technique is considerably higher with respect to the one furnished by the reference procedure and after sampling about the same number of points.
A cross comparison between Tables 1 and 3 on the Stanford bunny model highlights also the higher sparsity of the matrix A_k and a reduction of the corresponding condition number in the adaptive case. Moreover, despite the additional time required for the check driving the adaptive procedure, the computational time characterizing the standard approach is higher with respect to the one demanded by the adaptive algorithm.
Finally, the adaptive procedure is extremely robust also when dealing with the

complex data provided by the medical measurements in Section 3.2.2. As shown by Figures 7 and 8, the adaptive algorithm outperforms with respect to VMTK interpolation routines, by exhibiting a strong effectiveness even in the presence of details at different scales and of non-uniformly distributed data.

Section 4.2 confirms the good quality of the results provided by the regularized least-squares approach, which completely removes the noise and yields very smooth surfaces. A low sensitivity with respect to the selected set Y is another important feature detected by the numerical validation.

A possible follow-up of this work is an appropriate merging of the two proposed approaches to tackle large sparse and noisy data sets, as well as the proposal of an adaptive method to deal with sharp features. This last issue is of great interest in the current scientific community [6, 5].

6 Acknowledgments

The medical models derive from medical acquisitions at Ospedale di Borgo Trento di Verona (courtesy of G.B. Luciani, G. Puppini - Università degli Studi di Verona) and at Policlinico di Milano (courtesy of M. Domanin, G. Biondetti, L. Forzenigo - Policlinico di Milano e Università degli Studi di Milano).
The authors thank Elena Faggiano (Università di Pavia) for the help in elaborating medical data and for the brain model, Fabio Manganini and Luca Formaggia for the fruitful discussions.
Finally, the financial support of MIUR (Project "Innovative Methods for Water Resources under Hydro-Climatic Uncertainty Scenarios", PRIN 2010/2011) is gratefully acknowledged by the second author.

References

[1] The Stanford 3D Scanning Repository. http://graphics.stanford.edu/data/3Dscanrep/.

[2] J. Bloomenthal. Polygonization of implicit surfaces. *Comput. Aided Geo. Design*, 5(4):341–355, 1988.

[3] J. Bloomenthal. An implicit surface polygonizer. In *Graphics Gems IV*, pages 324–349. Academic Press Professional, Inc. San Diego, CA, USA, 1994.

[4] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *SIGGRAPH '01 Proceedings of the 28th annual conference on Computer Graphics and Interactive Techniques*, pages 67–76. ACM New York, NY, USA, 2001.

[5] G. Casciola, D. Lazzaro, L.B. Montefusco, and S. Morigi. Fast surface reconstruction and hole filling using positive definite radial basis functions. *Numer. Algorithms*, 39(1-3):289–305, 2005.

[6] G. Casciola, D. Lazzaro, L.B. Montefusco, and S. Morigi. Shape preserving surface reconstruction using locally anisotropic radial basis function interpolants. *Comput. Math. Appl.*, 51(8):1185–1198, 2006.

[7] N. A. Gumerov and R. Duraiswami. Fast radial basis function interpolation via preconditioned Krylov iteration. *SIAM J. Sci. Comput.*, 29:1876–1899, 2007.

[8] H. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, second edition, 2001.

[9] A.E. Hoerl and R.W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[10] A. Iske. *Multiresolution Methods in Scattered Data Modelling*. Springer-Verlag, Berlin, 2004.

[11] LAPACK, Linear Algebra PACKage. http://www.netlib.org/lapack.

[12] W. E. Lorensen and H. E. Cline. Marching cubes: a high resolution 3D surface construction algorithm. In *SIGGRAPH '87 Proceedings of the 14th annual conference on Computer Graphics and Interactive Techniques*, pages 163–169. ACM New York, NY, USA, 1987.

[13] B.S. Morse, T.S. Yoo, P. Rheingans, D.T. Chen, and K.R. Subramanian. Interpolating implicit surfaces from scattered surface data using compactly supported radial basis functions. In *IEEE, International Conference on Shape Modeling and Applications (SMI'01)*, pages 89–98, 2001.

[14] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H. P. Seidel. Multi-level partition of unity implicits. *ACM Transactions on Graphics (TOG)*, 22(3):463–470, 2003.

[15] Y. Ohtake, A. Belyaev, and H. P. Seidel. A multi-scale approach to 3d scattered data interpolation with compactly supported basis functions. In *IEEE, International Conference on Shape Modeling and Applications (SMI'03)*, pages 153–161, 2003.

[16] Y. Ohtake, A. Belyaev, and H. P. Seidel. 3d scattered data approximation with adaptive compactly supported radial basis functions. In *IEEE, International Conference on Shape Modeling and Applications (SMI'04)*, pages 31–39, 2004.

[17] OpenNL, Open Numerical Library. alice.loria.fr/index.php/software/4-library/23-opennl.html.

[18] T.J. Rivlin. *An Introduction to the Approximation of Functions.* Dover Publications, Inc., Mineola, New York, 2003.

[19] Y. Saad. *Iterative Methods for Sparse Linear Systems.* Society for Industrial and Applied Mathematics, Philadelphia, PA, second edition, 2003.

[20] G. Turk and J. F. O'Brien. Modelling with implicit surfaces that interpolate. *ACM Transactions on Graphics (TOG)*, 21(4):855–873, 2002.

[21] JG. Turk and J. F. O'Brien. Shape transformation using variational implicit surfaces. In *SIGGRAPH '99 Proceedings of the 26th annual conference on Computer Graphics and Interactive Techniques*, pages 335–342. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA, 1999.

[22] VMTK, The Vascular Modeling Toolkit. http://www.vmtk.org/.

[23] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial basis functions of minimal degree. *Adv. Comput. Math.*, 4:389–396, 1995.

[24] H. Wendland. *Scattered Data Approximation.* Cambridge University Press, Cambridge, 2005.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

01/2016 Domanin, M.; Buora, A.; Scardulla, F.; Guerciotti, B.; Forzenigo, L.; Biondetti, P.; Vergara, C.
Computational fluid-dynamic analysis of carotid bifurcations after endarterectomy: closure with patch graft versus direct suture

62/2015 Signorini, M.; Zlotnik, S.; Díez, P.
Proper Generalized Decomposition solution of the parameterized Helmholtz problem: application to inverse geophysical problems.

63/2015 Lancellotti, R.M.; Vergara, C.; Valdettaro, L.; Bose, S.; Quarteroni, A.
Large Eddy Simulations for blood fluid-dynamics in real stenotic carotids

61/2015 Tagliabue, A.; Dedè, L.; Quarteroni, A.
Fluid dynamics of an idealized left ventricle: the extended Nitsche's method for the treatment of heart valves as mixed time varying boundary conditions

59/2015 Menafoglio, A.; Guadagnini, A.; Secchi, P.
Stochastic Simulation of Soil Particle-Size Curves in Heterogeneous Aquifer Systems through a Bayes space approach

60/2015 Perotto, S.; Reali, A.; Rusconi, P.; Veneziani, A.
HIGAMod: A Hierarchical IsoGeometric Approach for MODel reduction in curved pipes

58/2015 Iapichino, L.; Rozza, G.; Quarteroni, A.
Reduced basis method and domain decomposition for elliptic problems in networks and complex parametrized geometries

57/2015 Wilhelm, M.; Dedè, L.; Sangalli, L.M.; Wilhelm, P.
IGS: an IsoGeometric approach for Smoothing on surfaces

54/2015 Canuto, C.; Nochetto, R. H.; Stevenson, R.; Verani, M.
Adaptive Spectral Galerkin Methods with Dynamic Marking

55/2015 Fumagalli, A.; Zonca, S.; Formaggia, L.
Advances in computation of local problems for a flow-based upscaling in fractured reservoirs