

## MULTISCALE ANALYSIS OF MULTIVARIATE DATA

**Angela Kunoth**

*Institut für Mathematik  
Universität Paderborn  
Warburger Str. 100  
33098 Paderborn  
Germany*

E-mail: kunoth@math.uni-paderborn.de

**Keywords:** Scattered data fitting, irregularly distributed data, multiscale basis, wavelet, adaptivity, least squares method, fast solution, regularization, outlier treatment, harmonic function, geodesy.

**Abstract.** *For many applications, nonuniformly distributed functional data is given which lead to large-scale scattered data problems. We wish to represent the data in terms of a sparse representation with a minimal amount of degrees of freedom. For this, an adaptive scheme which operates in a coarse-to-fine fashion using a multiscale basis is proposed. Specifically, we investigate hierarchical bases using B-splines and spline-(pre)wavelets. At each stage a least-squares approximation of the data is computed. We take into account different requests arising in large-scale scattered data fitting: we discuss the fast iterative solution of the least square systems, regularization of the data, and the treatment of outliers. A particular application concerns the approximate continuation of harmonic functions, an issue arising in geodesy.*

# 1 INTRODUCTION

Assume that there is a set  $X = \{x_i\}_{i=1,\dots,N}$  of irregularly spaced and pairwise distinct points  $x_i \in \overline{\Omega} := [0, 1]^n$  where  $n \in \{1, 2, 3\}$  denotes the spatial dimension and the amount of points  $N \in \mathbb{N}$  is large. For each  $i$  let  $z_i \in \mathbb{R}$  be corresponding data which is supposed to stem from a multivariate function and which is allocated in a set  $Z$ . The problem of scattered data fitting is the following. Given the cloud of points  $P := (X, Z) = \{(x_i, z_i), i = 1, \dots, N\}$ , find a function  $f : \overline{\Omega} \rightarrow \mathbb{R}$  that approximates  $P$  in a least squares sense with regularization, that is,  $f$  minimizes the functional

$$\mathcal{J}(f) := \sum_{i=1}^N (z_i - f(x_i))^2 + \nu \mathcal{F}(f). \quad (1)$$

Here  $\mathcal{F}(f)$  denotes some quadratic functional of  $f$  which will be specified below. The first term in (1) is denoted as *data fidelity term* and is responsible for approximating the data. The second term requires  $f$  to be in some smoothness space which is a subspace of  $L_2(\Omega) \cap C^0(\overline{\Omega})$ , therefore called *regularization term*. Finally, the parameter  $\nu \geq 0$  balances these two requirements.

We will construct an expansion of the minimizer  $f$  of (1) of the form

$$f(x) = \sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda(x), \quad x \in \overline{\Omega}. \quad (2)$$

There are numerous possibilities for selecting the expansion functions  $\psi_\lambda$ : B- or box splines on uniform or nonuniform grids, trigonometric functions, radial basis functions, different classes of multiscale basis or wavelets or multiwavelets. We want to choose those which best match the following ‘wish list’: First, the data set is presumably large and irregularly spaced, may represent a nonsmooth function and may contain outliers (false data). So we should insist on  $\psi_\lambda$  with a possibly small support in order to be able to localize the representation. Second, we envisage iterative solution methods for the least squares problem arising from (1) in view of a large  $N$ . So the  $\psi_\lambda$  should give rise to favorable spectral condition numbers of these systems. Third, we wish to devise a unified method for very different sets of data. Hence, approaches based, e.g., on triangulations of  $X$  as an auxiliary step or based on interpolating  $P$  are not considered, partly also due to the fact that one would have to adapt these ingredients repeatedly whenever an outlier in the data is identified. Fourth, the approximation scheme should allow for a data-driven multiscale approximation, allowing very different resolutions in the data and a final error depending only on the data. In this context, the expansion coefficients  $d_\lambda$  should have some meaning in the sense that large values indicate important contributions and small coefficients are negligible. Fifth, the functions  $\psi_\lambda$  should in some sense characterize the smoothness space enforced by the functional  $\mathcal{F}(f)$ .

In view of this ‘wish list’, we propose to use as building blocks for  $\psi_\lambda$  tensor products of certain boundary-adapted spline-prewavelets from [45], shortly called ‘wavelets’ here, see, e.g., [11] for a general introduction into wavelet theory. Their properties are listed below in Section 2. These wavelets are defined on structured grids and the set of all these wavelets constitute a Riesz basis for  $L_2(\Omega)$ . The algorithm proposed below in Section 3 works with a subset constructed in an adaptive coarse-to-fine fashion. Thus, the index set  $\Lambda$  employed in the representation (2) actually turns out to be a lacunary set of all possible indices so that the set  $\{\psi_\lambda\}_{\lambda \in \Lambda}$  is a particular multiscale basis adapted to the distribution and vertical variety of the data. The proposed scheme is of minimal complexity in the following sense. As we do *not* assume that

the data lies on a uniform grid, we do not want to associate an (often artificially chosen) ‘finest’ grid  $\Omega_J$  with the data. Therefore, any ‘adaptive’ method that requires in a first step to apply the Fast Wavelet Transform on such a uniform grid  $\Omega_J$  followed by some thresholding (as it was done in [24] for data given on uniform grids) is in this sense not an appropriate strategy: There the full complexity  $\#\Omega_J$  of this grid would enter into the algorithm independent of the cardinality or distribution of the data points. A particular feature of our new approach is that the selection of wavelet bases relevant for the approximation is independent of the particular  $P$  when compared to schemes based on, e.g., data dependent triangulations, see, e.g, [20]. Our scheme which was developed initially in [6] can be seen as one of the approaches on structured grids which are by many considered a universally applicable and well–established procedure in scattered data fitting, see, e.g., [26, 27, 30, 35, 37, 42, 43].

Multiscale data fitting with or without smoothing has been discussed widely in the literature. To name only a few, in [22] a coarse–to–fine strategy has been presented with hierarchical splines on structured grids which is suited for parameterized gridded data. In [35] scattered functional data is approximated by multilevel B–splines. For data fitting problems on unstructured grids, radial basis methods provide good approximating properties [40], a rich theoretical characterization [3, 39] and a mesh free approach, which appears quite natural when working with unorganized data [4]. However, corresponding data reduction strategies like thinning [21] or adaptive thinning [20] usually require the construction of some auxiliary triangulation in order to achieve a multiscale formulation of the problem. Interpolating multivariate scattered data in higher dimensions is a somewhat different task where radial basis functions or other meshless methods may be particularly useful, see, e.g., [3, 39, 40, 41, 44, 47]. An up-to-date account of issues and techniques from nonlinear approximation, multiscale analysis and adaptive algorithms for different applications such as curve fitting or partial differential equations is provided by [18].

Our algorithm operates in a coarse–to–fine fashion solely governed by the distribution of the data. It adaptively constructs least–squares approximations to the data. Followed by thresholding irrelevant wavelet coefficients, we achieve a representation of the data which is minimally sparse. We systematically exploit the concept of adaptivity which is by now a well–established and powerful methodology to finding sparse representations of large–scale scientific problems. In principle, our scheme allows also for approximating data from spatial dimensions higher than  $n = 3$  as long as storage permits.

The remainder of the material is organized as follows. In the next section, we recall the main ingredients of wavelet bases used here. Section 3 describes the adaptive data fitting algorithm. We address issues of numerical performance, regularization and robust treatment of outliers in the data in Sections 4 through 6. We conclude with an example from geophysics where the goal is to reconstruct a harmonic function which is done by including the Laplacian into the functional  $\mathcal{F}$  in (1).

## 2 WAVELET BASES

The main ingredient of our scattered data fitting algorithm will be an appropriate wavelet basis for a Hilbert space  $H \subseteq L_2(\Omega)$  living on a bounded domain  $\Omega \subset \mathbb{R}^n$ . A single multivariate wavelet function will be indexed by a parameter  $\lambda = (j, \mathbf{k}, \mathbf{e})$  which encodes different information such as the *resolution level* or *scale*  $|\lambda| := j$ , the *spatial location*  $\mathbf{k}$ , and the type  $\mathbf{e} \in \{0, 1\}^n$  of wavelet for  $n > 1$ . This accounts to the fact that in the multivariate case we will be using

tensor products, see, e.g., [13, 16, 45]. Since the domain  $\Omega$  is bounded, there is a coarsest resolution level denoted by  $j_0$ . Functions with index  $\mathbf{e} = \mathbf{0}$  only occur on the coarsest level; they are *scaling functions* and consist of tensor products of B-splines. Functions with index  $\mathbf{e} \neq \mathbf{0}$  represent detail information of higher frequencies. The infinite set of all possible indices will be denoted by  $\mathcal{I}$ . We call the collection of functions

$$\Psi := \{\psi_\lambda : \lambda \in \mathcal{I}\} \subset H \quad (3)$$

a *wavelet basis* provided it has the following properties:

**Riesz basis property (R):**  $\Psi$  constitutes a *Riesz basis* for  $H$ : every  $v \in H$  has a unique expansion in terms of  $\Psi$ ,

$$v = \sum_{\lambda \in \mathcal{I}} v_\lambda \psi_\lambda =: \mathbf{v}^T \Psi, \quad \mathbf{v} := (v_\lambda)_{\lambda \in \mathcal{I}}, \quad (4)$$

and its expansion coefficients satisfy a *norm equivalence* of the following form. There exist finite positive constants  $c_H \leq C_H$  such that for any  $\mathbf{v} = \{v_\lambda : \lambda \in \mathcal{I}\}$  one has

$$c_H \|\mathbf{v}\|_{\ell_2(\mathcal{I})} \leq \|\mathbf{v}^T \Psi\|_H \leq C_H \|\mathbf{v}\|_{\ell_2(\mathcal{I})}.$$

We will write this relation shortly as

$$\|\mathbf{v}\|_{\ell_2(\mathcal{I})} \sim \|\mathbf{v}^T \Psi\|_H. \quad (5)$$

This means that wavelet expansions induce isomorphisms between the function space  $H$  and the sequence space  $\ell_2(\mathcal{I})$ . For the applications considered here,  $H$  will be a Sobolev space  $H^\alpha(\Omega)$  of smoothness  $\alpha \geq 0$ . A particular feature of wavelet bases is that such norm equivalences hold, in fact, for a whole range of Sobolev spaces  $\alpha \in (-\tilde{\gamma}, \gamma)$  depending on the smoothness of the bases and the cancellation properties (CP) mentioned below, with an appropriate scaling of the expansion coefficients by a diagonal matrix, see, e.g., [13].

**Locality (L):** The functions  $\psi_\lambda$  are compactly supported: for each  $\lambda \in \mathcal{I}$  one has

$$\text{diam}(\text{supp } \psi_\lambda) \sim 2^{-|\lambda|}. \quad (6)$$

**Cancellation Property (CP):** There exists an integer  $\tilde{m}$  such that

$$\langle v, \psi_\lambda \rangle \lesssim 2^{-|\lambda|(n/2+\tilde{m})} |v|_{W_\infty^{\tilde{m}}(\text{supp } \psi_\lambda)}, \quad (7)$$

where  $\langle \cdot, \cdot \rangle$  denotes the dual form between  $H$  and its topological dual  $H'$  and  $|v|_{W_\infty^{\tilde{m}}(\text{supp } \psi_\lambda)}$  is the seminorm containing all weak derivatives of  $v$  of order  $\tilde{m}$  in  $L_\infty(\text{supp } \psi_\lambda)$ . Here  $a \lesssim b$  means that  $a$  can be estimated from above by a constant multiple of  $b$  independent of all parameters on which  $a$  or  $b$  may depend.

Property (CP) means that integration against a wavelet acts like taking an  $\tilde{m}$ th order difference which annihilates the smooth part of  $v$ . Thus, the expansion coefficients are small where  $v$  is smooth.

Together, properties (R), (L) and (CP) allow for the proof of numerous theoretical statements ranging from optimal spectral condition number estimates of elliptic operators in wavelet bases [14] to convergence results of adaptive wavelet methods for linear and nonlinear variational problems [12, 18] and PDE-constrained elliptic control problems [15, 34]. At the same time, these properties still allow to work with piecewise polynomials which is an essential advantage

for numerical computations. Among the various constructions of different classes of wavelets, it has turned out that for data fitting problems tensor products of the piecewise linear boundary-adapted B-spline pre-wavelets from [45] are a good compromise for meeting the ‘wish list’ from Section 1 since they have the additional property

**Semi-Orthogonality (O)** with respect to  $L_2(\Omega)$ , i.e., for  $|\lambda| \neq |\mu|$  one has

$$\int_{\Omega} \psi_{\lambda}(x) \psi_{\mu}(x) dx = 0. \quad (8)$$

The semi-orthogonality between functions from different resolution levels will be very useful for the fast iterative solution of the linear systems appearing in the scattered data context.

### 3 BASIC COARSE-TO-FINE ALGORITHM

Our algorithm described in this section was originally developed in [6] and is designed in such a way that we can handle very large data sets. Our goal is to work with as few wavelets as possible, taking into account only those basis functions which are necessary to resolve the data given by  $P$ . As only finitely many data are to be approximated, only finitely many indices from  $\mathbb{I}$  are essential for the representation of the minimizer  $f$  of (1). For a given cloud of points  $P$ , the determination of an appropriate index set  $\Lambda \subset \mathbb{I}$ ,  $\#\Lambda < \infty$ , is therefore the central issue. On the one hand,  $\Lambda$  should be rich enough to be able to represent all features of  $P$  while at the same time any *overfitting artefacts* as well as processing *redundant data* should be avoided. Our basic adaptive coarse-to-fine algorithm will be described for the case that  $P$  is not noisy and does not contain outliers, and we describe it for the case of the least squares functional (1) without regularization, i.e., for  $\nu = 0$ .

Before giving the technical description of our scheme, it may be helpful to list several properties. The choice of relevant indices in the index set  $\Lambda$  is determined in such a way that it progressively adapts to the data, without introducing at any point a ‘finest’ artificial full grid. The finest resolution is only determined by the data. In each step of the scheme, a least squares approximation is computed by solving the normal equations by conjugate gradient (CG) iterations (this issue is discussed in detail below in Section 4). After this step, an appropriate thresholding on the wavelet coefficients guarantees that only relevant wavelet coefficients are kept.

The scheme operates in a coarse-to-fine fashion as follows. On the coarsest level  $j_0$ , all indices contained in  $\mathbb{I}$  are selected; these indices correspond to the generator functions on this level. Thanks to the scale localization property (L), we progressively construct ‘from coarse to fine’ a sequence of (nested) index sets  $\Lambda_j \subset \mathbb{I}$ ,  $j \geq j_0$ , in such a way that we only consider those wavelets whose supports contain enough data points, parameterized by some number  $q$ . From a set of indices  $\Lambda_j$  we construct a new set  $\Lambda_{j+1}$  by considering only children of previously chosen wavelets, with the additional requirement that those children must have at least  $q$  points in their support. In this way, we generate a tree as data structure for  $\Lambda$ . Denote by  $\delta(\Lambda_j)$  the indices of newly allowable candidates and by  $S_q(\delta(\Lambda_j)) \subset \delta(\Lambda_j)$  their subset of indices for wavelets having at least  $q$  points in their support. After an appropriate index set  $\tilde{\Lambda}_{j+1}$  is found, the minimum of the least squares functional

$$\mathcal{J}(f_{j+1}) := \sum_{i=1}^N \left( z_i - \sum_{\lambda \in \tilde{\Lambda}_{j+1}} d_{\lambda}^{j+1} \psi_{\lambda}(x_i) \right)^2 \quad (9)$$

is computed by CG iterations on the normal equations. Recall that we want to find a minimally sparse representation of  $P$ . So we afterwards make in each step use of the norm equivalence (5) by *thresholding* to discard those wavelet coefficients which are smaller than a certain threshold  $\varepsilon$ .

ALGORITHM: FIX  $q$  AND A THRESHOLD  $\varepsilon$  AND DO FOR EACH  $j = j_0, \dots$

- (I) DETERMINE  $\delta(\Lambda_j)$  AND SELECT  $S_q(\delta(\Lambda_j))$ ,
- (II) IF  $S_q(\delta(\Lambda_j)) = \emptyset$  STOP, ELSE SET  $\tilde{\Lambda}_{j+1} := \Lambda_j \cup S_q(\delta(\Lambda_j))$ ,
- (III) COMPUTE  $\{d_\lambda^{j+1}\}_{\lambda \in \tilde{\Lambda}_{j+1}}$  BY MINIMIZING  $\mathcal{J}(f_{j+1})$  DEFINED IN (9),
- (IV) DETERMINE  $\Lambda_{j+1}^\varepsilon := \{\lambda \in \tilde{\Lambda}_{j+1} \setminus \Lambda_j : |d_\lambda^{j+1}| \geq \varepsilon\}$ ,
- (V) IF  $\Lambda_{j+1}^\varepsilon = \emptyset$  STOP, ELSE SET  $\Lambda_{j+1} = \Lambda_j \cup \Lambda_{j+1}^\varepsilon$  AND  $j = j + 1$ .

The algorithm terminates at a *finest level*  $J$  with an index set  $\Lambda_J$  which only depends on the density and the distribution of the data. Note that  $J$  is neither user-specified nor previously chosen. In view of (L), for given finite data  $X$  the algorithm always terminates since either  $S_q(\delta(\Lambda_j))$  or  $\Lambda_{j+1}^\varepsilon$  are at some point empty. The algorithm grows exactly one level in each step so that the highest resolution level contained in the index set  $\Lambda_J$  is  $J$ . The final index set  $\Lambda_J$  has tree structure and the representation is sparse in the sense that it provides a least squares approximation to the data with a minimal amount of degrees of freedom.

After the index set for possible wavelet candidates  $\Lambda = \tilde{\Lambda}_{j+1}$  in (2) is fixed, the necessary conditions for minimizing  $\mathcal{J}$  in (9) yield the *normal equations*

$$M\mathbf{d} = \mathbf{b}, \quad M := A^T A, \quad \mathbf{b} := A^T \mathbf{z}, \quad (10)$$

where  $\mathbf{d}$  and  $\mathbf{z}$  are the column vectors for the coefficients and the given values in  $Z$ , respectively. The *observation matrix*  $A \in \mathbb{R}^{N \times (\#\Lambda)}$  has entries  $A_{i,\lambda} = \psi_\lambda(x_i)$  so that the *cross product matrix*  $M = A^T A$  and right hand side  $\mathbf{b} = (b_\lambda)_{\lambda \in \Lambda}$  are given as

$$M_{\lambda,\lambda'} = \sum_{i=1}^N \psi_\lambda(x_i) \psi_{\lambda'}(x_i), \quad b_\lambda = \sum_{i=1}^N z_i \psi_\lambda(x_i). \quad (11)$$

Note that  $M$  can be interpreted as a discretized Gramian matrix since the  $M_{\lambda,\lambda'}$  have the form of a discretization of the  $L_2$  inner product. It is not immediately clear that  $M$  resulting from the above algorithm is indeed invertible. Sufficient conditions guaranteeing  $A$  to have full rank, which blend into the above scheme and which are based on the Theorem of Schoenberg and Whitney for B-Splines are provided in [5].

As an example, the performance of ALGORITHM is tested on the 3D data set from [5] displayed on the left of Fig. 1. The reconstruction with the full coarsest grid  $j = 3$  provides already a good reconstruction in smooth areas but fails to do so in the neighborhood of the ring where high gradients occur. New wavelets on level  $j = 4$  are only chosen in this area and, as can be seen on the right of Fig. 2, provide a more accurate reconstruction.

An interesting application of the above least squares data fitting approach based on adaptive wavelets for the approximate evaluation of *nonlinearities* is studied in [9]. There the data points  $P$  are chosen depending on the type of nonlinearity. This scheme is then used to compute the solution of Burgers' equation by means of an adaptive wavelet viscosity method [9].

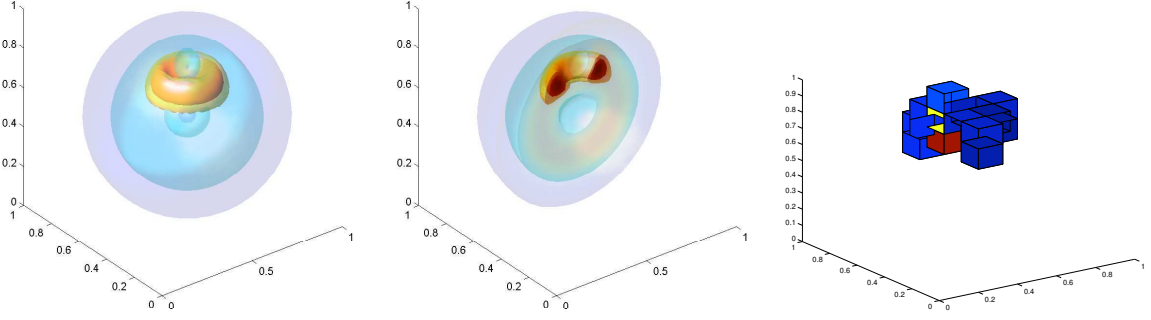


Figure 1: Left: Original data set consisting of  $41^3$  points displaying the result of a simulation of the two-body distribution probability of nucleon-nucleon correlations for  $^{16}\text{O}$  (Courtesy of VolVis and SFB 382 of the DFG). Middle: Isosurface at  $y = 0.5$ . Right: Distribution and size of wavelet coefficients for level  $j = 3$  and wavelet type  $e = (1, 1, 1)$ ; coefficients range from small (blue/dark) to large (red/light).

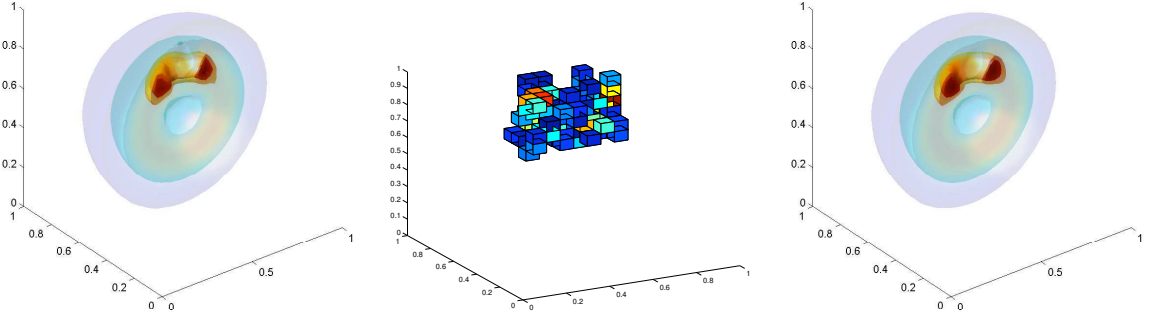


Figure 2: Left: Isosurface of reconstruction at level  $j = 3$ . Middle: Distribution and size of wavelet coefficients for level  $j = 4$  and wavelet type  $e = (1, 1, 1)$ . Right: Isosurface of reconstruction at level  $j = 4$  with 2107 new degrees of freedom as opposed to the 35937 possible ones.

So far we have not taken into account any statistical assumption on the distribution of the given data. It would be very interesting to combine the present approach with techniques from statistical learning theory where in addition to the regression function  $f$  the unknown probability distribution is to be estimated, see, e.g., [1].

## 4 NUMERICAL SOLUTION

A crucial step in our adaptive ALGORITHM is the computation of the minimizer of the least squares functional (9). Such least squares problems are often solved by computing a QR decomposition of  $A$ . In the present context, we prefer to solve the resulting normal equations (10) for the following reasons.

First, for very large data sets as the one considered in Fig. 1, we usually obtain  $\#\Lambda \ll N$ . This means that we need much less space for storing  $M$  which is of size  $\mathbb{R}^{(\#\Lambda) \times (\#\Lambda)}$  than for  $A$ , where the amount of data  $N$  enters as the number of rows. In this sense, forming the normal equations entails in fact some *data compression* effect, as a large  $N$  only appears as the size of the sums in the entries of  $M_{\lambda,\mu} = \sum_{i=1}^N \psi_\lambda(x_i) \psi_\mu(x_i)$  and not in the size of the matrix  $M$ .

Second, for the particular wavelet basis we use, the spectral condition number of  $M$  is very moderate. This may be attributed to the fact that  $M$  is a *discretization* of the Gramian matrix with entries  $(\psi_\lambda, \psi_\mu)_{L_2(\Omega)}$  which has in view of the  $L_2$ -stability of the wavelet basis (5) a uniformly bounded condition number independent of the finest level of resolution. For growing  $N$ , the entries of  $M$  ‘approximate’ the entries of the Gramian even better, meant as a trend. (Of course, a concrete approximation estimate between the matrix entries would depend on the distribution of points.) Consequently, iterative methods like a conjugate gradient (CG) scheme on  $M$  are expected to reduce the error in each step by a fixed amount and the number of iterations to reach a prescribed accuracy are envisaged to be very moderate. This has been confirmed in all numerical experiments we have conducted. We have also compared in [7] approximation errors of our results with the normal equations with approximation errors for a least squares solution computed using  $QR$  decomposition of  $A$  and found only negligible differences, while the iterations took much less time for the normal equations. Further details on the efficient numerical solution of least squares problems can be found in [2].

Third, in view of the tree structure generated by the adaptive algorithm, the matrix  $M$  on level  $j$  is a submatrix of  $M$  on level  $j + 1$ ; the latter only contains additional entries with respect to wavelet functions  $\psi_\lambda$  for  $|\lambda| = j + 1$ . Thus, one expects a *nested iteration* algorithm which takes the solution vector  $\mathbf{d}$  from level  $j$ , padded with zeroes, as initial guess for the CG iteration on level  $j + 1$  to perform particularly well. In addition, the ‘almost’ orthogonality between levels in  $M$  which is a result of the semi-orthogonality property (O) provides a strong decoupling between levels, so that a solution on level  $j$  is on its low frequencies only marginally altered by the iteration on level  $j + 1$ .

We illustrate these particular features of our wavelet basis by comparing it with another standard multiscale basis, the *hierarchical basis* which has proven to be very popular when solving elliptic PDEs in variational form [48]. In one dimension, it consists of the scaled versions of the linear B-spline, also called ‘hat function’. This basis is not stable with respect to  $L_2$  and, therefore, the spectral condition number of the Gramian grows like  $2^j$  [5]. Comparisons between the hierarchical B-Spline basis and a wavelet basis were made in [25] in the context of geometric modeling. In computer graphics, a hierarchical B-spline approach was studied in [22]. As an example for the different performance of the wavelet basis used here and the hierarchical basis, Fig. 3 from [5, 6] is displayed, where these issues have been discussed in detail. Very few CG iterations suffice for the wavelet basis to reach the saturation error  $10^{-2}$ .

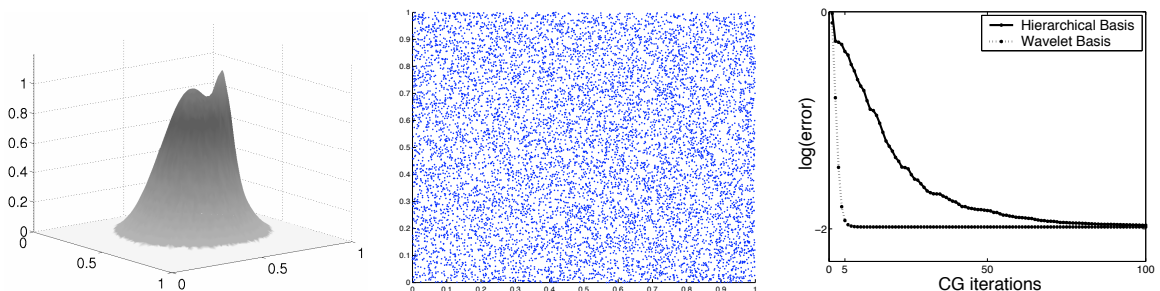


Figure 3: Error decay for  $J = 4$  (right) for the hierarchical basis and the wavelet basis using nested iteration for a data set of  $10^4$  randomly generated points (middle) generated from the surface on the left. For the semi-orthogonal wavelet basis, five iterations suffice to reach the saturation error  $10^{-2}$  while the hierarchical basis needs about 100 iterations.

One remark should be made concerning the different possible construction of tensor prod-



ucts from univariate functions. In principle, one can distinguish the isotropic case, where the supports of all wavelets from one level are cubes of approximately the same length in each direction, from the anisotropic case, see [5] for a precise description of the construction. In the numerical experiments also reported there, the spectral properties of  $M$  using one or the other construction are for the data fitting problem very similar and do not influence much the behavior of the CG scheme on the normal equation.

## 5 MULTILEVEL REGULARIZATION

Now we extend the least squares fitting strategy to include into the objective functional a Sobolev seminorm with positive smoothness index  $\alpha$  as a regularizing term to enforce a smooth approximation. This classical approach for  $\alpha \in \{1, 2\}$  was studied, e.g., in [19] for general splines, [27] for hierarchical splines, [29] for splines on triangulations, or [25] in a wavelet reformulation of the spline problem. Here we choose more generally  $\mathcal{F}(f) := \|f\|_{H^\alpha(\Omega)}^2$  so that we consider

$$\mathcal{J}(f) = \sum_{i=1}^N (z_i - f(x_i))^2 + \nu \|f\|_{H^\alpha(\Omega)}^2, \quad (12)$$

where a not necessarily integer  $\alpha > 0$  parameterizes the smoothing and  $\nu > 0$  balances the fidelity of the data and the smoothness. In view of the statements made after (5), the wavelet formulation allows to represent functions in Sobolev norms in terms of sequence norms of weighted coefficients from their wavelet expansions. Accordingly, following [10], we replace  $\|\cdot\|_{H^\alpha(\Omega)}^2$  by the corresponding sequence norm on the right hand side of (12),

$$\left\| \sum_{j,k,e} d_{j,k,e} \psi_{j,k,e} \right\|_{H^\alpha(\Omega)}^2 \sim \sum_{j,k,e} 2^{2\alpha j} |d_{j,k,e}|^2. \quad (13)$$

The minimization of (12) using this sequence norm is now equivalent to the solution of the normal equations

$$(M + \nu R)\mathbf{d} = \mathbf{b}. \quad (14)$$

Here  $R$  is a diagonal matrix with entries

$$R = \text{diag}(2^{2\alpha j_0} I, \dots, 2^{2\alpha J} I) \quad (15)$$

if  $J$  is the highest level of approximation and the sizes of the identity matrices depend on the cardinality of  $\Lambda_j$ . The representation of the smoothing term in wavelet coordinates has some interesting properties. First, the different dyadic scales are decoupled which means that the weak decoupling between levels in  $M$  is maintained, and the effect of the regularization term boils down to a penalization of the higher frequencies. Thus, the parameter  $\nu$  controls the balance between fidelity and regularization while  $\alpha$  controls the relative penalization across scales. Second, since (5) holds for the whole scale of fractional Sobolev spaces  $\{H^\alpha(\Omega)\}_{\alpha \in [0, \gamma]}$ , one has easy access to the entire scale which reduces to a simple diagonal scaling in the wavelet framework by means of the diagonal matrix  $R$ . Third, the formulation is independent of the wavelet family. Of course, the smoothness of the wavelets imposes a limit on the maximal value of  $\alpha$  for which (13) holds.

We have now two parameters in (12), the smoothness index  $\alpha$  and the weight parameter  $\nu$  to balance the data approximation and the smoothness part. This can be used to adjust as best

as possible to the given data. An example of an application from photogrammetry with fixed  $\alpha = 4$  and  $\nu = 0.01$  is provided in Figs. 4 and 5. It should be mentioned that for the linear pre-wavelets we employ, the norm equivalence (13) does not hold for this high value of  $\alpha$ . Nevertheless, the reconstruction in Fig. 5 gives the best result for this case.

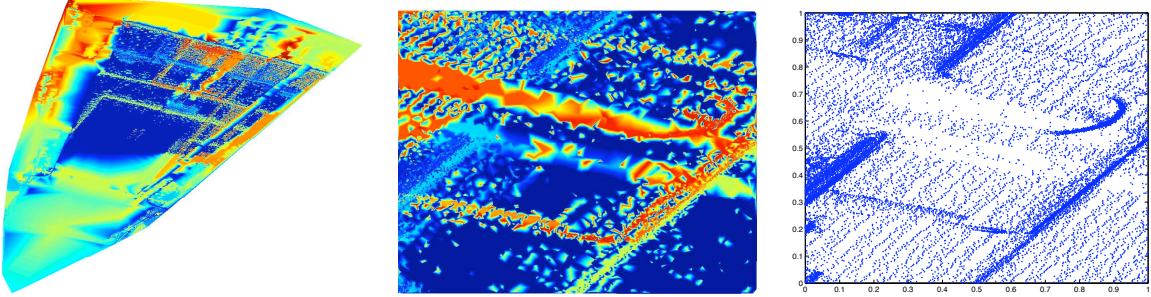


Figure 4: Left: Original data, a 3D point set consisting of 330.000 points of an industrial site (taken by Leica Cyrax 2500, Prof. Staiger, GH Essen). Middle: Vertical view of a section of the original data. Right: Sampling geometry of this section.

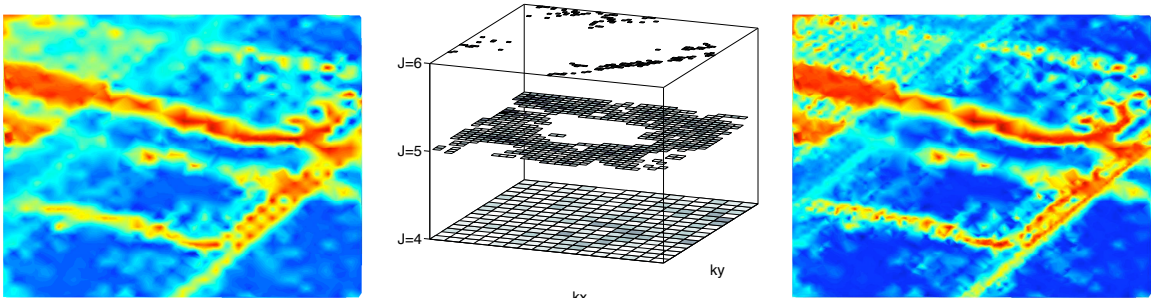


Figure 5: Approximation of the data from Fig. 4 by wavelet reconstruction with regularization for  $\nu = 0.01$ ,  $\alpha = 4$  and thresholding parameter  $\varepsilon = 10^{-3}$ . Left: Reconstruction for  $J = 4$ . Middle: Wavelet coefficients of type  $(1, 1)$ . Right: Reconstruction for  $J = 6$ . The final approximation contains  $\#\Lambda_6 = 2623$  coefficients as opposed to 16384 coefficients for the full grid.

Standard cross validation techniques start out with computing the weight parameter  $\nu$  for a fixed choice of  $\alpha$ , which requires to solve an additional system of normal equations. Below we once more exploit the multilevel framework provided by wavelets and introduce a multi-level version of the cross validation procedure. This scheme turns out to be both relatively inexpensive from a computational point of view as well as adjusting well to smoothness and to localization effects.

Let us first describe the classical situation where the smoothness parameter  $\alpha$  is fixed. Different ways to select an appropriate regularization parameter  $\nu$  in (12) were proposed, where one of the most popular choices are cross validation methods, see, e.g., [46]. Particularly, in the *Generalized Cross Validation* (GCV) method one computes  $\nu$  as the minimizer of the *GCV potential* with respect to the Euclidean norm  $\|\cdot\|$ , that is,

$$GCV(\nu) := \frac{\|I - H(\nu)\mathbf{z}\|^2}{(\text{tr}(I - H(\nu)))^2}, \quad (16)$$

where the *influence matrix*  $H(\nu)$  is defined as  $H(\nu) := A(M + \nu R)^{-1}A^T$ . The influence matrix is never computed explicitly, since only the application of  $H(\nu)$  to a vector needs to be carried

out. This in turn requires the solution of a linear system with matrix  $M + \nu R$  which is of a similar type as in (14). The trace in the denominator of (16) is usually estimated stochastically as  $\text{tr}(B) = \mathbf{u}^T B \mathbf{u}$  where  $\mathbf{u}$  is a random vector with entries 1 and  $-1$  and  $B$  is a quadratic symmetric matrix. Consequently, the evaluation of  $GCV(\nu)$  for each  $\nu$  requires the solution of two linear systems of the type (14). The total computation time then depends on the number of GCV evaluations which are needed to find the minimum of  $GCV(\nu)$ . We have displayed in [7] a series of bivariate examples where we have applied these GCV techniques and where the solution of each linear system costed less than one second.

The GCV technique yields an optimal  $\nu$  in the sense that it minimizes the variance under the assumption that the data has been corrupted by white noise. This theory developed for splines immediately carries over to the wavelet case [5]. Moreover, as shown in [7] the underlying heuristics can also be applied successfully to other data fitting approaches which require some regularization such as for data sets with a highly varying point distribution or for data with holes in the domain where typically overfitting artifacts occur.

The main issue in [7] is the following multilevel version of the GCV algorithm used within the context of the adaptive coarse-to-fine ALGORITHM. The idea is to simultaneously use the GCV with the hierarchical growth of the wavelet tree as the levels become higher. We have observed above that the wavelet representation penalizes the different dyadic levels separately with the same weight  $2^{2\alpha j}$ . In view of the structure of the diagonal matrix  $R$  given in (15), we prescribe instead a diagonal penalizing matrix with the same values  $\nu_j \geq 0$  for all entries of each level  $j$ , where the  $\nu_j$  are computed independently following a GCV criterion. Now at each level  $j$  the normal equations attain the form

$$(A_{\Lambda_j}^T A_{\Lambda_j} + R_j)d = A_{\Lambda_j}^T z, \quad (17)$$

where the diagonal matrix  $R_j$  is defined component-wise as  $(R_j)_{\lambda, \lambda'} := \delta_{\lambda, \lambda'} \nu_j$  for some set of scalars  $\{\nu_{j'}\}_{j_0 \leq j' \leq j}$ . These scalars are computed inductively at each level following a GCV criterion. At level  $j_0$ , we define the influence matrix

$$H(\nu_{j_0}) := A_{\Lambda_{j_0}} (A_{\Lambda_{j_0}}^T A_{\Lambda_{j_0}} + R_{j_0})^{-1} A_{\Lambda_{j_0}}^T,$$

and  $\nu_{j_0}$  is obtained by minimization of the corresponding GCV potential (16). At any subsequent level  $j$  the influence matrix is defined as

$$H(\nu_j; \nu_{j_0}, \dots, \nu_{j-1}) := A_{\Lambda_j} (A_{\Lambda_j}^T A_{\Lambda_j} + R_j)^{-1} A_{\Lambda_j}^T \quad (18)$$

to have  $\nu_j$  as only variable since the  $\nu_{j'}$  from previous levels  $j_0 \leq j' < j$  are already computed.  $\nu_j$  is then determined by minimizing (16). Recall that the influence matrices are never computed explicitly as they only need to be applied to a vector. The new penalizing term can no more be interpreted as stemming from a smoothing term  $\|f\|_{H^\alpha(\Omega)}^2$  since no relation between the penalizing parameters  $\{\nu_j\}_{j \geq j_0}$  is prescribed. Nevertheless, this approach offers some interesting advantages. It is easily built into the coarse-to-fine growth of the tree. One can attain higher flexibility for the smoothing effect. Moreover, overfitting artifacts are typically localized in scale which enables us to disentangle the different scales. Last, the method is much cheaper as the GCV method with both  $\alpha$  and  $\nu$  as free parameters from a computational point of view.

As illustration, we apply the multilevel GCV algorithm to the data set in Fig. 6 obtained from [38] which corresponds to a bathymetrical study of part of the sea floor of the Dominican Republic. As can be seen from the left image, measurements of the sea floor depth are irregularly

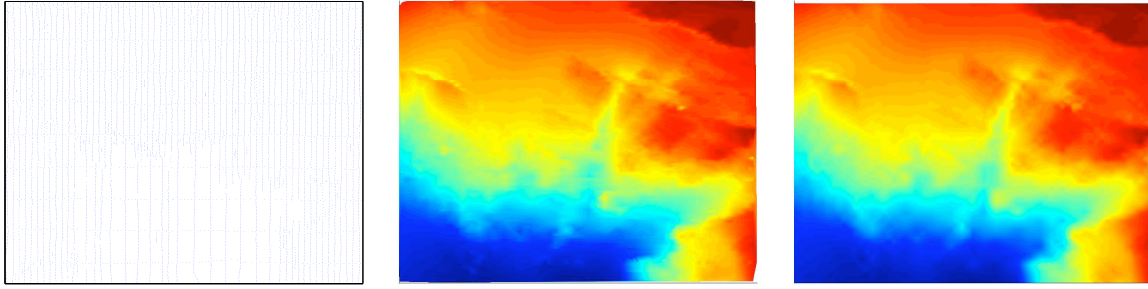


Figure 6: Left: Sampling geometry with  $N = 10732$  points. Middle: Piecewise linear interpolation. Right: Reconstruction with multilevel GCV.

distributed, forming lines, clusters and holes. The central plot shows a visualization of the depth of the full set of data points using piecewise linear interpolation. For the classical GCV, there is no obvious way to predict which  $\alpha$  may give a good reconstruction. In contrast, our multilevel GCV algorithm gives a sufficiently good reconstruction which is displayed in the right plot.

## 6 TREATMENT OF OUTLIERS

Next we discuss the issue how to treat *outliers* in the data. At first, the definition of an outlier has to be clarified. For the ‘human eye’, the presence of an outlier must create a ‘cusp’ in the approximating function; an outlier is recognized as an artifact which is extremely well localized in space *and* frequency. For this reason, the wavelet framework may be the method of choice for outlier detection. The method introduced in [5, 8] exploits the least squares approach described above by building outlier detection into it in the following way. The result of the data fitting procedure is the construction of a function  $f$  in the form (2) which represents the data. Thus, it seems natural to transfer the problem of outlier detection from the raw data point of view to the approximation  $f$ . If  $f$  catches the data features, the presence of an outlier must consequently create a local *jump* (or oscillation or peak) in  $f$ . Otherwise, if  $f$  is not affected by the outlier, the scheme does not have to take the outlier into account. Thus, we face the issues how to define these local jumps in a rigorous mathematical way which is at the same time easy to implement, and how can we distinguish jumps created by outliers and jumps contained in the data? The left plot in Fig. 7 illustrates four areas of different prototypes of outliers and various types of data which we need to be able to handle. In the leftmost area we see one point  $(x_{100}, z_{100})$  which is definitely classifiable as an outlier. The second area represents a cusp really represented in the data, and points in this area should not be marked as outliers as they represent a high frequency phenomenon in the data and accidental removal of points in this area would eliminate significant information about this local structure. The data in the third area present a spatially located feature of high frequency. Here the frequency is lower than in the second area and more data points are involved in its local representation of the local structure. A removal of some points from this area would not necessarily eliminate significant information, as the remaining points would certainly reproduce the local features of the data. The point  $(x_{800}, z_{800})$  in the fourth area is an outlier embedded into a highly energetic zone. These kind of points pose the hardest difficulties to outlier detection algorithms. On one hand, the neighborhood can mask the effect of the outlier. On the other hand, neighboring points carry significant information about the local structure on the data, and false removal in this area should be avoided by any means.

The wavelet representation of the data allows for a very practical answer to the questions

raised above. Concerning the identification of jumps, there is a vast amount of literature, see, e.g., [31]. There the basic idea is that the presence of a jump is reflected by large wavelet coefficients of  $f$ . However, jumps caused by outliers cannot be distinguished from data-inherent jumps solely by inspection of the wavelet representation of the data. In fact, one has to analyze how individual points influence the wavelet representation. In the approach described below, this task can be easily performed after having constructed the approximation since all the information needed to perform this analysis has already been processed in the data structures. Without treating the outliers, our ALGORITHM produces the approximation shown in the middle plot in Fig. 7. Clearly the outliers produce undesired artifacts in the first and the fourth area.

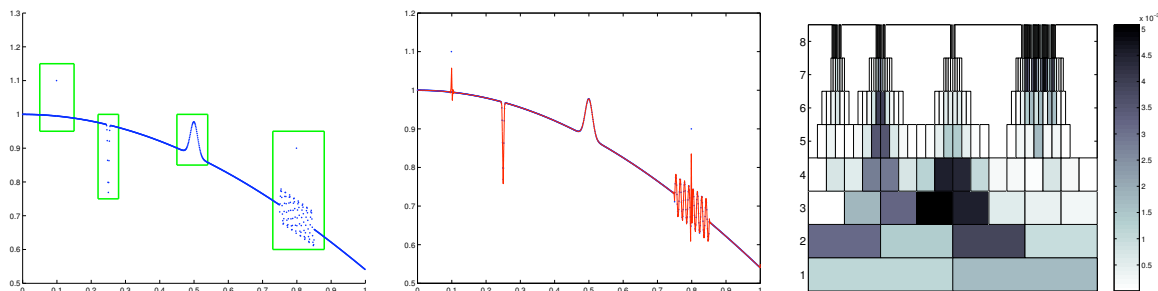


Figure 7: Left: Synthetic data set with  $N = 1000$  data points showing different types of outliers. Middle: Adaptive reconstruction of the data set with  $\#\Lambda = 190$  wavelets. Right: Resulting wavelet coefficients of the approximation. The  $y$ -axis shows the resolution levels and the  $x$ -axis displays the spatial location of the wavelet coefficients. Larger wavelet coefficients are darker. Highest resolution level in  $\Lambda$  is  $J = 8$ .

After having carried out ALGORITHM, we know the coefficients  $\{d_\lambda\}_{\lambda \in \Lambda}$  of the constructed approximation  $f$  of the form (2) as the main product of the algorithm and an efficient encoding of the observation matrix  $A = A_\Lambda$ . The main computational effort lies in the construction of  $A$ , which implicitly contains a complete analysis of the data. We will exploit this in the outlier-finding approach by separating the outlier finding procedure into three steps. First, we *locate* in which area an outlier can be present by simple inspection of the coefficients  $\{d_\lambda\}_{\lambda \in \Lambda}$ , as the presence of outliers must cause large wavelet coefficients at high levels, see the right plot in Fig. 7. We then *extract* out of the original  $N$  points those contained in the support of all wavelets  $\psi_\lambda$  classified in the first step. Third, we test all the points  $(x_i, z_i)$  in the areas identified in the first two steps, accumulated in the subset  $S \subset P$ , based on a *merit criterion*  $\omega(i)$  which serves as an ‘outlierness indicator’. Points whose  $\omega(i)$  is above a predefined threshold  $\tau$  are discarded from the data set. Next we need to describe how to construct meaningful and computationally fast outlierness profiles  $\omega(i)$ .

The first criterion is a *global refitting criterion* defined as follows. Consider one of the points  $(x_i, z_i)$  for  $i \in S$  and test if it is an outlier. The basis of this test is to measure to which extent its presence alters the shape of the approximation  $f$  which can be achieved in three steps. We first construct an approximation  $f$  to the whole data set. Then we construct an approximation to the whole set of data *except* the point  $(x_i, z_i)$  to be checked. In order to do so, we use the same configuration of wavelets which has been used to obtain the approximation to the whole

data set. Thus, we compute

$$f^{[i]} := \arg \min_{g = \sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda} \sum_{\ell \neq i} (z_\ell - g(x_\ell))^2. \quad (19)$$

Then we compare the behavior of  $f$  and  $f^{[i]}$  in the ‘neighborhood’ of  $x_i$ . The last step can be done in a natural way using the wavelet coefficients. We define the *influence set* of  $(x_i, z_i)$  starting from level  $j$  in the index set  $\Lambda$  as the subset of those  $\lambda$  which includes indices of all wavelets whose support contain  $x_i$ . We denote this neighborhood as

$$\Lambda_j^{[i]} := \{\lambda \in \Lambda, |\lambda| \geq j : x_i \in \text{supp } \psi_\lambda\}, \quad (20)$$

or shortly  $\Lambda^{[i]}$ . Then we compare the local behavior of  $f$  and  $f^{[i]}$ . The *local energy* of a function  $f$  by means of a weighted summation of a subset of its wavelet coefficients is defined as follows: for a function  $g = \sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda$  on  $\Omega$  and a set  $\Lambda' \subset \Lambda$ , let

$$E_\alpha^{\Lambda'}(g) := \sum_j 2^{2j\alpha} \sum_{\mathbf{k}, \mathbf{e} \in \Lambda'} |d_{j, \mathbf{k}, \mathbf{e}}|^2. \quad (21)$$

Apparently this definition relies on the norm equivalence (13), compare [28] for the more general case of norm equivalences for Besov seminorms. If  $(x_i, z_i)$  is an outlier, in the neighborhood of  $x_i$  the local energy of  $f^{[i]}$  should be much smaller than the local energy of  $f$ . We define the *merit profile* of  $(x_i, z_i)$  according to a *global criterion* as

$$\omega_{\text{global}}(i) := \log \left( \frac{E_\alpha^{\Lambda^{[i]}}(f)}{E_\alpha^{\Lambda^{[i]}}(f^{[i]})} \right). \quad (22)$$

In our experiments we found that a typical thresholding value  $\tau$  should be in the order of magnitude of 1. Points  $(x_i, z_i)$  for which  $\omega_{\text{global}}(i) \geq \tau$  are then classified as an outlier. Thus, in the proposed model we expect the presence of an outlier to cause a noticeable increase of the local energy.

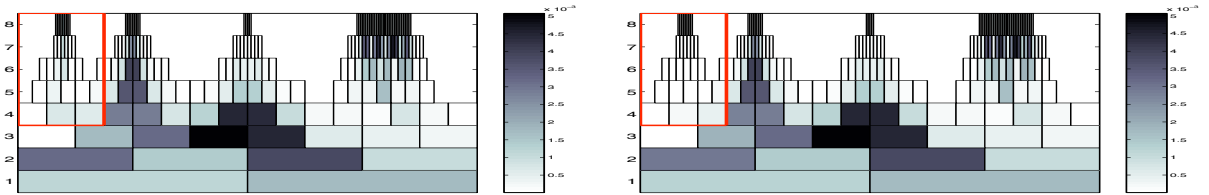


Figure 8: Global removal criterion applied to  $(x_{100}, z_{100})$ . Wavelet coefficients of  $f$  (left) and of  $f^{[100]}$  (right).

Revisiting the synthetic data set from Fig. 7, we explore next how this method works for the different areas represented in the data. If we take the outlier  $(x_{100}, z_{100})$  and compute the global approximations  $f$  and  $f^{[100]}$ , we obtain the wavelet coefficients displayed in Fig. 8. As expected, no difference is visible outside the indicated box in the upper left corner of the wavelet coefficients starting at level 4. The presence of the outlier really does act locally. This is shown in detail in Fig. 9. In the left plot we see  $f$  (dashed line) and  $f^{[100]}$  (solid line) in the neighborhood of  $(x_{100}, z_{100})$ , where  $f$  shows an oscillation created by the point  $(x_{100}, z_{100})$ .

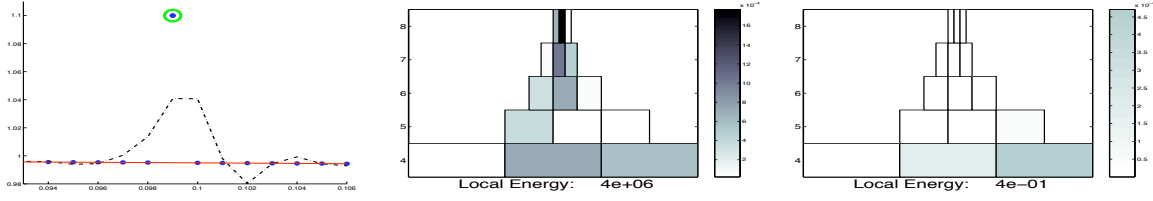


Figure 9: Global removal criterion applied to  $(x_{100}, z_{100})$ . Left: Local view of  $f$  (dashed line) and  $f^{[100]}$  (solid line). Middle: Coefficients of wavelets in  $\Lambda^{[100]}$  for  $f$ . Right: Coefficients of wavelets in  $\Lambda^{[100]}$  for  $f^{[100]}$ .

This is reflected by the coefficients of the wavelets in  $\Lambda^{[100]}$ . After eliminating  $(x_{100}, z_{100})$  from the data to be fitted, the energy content of  $\Lambda^{[100]}$  is practically empty. The criterion (22) here yields the value  $\omega_{\text{global}}(100) = 7.02$ . The same computation for the regular neighboring point  $(x_{102}, z_{102})$  gives  $\omega_{\text{global}}(102) = 0.0078$ , as its removal does not critically vary the local energy.

Although these results are convincing, the proposed strategy has an obvious drawback since it requires the computation of  $f^{[i]}$  for every suspicious point  $(x_i, z_i)$ : we have to construct and solve a different set of normal equations for every  $i$ . Fortunately, the structure of the system  $M\mathbf{d} = \mathbf{b}$  allows to simplify the computations essentially by subtracting only a matrix of rank one as described in [8]. Moreover, the outlier detection strategy can be further sped up by considering *local energy criteria* which boils down to a *local reconstruction* of  $f$  and  $f^{[i]}$ , see the results displayed in Fig. 10. Point  $(x_{800}, z_{800})$  is an outlier in the highly energetic environment in the fourth area from Fig. 7 while the neighboring point  $(x_{802}, z_{802})$  is not. We dispense with a precise description of the local energy criteria here and refer to [8].

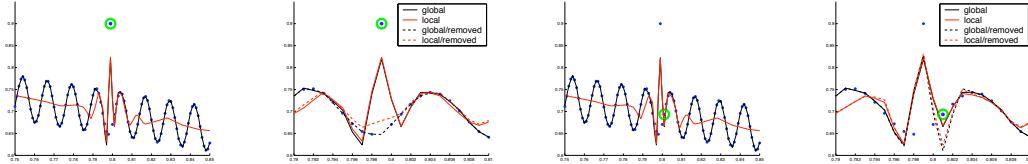


Figure 10: Left two plots: Global and local reconstructions near  $(x_{800}, z_{800})$ . Local view of  $f$  and  $f^{[800]}$  (left) and removal of  $(x_{800}, z_{800})$  in global and local approximations. Right two plots: Global and local reconstructions near  $(x_{802}, z_{802})$ . Local view of  $f$  and  $f^{[802]}$  (left) and removal of  $(x_{802}, z_{802})$  in global and local approximations.

The above methodology can naturally be extended to deal in addition with noisy data and also works in higher spatial dimensions and for a large number of outliers, see [8], where also extensive tables displaying numerical results and percentages of outlier detection rates can be found.

As a final example, we analyze a geophysical data set taken again from [38]. The set includes 18634 points ordered in a square grid which is displayed in the left of Fig. 11. We add 1000 randomly distributed outliers to this data, yielding the data on the right. In the left plot of Fig. 12 we see the performance of our algorithm after a first run which eliminates 75% of the outliers, while the data eliminated by false detection does not appear to damage the reconstruction. A second run of the algorithm, that is, another iteration on the cleaned data results in the reconstruction on the right of the figure. As we start from a situation where the density of outliers has been reduced, further outliers that were previously concealed by neighboring ones are now successfully detected.

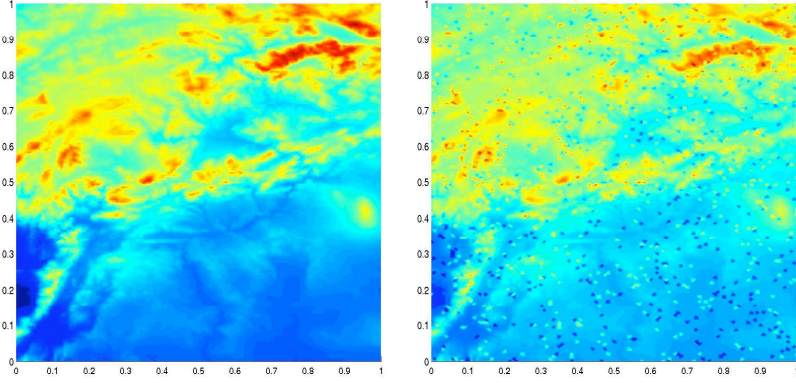


Figure 11: Geophysical data set with  $N = 18605$  data points. Left: Vertical view of original data. Right: Data set corrupted with 1000 outliers which amount to 5.6% of the data.

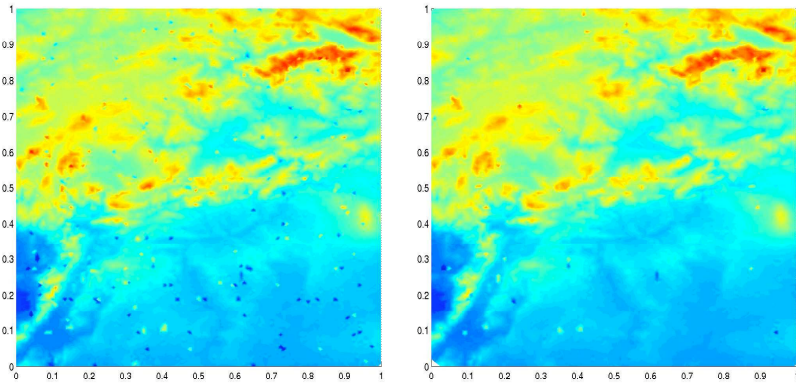


Figure 12: Reconstruction of the geophysical data from Fig. 11 after robust cleaning. Left: Results after first run. Right: Approximation after second run.

## 7 APPROXIMATE CONTINUATION OF HARMONIC FUNCTIONS

Finally, first preliminary results from [33] on the following continuation problem arising in the context of the computation of the gravity potential are addressed. Assuming that the data set stems from a *harmonic function*, that is, a function  $f : \bar{\Omega} \rightarrow \mathbb{R}$  satisfying  $\Delta f = 0$  where  $\Delta$  denotes the Laplacian, we study how well  $f$  can be reconstructed by computing the minimizer of

$$\mathcal{J}(f) := \sum_{i=1}^N (z_i - f(x_i))^2 + \nu \|\Delta f\|_{L_2(\Omega)}^2. \quad (23)$$

Here  $\mathcal{F}(f) := \|\Delta f\|_{L_2(\Omega)}^2$  in (1) is chosen. Usually such harmonic continuations are based on *interpolating* the data in terms of harmonic Ansatz functions like the spherical harmonics or wavelet-type approximations, see, e.g., [23]. These functions, however, are globally supported so that the corresponding linear equations stemming from the interpolation problem involve fully populated matrices. Consequently, their numerical solution is prohibitively expensive, in addition to the huge storage space required for the system matrix.

A first attempt to test the quality of the reconstruction uses tensor products of cubic B-splines based on a uniform grid of spacing  $2^{-j}$ ; the harmonicity is enforced by the regularization term in (23). This was originally proposed in [36] where also first promising experiments based on higher order finite elements were conducted. The normal equations derived from (23) are  $(M + \nu G)\mathbf{d} = \mathbf{b}$  where the matrix  $G$  represents the regularization term. The minimizer of (23)



will be denoted by  $f_\nu$ .

As an example, we consider

$$f(x, y) = \frac{1}{2} (e^{-10x} \sin(10y) + e^{-10(1-x)} \sin(10(1-y)) + 1) \quad (24)$$

on  $\Omega = (0, 1)^2$  from which a set of data  $P$  is extracted by sampling  $f$  on the boundary with respect to a sampling density  $2^{-\ell}$  for  $\ell \in \{6, 7, 8\}$  displayed in Fig. 13.

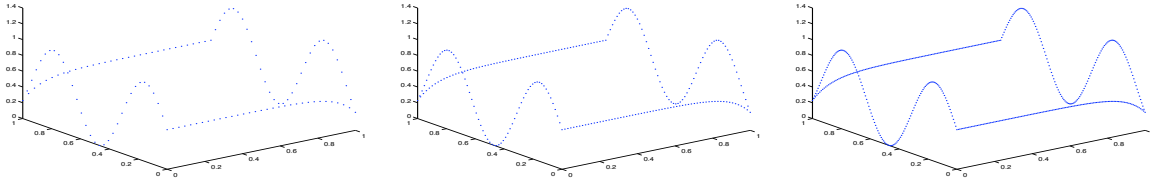


Figure 13: Geometry of points for resolution  $2^{-\ell}$ ,  $\ell = 5, 6, 7$ .

For tensor products of cubic B-splines on resolution  $2^{-4}$ , the results for the reconstruction for data resolved at  $2^{-5}$  are shown in Fig. 14. Naturally the reconstruction error is larger at the boundaries where the amplitude of  $f$  is also larger than in the interior.

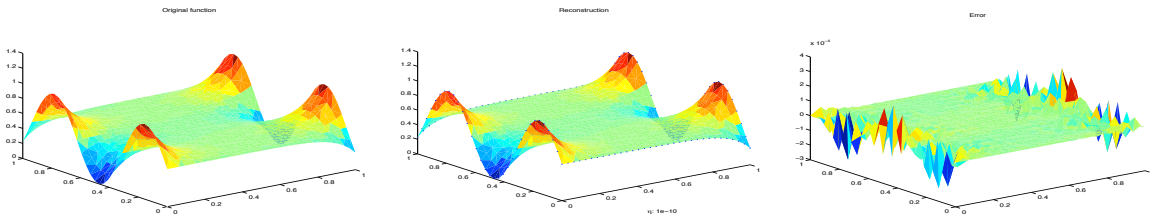


Figure 14: Left: Original function  $f$ . Middle: Reconstruction  $f_\nu$  for  $\nu = 1.0\text{e-}10$  and data points. Right: Reconstruction error  $f - f_\nu$ .

Many numerical studies and computations with satellite data from actual satellite missions are provided in [33], together with extensions on using multiscale bases and adaptivity along the lines in this article.

## ACKNOWLEDGEMENT

This work was supported by the Deutsche Forschungsgemeinschaft, Grant KU 1028/7-1, and by the SFB-TR32, Universität Bonn.

## REFERENCES

- [1] A.R. Barron, A. Cohen, W. Dahmen and R. DeVore, Approximation and learning by greedy algorithms, *Ann. Statist.* **36**(1), 2008, 64–94.
- [2] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

- [3] M. Buhmann, Radial basis functions: the state-of-the-art and new results, *Acta Numerica* **9** (2000), 1–37.
- [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: *Computer Graphics, Proc. SIGGRAPH 2001* (L. Poccock et al., Eds.), ACM Press, New York, 2001, 67–76.
- [5] D. Castaño, *Adaptive Scattered Data Fitting with Tensor Product Spline–Wavelets*, PhD Dissertation, Universität Bonn, 2005.  
[http://hss.ulb.uni-bonn.de/diss\\_online/math\\_nat\\_fak/2005/castano-diez\\_daniel](http://hss.ulb.uni-bonn.de/diss_online/math_nat_fak/2005/castano-diez_daniel)
- [6] D. Castaño and A. Kunoth, Adaptive fitting of scattered data by spline-wavelets, in: *Curves and Surfaces* (A. Cohen et al.), Nashboro Press, Brentwood, 2003, 65–78.
- [7] D. Castaño and A. Kunoth, Multilevel regularization of wavelet based fitting of scattered data – Some experiments, *Numer. Algorithms* **39** (1-3) (2005), 81–96.
- [8] D. Castaño and A. Kunoth, Robust regression of scattered data with adaptive spline–wavelets, *IEEE Trans. Image Proc.*, *IEEE Trans. Image Proc.* **15** (6), 2006, 1621–1632.
- [9] D. Castaño, M.D. Gunzburger and A. Kunoth, An adaptive wavelet viscosity method for hyperbolic conservation laws, *Numer. Meths. for PDEs* **24** (6), 2008, 1388–1404.
- [10] A. Chambolle, R.A. DeVore, N.-Y. Lee and B.J. and Lucier, Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Trans. Image Proc.* **7**(3) (1998), 319–335.
- [11] C.K. Chui, *An Introduction to Wavelets*, Vol. 1, Academic Press, Boston, 1992.
- [12] A. Cohen, W. Dahmen and R. DeVore, Adaptive wavelet schemes for nonlinear variational problems, *SIAM J. Numer. Anal.* **41** (2003), 1785–1823.
- [13] W. Dahmen, Wavelet and multiscale methods for operator equations, *Acta Numerica* **6** (1997), 55–228.
- [14] W. Dahmen and A. Kunoth, Multilevel preconditioning, *Numer. Math.* **63** (1992), 315–344.
- [15] W. Dahmen and A. Kunoth, Adaptive wavelet methods for linear–quadratic elliptic control problems: Convergence rates, *SIAM J. Contr. Optim.* **43** (2005), 1640–1675.
- [16] W. Dahmen, A. Kunoth and K. Urban, Biorthogonal spline-wavelets on the interval – Stability and moment conditions, *Applied Comp. Harmonic Analysis* **6** (1999), 132–196.
- [17] R.A. DeVore, Nonlinear approximation, *Acta Numerica* **7** (1998), 51–150.
- [18] R.A. DeVore and A. Kunoth, *Multiscale, Nonlinear and Adaptive Approximation*, Springer, to appear September 2009.
- [19] P. Dierckx, *Curve and Surface Fitting with Splines*, Oxford University Press, Oxford, 1993.

- [20] N. Dyn, M.S. Floater and A. Iske, Adaptive thinning for bivariate scattered data, *J. Comput. Appl. Math.* **145** (2002), 505–517.
- [21] M.S. Floater and A. Iske, Thinning, inserting and swapping scattered data, in: *Surface Fitting and Multiresolution Methods* (A. Le Méhauté et al.), Vanderbilt University Press, Nashville, 1996, 139–144.
- [22] Forsey, D. R. and Bartels, R. H., Hierarchical B–spline refinement, *Computer Graphics* **22** (1988), 205–212.
- [23] W. Freeden and V. Michel, *Multiscale Potential Theory with Applications to Geoscience*, Birkhäuser, Basel, 2004.
- [24] Th. Gerstner, H.-P. Helfrich and A. Kunoth, Wavelet analysis of geoscientific data, in: *Dynamics of Multiscale Earth Systems* (H. J. Neugebauer and C. Simmer, Eds.), Lecture Notes in Earth Sciences, Springer, New York, 2003, 69–88.
- [25] S.J. Gortler and M. Cohen, Hierarchical and variational geometric modeling with wavelets, in: *Proc. Symposium on Interactive 3D Graphics* (M. Zyda et al., Eds.), ACM Press, New York, 1995, 35–42.
- [26] B.F. Gregorski, B. Hamann and K.I. Joy, Reconstruction of B-spline surfaces from scattered data points, in: *Proceedings of Computer Graphics International 2000* (N. Magnenat-Thalmann et al.), IEEE Computer Society, Washington DC, 163–170.
- [27] G. Greiner and K. Hormann, Interpolating and approximating scattered 3D-data with hierarchical tensor product splines, in: *Surface Fitting and Multiresolution Methods* (A. Le Méhauté, et al.), Vanderbilt University Press, Nashville, 1996, 163–172.
- [28] W. Härdle, G. Kerkycharian, D. Picard and A. Tsybakov, *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statistics, Springer, New York, 1998.
- [29] M. Hegland, S. Roberts and I. Altas, Finite element thin plate splines for surface fitting, in: *Computational Techniques and Applications*, Proc. CTAC97 (B. J. Noye, M. D. Teubner and A. W. Gill, Eds.), World Scientific, Singapore, 1997, 289–296.
- [30] K. Höllig and U. Reif, Nonuniform web-splines, *Computer Aided Geometric Design* **20** (2003), 277–294.
- [31] P.J. Huber, *Robust Statistics*, John Wiley & Sons, New York, 1981.
- [32] A. Iske and J. Levesley, Multilevel scattered data approximation by adaptive domain decomposition, *Numer. Algorithms* **39** (1-3) (2005), 187-198.
- [33] G. Jager, *Approximate Continuation of Harmonic Functions: A Weighted Least Squares Approach Based on Multiscale Bases, Adaptive Extensions, and Applications to Geodesy*, PhD Dissertation, Universität Paderborn, in preparation.
- [34] Kunoth, A., Adaptive wavelet schemes for an elliptic control problem with Dirichlet boundary control, *Numer. Algorithms* **39** (1-3) (2005), 199-220.

- [35] S. Lee, G. Wolberg, S.Y. Shin, Scattered data interpolation with multilevel B-splines, *IEEE Trans. Visualization and Computer Graphics* **3** (3) (1997), 228–244.
- [36] P. Meissl, The use of finite elements in physical geodesy, Report No. 313, Dept. of Geodetic Science, Ohio State University, 1981.
- [37] V. Pereyra and G. Scherer, Large scale least squares scattered data fitting, *Appl. Numer. Math.* **44** (1–2) (2002), 73–86.
- [38] The Puerto Rico Tsunami Warning and Mitigation Program. Data obtainable at <http://poseidon.uprm.edu>
- [39] M.J.D. Powell, Radial basis functions for multivariate interpolation: a review, in: *Algorithms for Approximation* (J. C. Mason et al.), Clarendon Press, Oxford, 1987, 143–167.
- [40] R. Schaback, Multivariate interpolation by polynomials and radial basis functions, *Constr. Approximation* **21** (2005), 293–317.
- [41] R. Schaback and H. Wendland, Adaptive greedy techniques for approximate solution of large rbf systems, *Numer. Algorithms* **24** (2000), 239–254.
- [42] V. Scheib, J. Haber, M.C.Lin and H.P. Seidel, Efficient fitting and rendering of large scattered data sets using subdivision surfaces, In: *Computer Graphics Forum*, Proc. Eurographics 2002, Eurographics Publ., 2002, 353–362.
- [43] L.L. Schumaker, Fitting surfaces to scattered data, in: *Approximation Theory II* (G. G. Lorentz et al.), Academic Press, New York, 1976, 203–268.
- [44] M.A. Schweitzer, *A Parallel Multilevel Partition of Unity Method for Elliptic Partial Differential Equations*, Lecture Notes in Computational Science and Engineering, Vol. 29, Springer, New York, 2003.
- [45] E.J. Stollnitz, T.D. DeRose, and D.H. Salesin, *Wavelets for Computer Graphics*, Morgan Kaufmann Publishers, San Mateo, 2000.
- [46] G. Wahba, *Spline Models for Observational Data*, CBMS-NSF Reg. Conf. Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, 1990.
- [47] H. Wendland, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, Vol. 17, Cambridge University Press, Cambridge, 2005.
- [48] Yserentant, H., On the multilevel splitting of finite element spaces, *Numer. Math.* **49** (1986), 379–412.