Robust Regression of Scattered Data with Adaptive Spline–Wavelets

Daniel Castaño and Angela Kunoth

***Abstract*— A coarse-to-fine data fitting algorithm for irregularly spaced data based on boundary–adapted adaptive tensor–product semi–orthogonal spline–wavelets has been proposed in [CK1]. This method has been extended in [CK2] to include regularization in terms of Sobolev and Besov norms. In this paper, we develop within this least–squares approach some statistical robust estimators to handle outliers in the data. Our wavelet scheme yields a numerically fast and reliable way to detect outliers.**

***Index Terms*— Scattered data, outlier detection, robust estimation, adaptive wavelets, coarse–to–fine algorithm.**

I. Introduction

In [CK1], we have proposed an adaptive method of least squares data fitting based on wavelets which progresses on a coarse–to–fine basis. The algorithm works on structured grids and is, therefore, particularly useful to implement using tensor products in more than one spatial dimensions. A favourable feature of the scheme is that there is no explicit grid construction like when using, e.g., data–dependent triangulations. In this sense, the data representation is independent of the original cloud of data points. These advantages have made approaches on structured grids a well established, much followed and universally applicable procedure, covering the whole range from strictly mathematical analysis to applications in various areas of sciences, see, e.g., [GG], [GHJ], [GH], [He], [HPMM], [HR], [LWS], [PS], [SHLS], [Sch], [Z]. Wavelets as basis functions for the representation of scattered data provide additional features in the problem formulation regarding computational efficiency as well as sparseness of the representation, such as good conditioning and a natural built–in potential for adaptivity (see, e.g., [Ch] for an introduction to basic wavelet theory).

We briefly wish to recall for further reference our approach in [CK1], [CK2] and some properties of the wavelets we employ here. Suppose we are given some set $X = \{x_i\}_{i=1,\dots,N}$, consisting of irregularly spaced and pairwise disjoint points $x_i \in \Omega := [0,1]^n$, $n \in \{1,2,3\}$. For each i, we denote by $z_i \in I\!R$ the corresponding data making up the set Z. The problem of *scattered data fitting* with regularization (of Tikhonov type) can be formulated as follows: Find a function $f : \Omega \to I\!R$ that approximates the cloud of points (X, Z) in a least squares sense, that is, f is to minimize the functional

$$J(f) := \sum_{i=1}^{N} \left(z_i - f(x_i)\right)^2 + \nu \|f\|_Y^2. \tag{I.1}$$

Here Y may be a Sobolev space $Y = H^\alpha$ or a Besov space with smoothness index α, and the parameter $\nu \geq 0$ balances the data fit with a user–specified regularity of the representation f. In particular, we want to construct an expansion of f of the form

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

The set of basis functions $\{\psi_\lambda\}_{\lambda\in\Lambda}$ is a subset of the wavelets described in [SDS]. They have the following properties: each ψ_λ is a tensor product of a certain boundary–adapted linear combination of linear B–splines, denoted as (pre)wavelets or shortly *wavelets*. From a computational point of view, this is very advantageous since we can practically work with piecewise polynomials. The index set Λ is a lacunary set of indices resulting from an adaptive coarse–to–fine procedure explained below. We denote the infinite set of all possible indices by $I\!I$. Each index $\lambda \in I\!I$ is of the form $\lambda = (j, \mathbf{k}, \mathbf{e})$, where $j =: |\lambda|$ denotes the *level of resolution* or *refinement scale*, $\mathbf{k}$ the spatial *location*, and $\mathbf{e} \in \{0,1\}^n$ the type of wavelet in more than one spatial dimension which is induced by tensor products, see e.g. [D1], [DKU], [SDS]. Each wavelet ψ_λ has compact support satisfying $\operatorname{diam}(\operatorname{supp}\psi_\lambda) \sim 2^{-|\lambda|}$. The relation $a \sim b$ always is to mean that a can be estimated from above and below by a constant multiple of b independent of all parameters on which a or b may depend. In view of the finite domain and the compact support of the basis functions, there is by construction a coarsest refinement level j_0 ($j_0 = 1$ in the case of piecewise linear wavelets considered here). Basis elements with multi–index $\mathbf{e} = (0,0)$ only occur on level j_0. They are called *scaling functions* and are tensor products of piecewise linear B–Splines. For $\mathbf{e} \neq (0,0)$, ψ_λ represents detail information of higher frequencies.

Consequently, $f(x)$ in (I.2) can be split into a *scaling function term* and a *wavelet term*,

$$f(x) = \sum_{\lambda\in\Lambda,\ j=j_0,\ \mathbf{e}=(0,0)} d_\lambda \, \psi_\lambda(x) + \sum_{\lambda\in\Lambda,\ j\geq j_0,\ \mathbf{e}\neq(0,0)} d_\lambda \, \psi_\lambda(x). \tag{I.3}$$

The complete collection of scaling functions and wavelets $\{\psi_\lambda : \lambda \in I\!I\}$ constitutes a *Riesz basis* for $L_2(\Omega)$. Moreover, one has *norm equivalences* for functions in Sobolev spaces

Manuscript received March 11, 2005. This work was supported by the Deutsche Forschungsgemeinschaft, Grant KU 1028/7–1, and by the SFB 611, Universität Bonn. Daniel Castaño was also supported by a BFI grant of the Basque Government.

European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany, castano@embl.de

Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany, kunoth@ins.uni-bonn.de.

$H^\alpha = H^\alpha(\Omega)$ (or, even more general, in Besov spaces [DV]) in the range $\alpha \in [0, 3/2)$ of the form

$$\| \sum_{\lambda=(j,\mathbf{k},\mathbf{e})\in I\!I} d_\lambda \, \psi_\lambda \|^2_{H^\alpha(\Omega)} \;\sim\; \sum_{j\geq j_0} 2^{2\alpha j} \sum_{\mathbf{k},\mathbf{e}} |d_{j,\mathbf{k},\mathbf{e}}|^2. \quad \text{(I.4)}$$

Since such smoothness spaces can be characterized in this way, together with their compact support wavelets suggest themselves as a powerful analysis tool for many purposes, see, e.g., [D1], [CDLL]. In addition, the wavelets we employ here are *semi–orthogonal* with respect to $L_2(\Omega)$, i.e., for $|\lambda| \neq |\mu|$ we always have $\int_\Omega \psi_\lambda(x)\, \psi_\mu(x)\, dx = 0$.

Returning to the least squares fitting problem (I.1), adaptation to the given data is achieved in [CK1], [CK2] through the construction of an index set $\Lambda \subset I\!I$ as follows. Starting on the coarsest level j_0, we choose here the set Λ_{j_0} of indices of all scaling functions and wavelets on this level. An initial fitting function $f^{j_0}(x) := \sum_{\lambda\in\Lambda_{j_0}} d_\lambda \, \psi_\lambda(x)$ is constructed on this index set by minimizing $J(f^{j_0})$ or, equivalently, solving the normal equations

$$(A^T_{\Lambda_{j_0}} A_{\Lambda_{j_0}} + \nu R_{j_0}) d = A^T_{\Lambda_{j_0}} z \quad \text{(I.5)}$$

iteratively using the conjugate gradient method. The matrix R_{j_0} is in the case of $Y = H^\alpha$ a diagonal matrix of the form $R_{j_0} = \mathrm{diag}(2^{2\alpha\, j_0})$. The *observation matrix* $A_{\Lambda_{j_0}}$ has entries

$$(A_{\Lambda_{j_0}})_{i,\lambda} := \psi_\lambda(x_i),\ i = 1, \ldots, N,\ \lambda \in \Lambda_{j_0}, \quad \text{(I.6)}$$

and z and d in (I.5) are vectors comprising the right hand side data $\{z_i\}_{i=1,\ldots,N}$ and the expansion coefficients $\{d_\lambda\}_{\lambda\in\Lambda_{j_0}}$.

The norm equivalence (I.4) for $\alpha = 0$ and the locality of the ψ_λ reveals that for $\mathbf{e} \neq (0,0)$, the absolute value of a coefficient d_λ is a measure of the spatial variability of f^{j_0} on $\mathrm{supp}\, \psi_\lambda$: a large value of d_λ is understood as an indicator that further resolution in this area of the domain is desired. On the other hand, in order to keep control over the computational complexity, if d_λ is below a certain user–defined threshold, this coefficient is considered irrelevant and will be discarded from the approximation in order not to spoil the complexity. The index set is modified accordingly. This motivates us to construct in the next step a refined index set Λ_{j_0+1} by including those children of the wavelets indexed by Λ_{j_0} whose coefficients are above some prescribed thresholding value *and* in whose support there are more than a fixed number q of data points. Note that this strategy generates an approximation on a *tree* as index structure. The procedure is repeated until at some dyadic highest resolution level J either all the computed coefficients are smaller than the thresholding value, or none of the children whose supports shrinks with each refinement step contains enough data points in their support. Then the algorithm stops growing the tree. As the data set is finite, the algorithm finishes in finitely many steps. The final index set is then a lacunary tree index set $\Lambda = \Lambda_J$. We wish to point out that the highest resolution level J is solely determined by the data to be fitted. As to the choice of q, one can determine conditions on the number of points and their distribution which guarantee by the Theorem of Schoenberg-Whitney that the observation matrix A_{Λ_j} has full rank [Ca]. A smaller number q (larger than the minimal) may generate a tree of greater depth while a larger number does the opposite. (We found that the best compromise between generating sufficient hierarchy in the tree, quality of the approximation and performance of the algorithm is achieved for $q = 8$ in 1D and $q = 100$ in 2D which we have used in all the subsequent numerical examples.)

Note that we have in the least squares functional (I.1) in fact two parameters, the smoothness index α and the weight parameter ν balancing the approximation and the smoothness part, to adjust as best as possible to the given data. In order to compute the weight parameter ν, often a generalized cross validation technique is employed. This requires solving an additional system of the type of the normal equations. In [Ca], [CK2], we have further exploited the multilevel framework provided by wavelets: we have introduced a multilevel 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 nicely to smoothness requirements as well as to localization effects.

It should be mentioned that we always solve the least squares problem using normal equations as in (I.5). We have observed in our numerical experiments that the condition numbers of $A^T_\Lambda A_\Lambda$ are relatively moderate, which is inherited from the well–conditioning of the wavelet basis relative to the L_2 inner product. Together with employing a nested iteration strategy in the coarse–to–fine algorithm, taking the solution from the previous level as initial guess for the refined index sets, we have documented in the experiments in [CK1] that iteration numbers for a conjugate gradient method on the normal equations are very moderate. Also we have found that the approximation error, comparing the reconstruction with the exact value of $J(f)$ defined in (I.1), is acceptable [CK2]. We have, in addition, compared approximation errors of our results with the normal equations with approximation errors for a least squares solution computed using QR decomposition of A_Λ and we found only negligible differences, while the iterations took much less time for the normal equations. As a third point worth mentioning, under these conditions, for $\#\Lambda \ll N$ which is mostly the case here, 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 $(A^T_\Lambda A_\Lambda)_{\lambda,\mu} = \sum_{i=1}^N \psi_\lambda(x_i)\psi_\mu(x_i)$ and not in the size of the matrix of the normal equations.

Multiscale data fitting schemes which may or may not include a smoothing term have been discussed in many other references, see [CK2] for further details. Previously to the studies presented here, we have already gained some experience with wavelet analysis of geoscientific data on *uniform grids* using the Fast Wavelet Transform (FWT) [GHK]. We wish to stress that for the scattered data fitting procedure considered in this paper, employing the FWT on uniform grids would require the introduction of some artificial uniform grid which would spoil the computational complexity. All the above mentioned aspects are described in detail in the dissertation [Ca] where also the employed software has been documented.

After describing the essential features of our algorithm, this paper is devoted now to statistical robust estimators to handle *outliers* in the data. Again we will see that we can take advantage of the multilevel structure of the wavelet setup.

The remainder of this paper is structured as follows. In Section II, we propose an LSW–specific (Least Squares–Wavelet) methodology for robust estimation of outliers. Its basics are described in Section III. Sections IV through VI discuss global and local outlier detection criteria. Section VII is devoted to the issue how to detect a large number of outliers in the data. Then Section VIII discusses other forms of energy criteria to measure the performance of our algorithm. Finally, we describe in Section IX the extension of our method to higher spatial dimensions and conclude with its performance on geophysical data sets.

II. An LSW–Specific Methodology for Robust Data Fitting

We will develop an LSW (Least Squares–Wavelet) scheme to detect outliers which exploits the least squares approach described above. First, we have to discuss how an outlier can be defined. We try to mimic the process by which the 'human eye' tells us what an outlier is: the presence of an outlier must create a 'cusp' in the approximating function. Thus, an outlier is recognized as an artifact which is extremely well localized in space *and* frequency. This is one of the reasons why we think that the wavelet framework is the method of choice to detect outliers. Other frameworks separate the removal of outliers and the wavelet analysis of the data. In contrast, here we rather want to build the outlier detection into the wavelet representation and take advantage of it. As described above, the result of the data fitting procedure is the construction of a function f in the form (I.2) which represents the data. Thus, it seems natural to transfer the problem of the outlier definition from the point of view of the raw data to the point of view of approximating f. If the approximating function correctly catches the data features, the presence of an outlier must consequently create a local jump in the approximating function. Otherwise, if f is not affected by the outlier, the scheme does not have to take the outlier into account.

This leads to reformulate the problem as follows: how can one define these jumps in a rigorous mathematical way which is at the same time easy to implement, and how can one distinguish jumps created by outliers and jumps contained in the data ? Figure 1 illustrates four areas of different prototypes of outliers and various types of data which we need to be able to handle. In the first area from the left we find a point which is definitely classifiable as an outlier. This helps us to fix the criterion that a well–performing method should mark the corresponding point (x_{100}, z_{100}) as an outlier. The second area represents a cusp really represented in the data. The points in this area are not to be marked as outliers. Moreover, they are the representatives of a high frequency phenomenon of the data. The accidental removal of points in this area would eliminate significant information about this local structure. Like in Area 1, the data in the third area presents a spatially located high frequency feature. In this case the frequency is lower than in Area 2, so that more points from the data set are involved in the representation of the local structure. This could represent a noncritical area of the domain. No outlier is present. A removal of some points on this area does 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 Area 4 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.

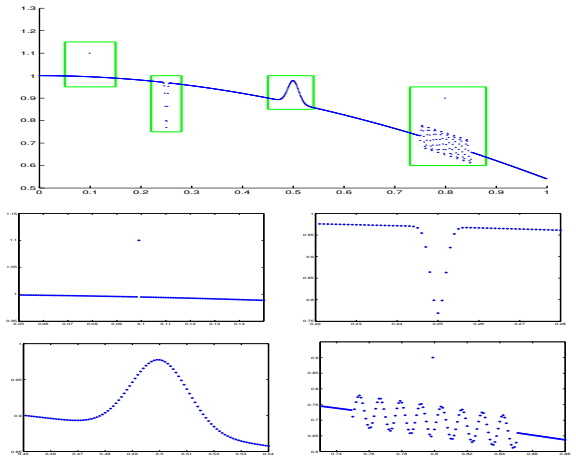

Fig. 1. Synthetic Data Set 1 with $N = 1000$ data points showing different types of outliers (left). Local details of the data (right).

The wavelet representation of the data allows for an extremely practical answer to both questions raised above:

- Concerning the identification of *jumps*, there is a vast amount of literature, see, e.g., [Co], [Hu], [LW].The basic idea is that the presence of a jump is reflected by large wavelet coefficients of the approximation.
- Outlier–caused jumps cannot be distinguished from data–inherent jumps solely by inspection of the wavelet representation of the data. In fact, one has to inspect the point themselves and analyze how individual points influence the wavelet representation. We find below that this task can be easily performed after having constructed the LSW–approximation to the data described below, as all the information needed to perform this analysis has already been processed in the data structures.

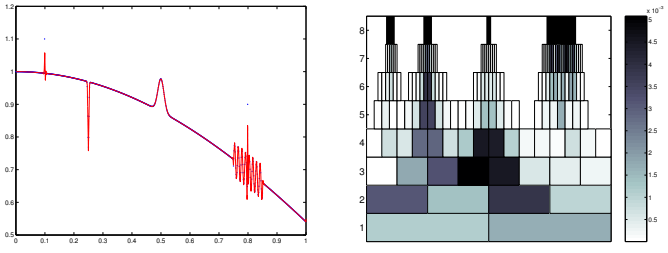

Fig. 2. Adaptive reconstruction of synthetic Data Set 1 with $\#\Lambda = 190$ wavelets (left) and resulting wavelet coefficients (right: on the y–axis the resolution levels are shown, the x–axis displays the spatial location of the wavelet coefficients. Larger wavelet coefficients are darker in color). Highest resolution level in Λ is $J = 8$.

III. Basics of the LSW–Approach

Without treating the outliers in any specific way, our algorithm produces the results shown in Figure 2. Clearly the outliers produce undesired artifacts in Areas 1 and 4.

After performing the LSW–data fitting algorithm, we end up with two objects:

1) the coefficients $\{d_\lambda\}_{\lambda\in\Lambda}$ of the constructed approximation f (I.2) as the main product of the algorithm and
2) an efficient encoding of the observation matrix $A = A_\Lambda$, which was used and updated in the process.

The main computational effort lies in the construction of A, which implicitly contains a *complete analysis* of the data. This will be exploited in our approach which we propose next. We separate the outlier finding procedure into three steps:

I. *Locate the areas* in which an outlier can be present. As explained previously, this can be performed by simple inspection of the coefficients $\{d_\lambda\}_{\lambda\in\Lambda}$, as the presence of outliers must cause large wavelet coefficients at high levels. One evident possibility could be to choose a thresholding parameter ε and a smallest refinement level $j_{\rm cut}$ which models the expected penetrating depth of the outliers, and to identify a set of wavelets indices

$$\Lambda_{\varepsilon,j_{\rm cut}} := \{\lambda\in\Lambda:\ |d_\lambda|\geq\varepsilon \text{ and } |\lambda|\geq j_{\rm cut}\}. \quad \text{(III.1)}$$

II. Next, we could extract out of the original N points those embedded in the support of a wavelet ψ_Λ such that $\lambda\in\Lambda_{\varepsilon,j_{\rm cut}}$, and record their indices in a set S:

$$i\in S :\Longleftrightarrow\ \lambda\in\Lambda_{\epsilon,j_{\rm cut}} \text{ so that } x_i\in\operatorname{supp}\psi_\lambda. \quad \text{(III.2)}$$

This step serves only to reduce the number of points in outlier–affected areas.

III. Test *all* the points in the areas identified in the first two steps. Basically, a *merit criterion* $\omega(i)$ is computed for each point included in S, which serves as an 'outlierness indicator'. Points whose ω is above a predefined threshold τ are discarded from the data set.
This is the decisive part of the algorithm, as the concrete mathematical translation of the outlier concept has to be built into the computation of the merit criterion.

We have actually set $\varepsilon = 0$ in order for the pointwise criterion in III. to be most effective. Remarks on the choice of $j_{\rm cut}$ are given at the end of Section V. In the next sections we describe and analyze several ways to construct meaningful and computationally fast outlierness profiles $\omega(i)$ for $i\in S$.

IV. Global Refitting Criterion

We start with a set S of indices of points which have been identified as possible outliers by the first two steps of the previous section, as they lie in the support of wavelets with large coefficients. Consider one of these points, (x_i, z_i) for $i\in S$ and test if it is an outlier. As mentioned above, in our approach, the basis of this test would be to measure to which extent its presence alters the shape of the approximation. This goal can be attained in three steps:

1) Construct an approximation to the whole data set.
2) 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 compute the approximation to the whole data set. Thus, we compute

$$f^{[i]} := \arg\min_{g=\sum_{\lambda\in\Lambda} d_\lambda\psi_\lambda} \sum_{\ell\neq i} \left(z_\ell - g(x_\ell)\right)^2. \quad \text{(IV.1)}$$

3) Compare the behavior of f and $f^{[i]}$ in the 'neighborhood' of x_i.

Elaborating this last point, this can be done in a natural way using the wavelet coefficients. First of all, we need an interpretation of the concept of *neighborhood*. To this end, we define the *influence set* of a point (x_i, z_i) starting from level j in the index set Λ as the subset of those Λ which includes indices of all wavelets whose support contain x_i. We denote this neighborhood as

$$\Lambda_j^{[i]} := \{\lambda\in\Lambda,\ |\lambda|\geq j:\ x_i\in\operatorname{supp}\psi_\lambda\}, \quad \text{(IV.2)}$$

or shortly $\Lambda^{[i]}$ when the subscript j is clear or irrelevant. Now, we compare the local behavior of f and $f^{[i]}$. We define the *local energy* of a function by means of a weighted summation of a subset of its wavelet coefficients as follows: for a function $g=\sum_{\lambda\in\Lambda} d_\lambda\psi_\lambda$ on Ω and a set $\Lambda'\subset\Lambda$, we define

$$E_{\alpha,p,r}^{\Lambda'}(g) := \sum_j \left(2^{j(\alpha+n/2-n/p)} \left(\sum_{\mathbf{k},\mathbf{e}\in\Lambda'} |d_{j,\mathbf{k},\mathbf{e}}|^p \right)^{\frac{1}{p}} \right)^r. \quad \text{(IV.3)}$$

This definition relies apparently on the norm equivalence relation between Besov seminorm and wavelet coefficients, similar to (I.4), compare [HKPT]. Although it may be interesting to choose different values of p and r, we have always taken $p = r = 2$ in which case (IV.3) is equivalent to a Sobolev semi-norm and we abbreviate $E_\alpha^{\Lambda'}(g) := E_{\alpha,2,2}^{\Lambda'}(g)$. In view of the norm equivalence (I.4), if (x_i, z_i) is indeed an outlier, in the neighborhood of x_i the local energy of $f^{[i]}$ should be much smaller than the local energy of f. This motivates the following definition: we define the *merit profile* of point (x_i, z_i) according to a *global criterion* as

$$\omega_{\rm global}(i) := \log\left(\frac{E_\alpha^{\Lambda_j^{[i]}}(f)}{E_\alpha^{\Lambda_j^{[i]}}(f^{[i]})} \right). \quad \text{(IV.4)}$$

With this definition, in view of Step (III) from Section III, we found in our experiments that a typical thresholding value denoted by τ which performs well should be in the order of magnitude of 1. Points (x_i, z_i) for which $\omega_{\rm global}(i)\geq\tau$ are then classified as an outlier, whereas points for which $\omega_{\rm global}(i)<\tau$ are not. This means that in our model we expect the presence of an outlier to cause a noticeable increase of the local energy.

Revisiting the synthetic Data Set 1 from Figure 1, we next explore in detail 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 illustrated in Figure 3. As expected, no

difference is visible outside the red box in the upper left corner of the wavelet coefficients starting at level 4. The presence of the outlier really does act locally. We illustrate this in more

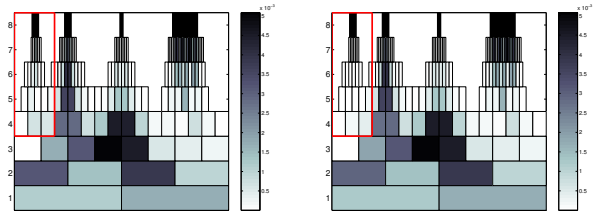

Fig. 3. Global removal criterion applied to (x_{100}, z_{100}). Wavelet coefficients of f (left) and of $f^{[100]}$ (right).

detail in Figure 4. In the first plot we compare f (dashed line) and $f^{[100]}$ (solid line) in the neighborhood of (x_{100}, z_{100}). The former obviously does not need to create the jump that tries to reproduce the point (x_{100}, z_{100}). This is reflected by the coefficients of the wavelets in $\Lambda^{[100]}$, as one can see in the following plots of the same figure: the energy content of the set $\Lambda^{[100]}$ is practically empty after subtracting (x_{100}, z_{100}) from the data. In computing the energy, a cutting level $j_{\text{cut}} = 3$ was chosen. The local measure of energy has parameters $\alpha = 5$. Criterion (IV.4) would give a merit criterion of $\omega_{\text{global}}(100) = 7.02$.

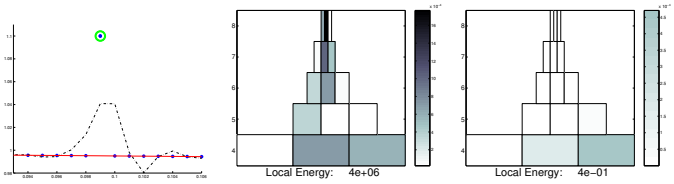

Fig. 4. Global removal criterion on (x_{100}, z_{100}). Local view of f (dashed line) and $f^{[100]}$ (solid line) (figure on the left). Coefficients of wavelets in $\Lambda^{[100]}$ for f (middle) and $f^{[100]}$ (right).

The same computation for the regular neighboring point (x_{102}, z_{102}) gives the coefficient to measure the 'amount of outlierness' as $\omega_{\text{global}}(102) = 0.0078$, as its removal does not critically vary the local energy, see Figure 5.

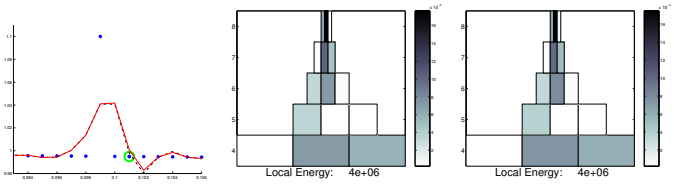

Fig. 5. Global removal criterion applied to (x_{102}, z_{102}). Local view of f and $f^{[102]}$ (left). Coefficients of wavelets in $\Lambda^{[102]}$ for f (middle) and $f^{[102]}$ (right).

Furthermore, the point (x_{251}, z_{251}) is located in the middle of a fine structure feature of the data. The removal of the point damages but does not destroy the structure, as can be seen from Figure 6. The similarity of the two approximations gives an outlierness coefficient of $\omega_{\text{global}}(251) = -0.27$. Thus, the algorithm will mark it as a regular point, preventing the loss of information.

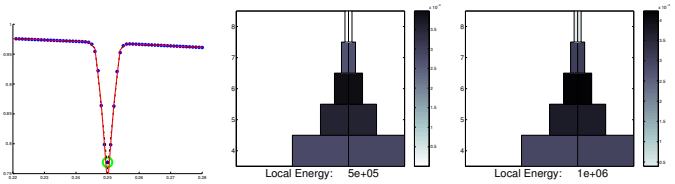

Fig. 6. Global removal criterion applied to (x_{251}, z_{251}). Local view of f and f_i (left). Coefficients of wavelets in Λ_{251} for f (middle) and f_{251} (right).

Finally, we consider the outlier (x_{800}, z_{800}). As it is located in a highly energetic environment, it is more complicated to disentangle its effects from those of the data features. In Figure 7 one can see that the local energy decay caused by the subtraction of the point is not so dramatic as in the case of the outlier in a flat background, compare with Figure 4. Nevertheless, (x_{800}, z_{800}) attains a merit value of $\omega_{\text{global}}(800) = 1.2$, so that the method would classify the point correctly as an outlier.

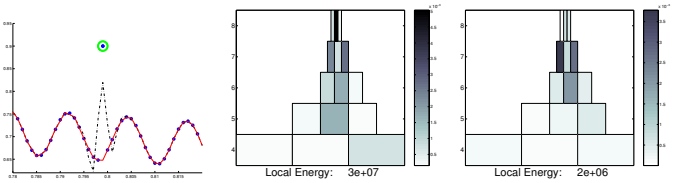

Fig. 7. Global removal criterion applied to (x_{800}, z_{800}). Local view of f and f_i (left). Coefficients of wavelets in Λ_{251} for f (middle) and f_{251} (right).

Removal of Points in Normal Equations. Despite these convincing results, the proposed strategy has an obvious drawback: it requires the computation of $f^{[i]}$ for every suspicious point (x_i, z_i). This amounts to construct and solve a different set of normal equations for every i. Fortunately, the structure of the problem allows for some possible simplifications if one considers the linear system which is to be solved for the data fitting procedure. Abbreviating the normal equations of the original problem (I.5) by $Md = b$ and the normal equations of the problem with the point (x_i, z_i) removed by $M^{[i]}d^{[i]} = b^{[i]}$, the relation between the two systems can be expressed as $M^{[i]} = M - a_i^T a_i$, and $b^{[i]} = b - a_i^T z$, where a_i denotes the i–th row of the observation matrix A.

This structure makes the construction of the new normal equations a trivial task and allows for two possible ways to simplify the solution process:

1) Under our assumption of an outlier having only an effect local in scale and space, one expects d to be a good approximation to $d^{[i]}$. Thus, an iterative method needs only few iterations to find $d^{[i]}$ having d as initial guess. This is particularly so if the matrix M is well–conditioned, as it is in our case here when working with our boundary–adapted spline–wavelets [CK1].
2) The inverse of $M^{[i]}$ is available by means of the inverse

of M, using the Sherman–Morrison identity

$$(M^{[i]})^{-1} = (M - a_i^T a_i)^{-1} = M^{-1} + \frac{M^{-1} a_i^T a_i M^{-1}}{1 - a_i M^{-1} a_i^T}.$$

Of course, a direct application of this formula is not advisable in general, as it requires the inversion and storage of the inverse of M (or of its QR factorization). In view of the sparsity structure of M, this results in a computational complexity of order $O((\#\Lambda)^2 \log(\#\Lambda))$. In cases where we can take full advantage of adaptivity, namely, when a large amount of data can be described by only a small number of wavelets, i.e., $\#\Lambda \ll N$, the computational complexity would therefore still be relatively small.

We exploit this later in Section VI where we use the Sherman–Morrison formula on the normal equations.

V. Local Criterion

The method described in the previous section seems to give a very reliable characterization of outliers. It attains high outlier detection rates in all our experiments. The drawbacks of this method are therefore not of conceptual but merely of computational nature: it can involve severe costs when recomputing a new approximating function $f^{[i]}$ for every point (x_i, z_i) to be checked.

In this section we present a modification of the above algorithm to speed up the computations which uses the semi–orthogonality of the wavelet basis.

Recall that the common ground for outlier detection is the comparison of f and $f^{[i]}$. This comparison is done locally in space, selecting wavelets located around the location of the outlier, and frequency, picking the higher scales. According to this implicit characterization of outliers based on *local* features, the idea now is to replace the comparison of f and $f^{[i]}$ with a comparison of functions which approximate f and $f^{[i]}$ *locally*. To this end, we consider an approximation of the data up to some coarse level $j < J$, where J is the maximal resolution level included in the original set Λ, by simple restricting the original function f to a maximal level j

$$f^j := \sum_{\lambda \in \Lambda;\ |\lambda| \le j} d_\lambda \psi_\lambda. \tag{V.1}$$

We construct now local approximations to f and $f^{[i]}$ in the neighborhood of x_i by keeping the coefficients of f up to level j and adding only *local wavelets* to describe higher detail features. These local wavelets are the ones which we included in the set $\Lambda_j^{[i]}$ defined in (IV.2). The method of the previous section is now mimicked by building two approximations, one that takes into account the point x_i, $f^{j,i}$, and one that does not, $f^{j,[i]}$. That is, we define

$$f^{j,i} := \sum_{\lambda \in \Lambda;\ |\lambda| \le j} d_\lambda \, \psi_\lambda + \sum_{\lambda \in \Lambda_j^{[i]}} d_\lambda^{j,i} \, \psi_\lambda. \tag{V.2}$$

where the vector $d^{j,i} = \{d_\lambda^{j,i}\}_{\lambda \in \Lambda_j^{[i]}}$ is computed such that

$$\sum_{\ell=1}^{N} \left(z_\ell - f^{j,i}(x_\ell)\right)^2 \tag{V.3}$$

attains its minimum. Note that the locality of wavelets included in $\Lambda_j^{[i]}$ forces most terms in the summation to be zero, and the non zero elements are directly accessible from the sparsity pattern in which we have coded the observation matrix. Likewise, we define

$$f^{j,[i]} := \sum_{\lambda \in \Lambda;\ |\lambda| \le j} d_\lambda \, \psi_\lambda + \sum_{\lambda \in \Lambda_j^{[i]}} d_\lambda^{j,[i]} \psi_\lambda, \tag{V.4}$$

where the vector $d^{j,[i]} = \{d_\lambda^{j,[i]}\}_{\lambda \in \Lambda_j^{[i]}}$ is the LSW–fit of

$$\sum_{\ell \ne i} \left(z_\ell - f^{j,[i]}(x_\ell)\right)^2. \tag{V.5}$$

Consequently, we can define a merit criterion based on this *local criterion* as

$$\omega_{\text{local}}(i) := \log \left(\frac{E_\alpha^{\Lambda_j^{[i]}}(f^{j,i})}{E_\alpha^{\Lambda_j^{[i]}}(f^{j,[i]})} \right). \tag{V.6}$$

One can interpret this process as 'freezing' f^j and 'gluing' onto it a local approximation to the set $\{(x_i, z_i - f^j(x_i))\}_{i=1,\dots,N}$. The implicit assumption behind this is that 'freezing' and 'gluing' will maintain a similar spectrum of local energies which is justified by our use of a wavelet basis. The semi–orthogonality property allows us to operate this 'level surgery' and thereby treating different scales separately.

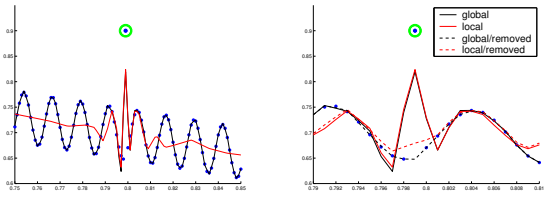

Fig. 8. Global and local reconstructions near (x_{800}, z_{800}). Local view of f and $f^{3,800}$ (left) and removal of (x_{800}, z_{800}) in global and local approximations.

Let us illustrate this procedure in Figure 8. On the left, the black line represents the global approximation to the data f and the red line represents the local approximation $f^{3,800}$ to the data. We observe that the degrees of freedom used in the construction of $f^{j,i}$ seems to perfectly fit our purposes, as the local approximation near the outlier artifact is nearly indistinguishable from the global approximation. Obviously this diverges from it outside of this narrow area. On the right, we see in a close–up the different functions near the point. Solid lines include information of the point and dashed lines do not. We observe that global (black) and local (red) approximations yield similar results (dashed and solid lines diverge near the outlier and approach each other away from it). In addition, as one could expect in view of the plot, the two 'outlierness criteria' are correspondingly similar, $\omega_{\text{global}}(800) = 1.2$ and $\omega_{\text{local}}(800) = 1.7$. These coefficients reflect a decrease of more than one order of magnitude in the respective local energies by removal of the point (x_{800}, z_{800}). The merit criterion for the regular neighboring point (x_{802}, z_{802}) is $\omega_{\text{global}}(802) = -0.15$ by the

global criterion (IV.4) and $\omega_{\text{local}}(802) = -0.13$ by the local criterion (V.6). This indicates that the presence of this point causes just a minor readjustment of the local reconstruction in Figure 9, amounting only to a slight local energy variation.

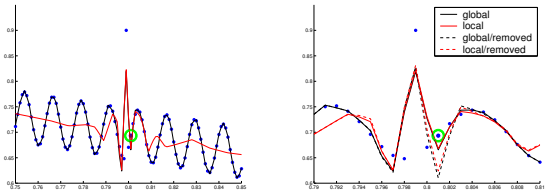

Fig. 9. Global and local reconstructions near (x_{802}, z_{802}). Local view of f and $f^{3,802}$ (left) and removal of (x_{802}, z_{802}) in global and local approximations.

This procedure obviously reduces the computation costs of the global method, as the number of degrees of freedom involved in the computation of $f^{j,i}$ and $f^{j,[i]}$ is just $\#\Lambda_j^{[i]}$, a fraction of the total number of wavelets $\#\Lambda$.

A further property implicit in this procedure is that the coarse scale projection f^j is in fact not affected by the outlier. This means that the effect of every outlier is restricted to dyadic levels higher than j. This makes the selection of j an important issue. A too small value j results in including in the local configuration $\Lambda_j^{[i]}$ wavelets of low frequency that do not notice the effect of the outlier. If the coefficients of these wavelets are much larger than the coefficients of the high frequencies, they may mask the effect of the removal of the outlier. In the global algorithm, the importance of a good selection of j is relative, as it only concerns the way in which the functions f and $f^{[i]}$ are compared, not the functions themselves. In the local criterion, an overestimated j can lead to an underestimation of the outlierness of the point: if we accept as coarse description of the data the one constructed by taking into account frequencies affected by the outlier, we are masking its effect. Precisely for this reason we develop in the next section a local-corrected criterion, which tries to represent a trade-off between the computational cheapness of the local criterion and robustness against underestimation of j of the global criterion.

VI. Local Corrected Criterion

We have noticed that the main drawback of the local relaxation method is the possibility of the effect of outliers filtering down to the coarse frequencies so that a local approximation to the data cannot reveal outliers. The obvious way to circumvent this problem is to construct a coarse level approximation which is not influenced by the suspicious point. Thus, we seek for a

$$\hat{f}^{j,i} := \sum_{\lambda\in\Lambda;\ |\lambda|\le j} \hat{d}_\lambda^{j,i}\psi_\lambda \tag{VI.1}$$

which minimizes

$$\sum_{\ell=1}^{N}\left(z_\ell - \hat{f}^{j,i}(x_\ell)\right)^2. \tag{VI.2}$$

The next step is like in the previous section: extend locally the degrees of freedom and compute for this configuration an approximation to the data with and without the point (x_i, z_i). That is, define

$$\tilde{f}^{j,i} := \sum_{\lambda\in\Lambda;\ |\lambda|\le j} \hat{d}_\lambda\psi_\lambda + \sum_{\lambda\in\Lambda_j^{[i]}} \tilde{d}_\lambda^{j,i}\psi_\lambda. \tag{VI.3}$$

where the vector $\tilde{d}^{j,i} = \{\tilde{d}_\lambda^{j,i}\}_{\lambda\in\Lambda_j^{[i]}}$ is computed such that it minimizes

$$\sum_{\ell=1}^{N}\left(z_\ell - \tilde{f}^{j,i}(x_\ell)\right)^2 \tag{VI.4}$$

Likewise, we define

$$\tilde{f}^{j,[i]} := \sum_{\lambda\in\Lambda;\ |\lambda|\le j} \hat{d}_\lambda\psi_\lambda + \sum_{\lambda\in\Lambda_j^{[i]}} \tilde{d}_\lambda^{j,[i]}\psi_\lambda, \tag{VI.5}$$

where the vector $\tilde{d}^{j,[i]} = \{\tilde{d}_\lambda^{j,[i]}\}_{\lambda\in\Lambda_j^{[i]}}$ results from minimizing

$$\sum_{\ell\neq i}\left(z_\ell - \tilde{f}^{j,[i]}(x_\ell)\right)^2. \tag{VI.6}$$

Then, one can compute the local energies and state a merit figure based on this local corrected fitting of the data like in the previous sections, i.e.,

$$\omega_{\text{local}+}(i) := \log\left(\frac{E_\alpha^{\Lambda_j^{[i]}}(\tilde{f}^{j,i})}{E_\alpha^{\Lambda_j^{[i]}}(\tilde{f}^{j,[i]})}\right). \tag{VI.7}$$

In the plots in Figure 10 approximations for the environment of (x_{100}, z_{100}) from Data Set 1 are shown. We have visualized an example of the three methods presented so far: the global, the local and the local–corrected procedure, from left to right. In the two former the level j_{cut} is fixed as 7. The

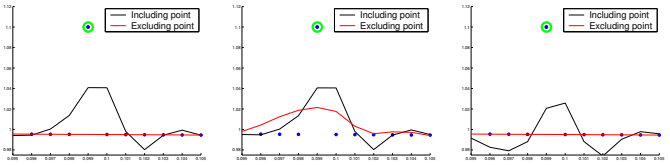

Fig. 10. Comparison of criteria for (x_{100}, z_{100}) with $j_{\text{cut}} = 7$. Global removal criterion at (x_{100}, z_{100}) (left), local criterion (middle) and local corrected criterion (right).

global approximation criterion yields $\omega_{\text{global}}(100) = 8.4$. The local approximation criterion gives a much smaller value $\omega_{\text{local}}(100) = 1.7$. The reason can be inferred from the central plot of the figure: the dashed (red) line, which ideally should not be affected by the outlier, is clearly affected by the presence of it, as it is constructed starting from a coarse–scale approximation to the *whole* data set, also including the outlier. We could formulate this effect by saying that the prescribed $j_{\text{cut}} = 7$ does not correspond to the actual penetration depth of the outlier. One possible solution could be to vary j_{cut}. The other solution is the one presented in this section: the use of an outlier–free coarse–scale approximation. The result is given in the right plot in Figure 10. Note that the dashed (red) line does not result in any undesirable influence from (x_{100}, z_{100}) and succeeds to produce no artifacts in its environment. The outlier coefficient arising from this criterion is $\omega_{\text{local}+}(100) = 8.3$.

Obviously, we are re–introducing some amount of computational overhead, as now $f^{j,[i]}$ has to be computed for every $i \in S$. The point is that this overhead is much more affordable, as it affects a reduced number of coefficients. The observations made in Section IV about point removal procedures from the normal equations can now be exploited.

VII. Robust Approximation in Highly Energetic Environments

In this section we want to envisage some extremal cases of the proposed methods. These refer to the very definition of the outlier. The basic intuition of the outlier concept is that it is a point whose presence in the data creates an outburst of the local energy of the approximating functions. This perspective carries the obvious consequence that outliers would not be detected which lie in domain areas where regular other points also produce similar energy variations. Thus, some further refinement might be necessary in the outlier defining criterion. There are several situations in which this effect may occur:

- outliers embedded in areas of rapid spacial variability;
- simultaneous presence of other noise sources;
- a high rate of outlier contamination.

A. Presence of Noise

We illustrate this with the following experiment. We add to the data in Figure 1 some quantity of random noise, whose amplitude is controlled by the parameter σ. For different noise amplitudes σ, we obtain, for instance, the data distributions shown in Figure 11. In the cases $\sigma = 0.01$ and $\sigma = 0.05$ the

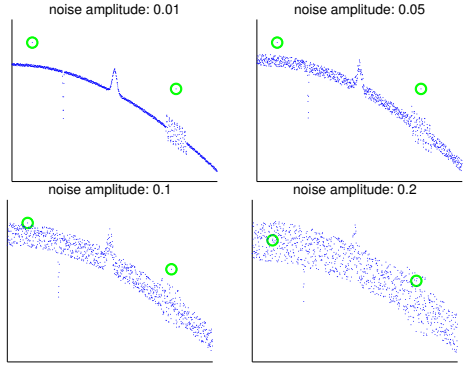

Fig. 11. Noisy data. Parameter $\sigma \in \{0.01, 0.05, 0.1, 0.2\}$ from left to right.

two outliers (x_{100}, z_{100}) and (x_{800}, z_{800}) are perfectly distinguishable from the noisy background, and it is to be expected that the criteria mentioned above will give good estimations. In the case $\sigma = 0.1$ point (x_{100}, z_{100}) is embedded into the noise. A correct and robust outlier finding criterion should not identify this point as an outlier, whereas (x_{800}, z_{800}) should be marked. Finally, in the case $\sigma = 0.2$ both original outliers are not distinguishable from the point cloud, and an outlier identification would not make sense. In correspondence with the above description of rough features of the data, we apply in Figure 12 the global criterion defined in (IV.4) for the whole bunch of points. We see that in the two extreme cases the criteria works well: $\sigma = 0.01$ gives good results although the

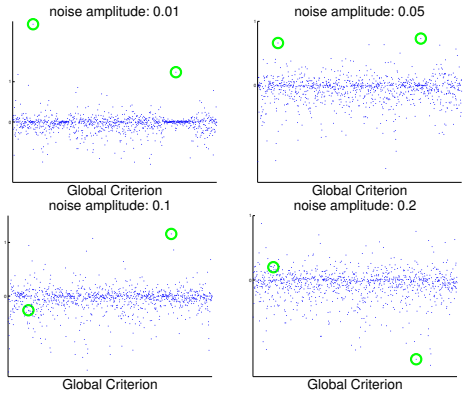

Fig. 12. Values of ω_{global} for data sets in Figure 11; parameter $\sigma \in \{0.01, 0.05, 0.1, 0.2\}$ from left to right.

discrimination of outliers is not so clear as in the previous case. The results for the case $\sigma = 0.2$ also fit the idea of 'outlierness' as the criterion does not recognize any special feature in the marked points, according to their inclusion into the noisy background. The intermediate cases also work: at $\sigma = 0.05$ both outliers are marked, whereas at $\sigma = 0.1$ point (x_{100}, z_{100}) is correctly ignored by the criterion, as it is embedded into the surrounding noise. Point (x_{800}, z_{800}) is also successfully provided with a large value of ω_{global}.

However, in these cases the discrimination of outliers is not so clear as in the previous case: observe that a number of regular data points also attain a large outlierness criterion. This is not a mayor problem: by its very definition, a false detection amounts to eliminate a data point which carries redundant information. As long as the remaining points still reproduce the whole set of significant data features, the loss of a moderate number of data points can be considered admissible. This issue is revisited in the following sections.

In any case, the criterion can be refined in order to reduce the loss of actual information. False detections affect points whose removal of the data yields a noticeable decrease of the local energy. As our criterion measures this decrease in relation to the original energy, in areas where this is very small, the local reconfiguration of the wavelet spectrum after removal of a data point can happen to produce a still lower local energy, without this decrease being significant. We can cope with this situation in different ways. Firstly, we can impose stricter thresholding policies in the processing step of Section III to rule out points lying in flatter areas. In the present case, where the data is corrupted by high–frequency noise, one should filter it with a classical wavelet smoothing procedure, as long as one has a statistical model for this noise. A second strategy would be to simply build the local energy factor into the criterion. In Figure 13, we see the values of $e_{\text{local}}(i) := E_{\alpha}^{\Lambda_{j_{cut}}^{[i]}}(f)$ for each point of the data set in the four level–of–noise scenarios. If we multiply the global criterion profile with the local energy profile we get the plots of Figure 14, where the discrimination of outliers appears much clearer than in Figure 12.

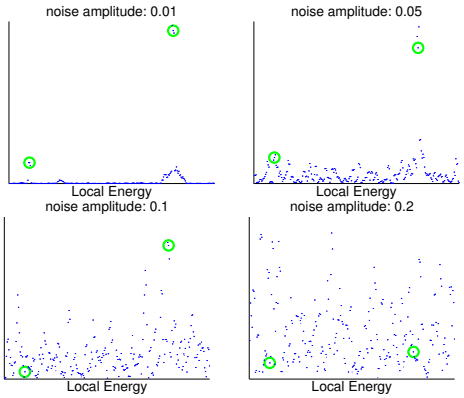

Fig. 13. e_{local} profile for data sets in Figure 11; parameter $\sigma \in \{0.01, 0.05, 0.1, 0.2\}$ from left to right.

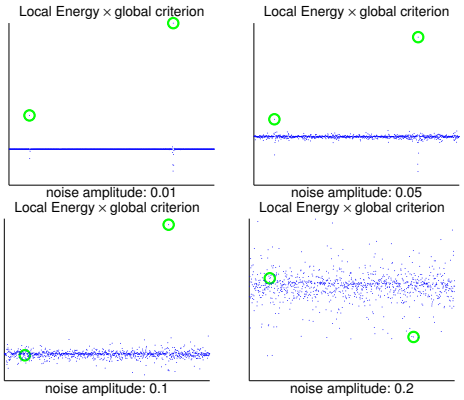

Fig. 14. Profile of the product $\omega_{\mathrm{global}} \cdot e_{\mathrm{local}}$ for data sets in Figure 11; parameter $\sigma \in \{0.01, 0.05, 0.1, 0.2\}$ from left to right.

B. Large Number of Outliers

Yet another possible source of problems is the outlier density. The capability of the method to disentangle outliers from surrounding signal lies in a characterization of the local energy of this surrounding signal. If further outliers are present in the neighborhood, this characterization fails, and consequently the outlier marking criterion fails as well.

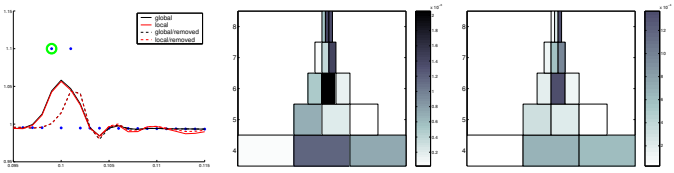

Fig. 15. Proximity of outliers: analysis on (x_{100}, z_{100}). Local view behavior of approximants (left), coefficients of wavelets in $\Lambda^{[100]}$ (middle), coefficients of wavelets in $\Lambda^{[100]}$ after removal of (x_{100}, z_{100}) (right).

We see an example in Figure 15. We add to the original data a new outlier by imposing the value $y_{102} = 1.1$. This represents an outlier in the immediate neighborhood of the original outlier (x_{100}, z_{100}). If we compute now the outlierness profile for each of these points, we find that the 'outlierness criterion' of (x_{100}, z_{100}) in this data set is 0.3, computed by both global and local criteria. Recall that this point enforced an 'outlierness coefficient' $\omega_{\mathrm{global}}(100) = 7.2$ when it was isolated. The reason for the low values ω can be read from the reconstructions given in the figure: the suppression of the outlier (x_{100}, z_{100}) does not locally relax the approximation, as the remaining outlier (x_{102}, z_{102}) still twists the approximation toward this point. There is indeed some energy decay, as one can deduce from the wavelet spectrum in the central and right plots of the same Figure, but not as severe as when no further outlier corrupts the background: compare Figure 15 with Figure 3. This means that the number of outliers which

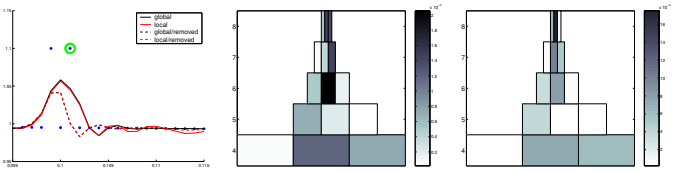

Fig. 16. Proximity of outliers: analysis on (x_{102}, z_{102}). Local behavior of approximants (left), coefficients of wavelets in $\Lambda^{[102]}$ (middle), coefficients of wavelets in $\Lambda^{[102]}$ after removal of (x_{102}, z_{102}) (right).

can be present in a data set without corrupting it depends, evidently as well as on the data set, on their *distribution profile*, as it is the proximity of outliers to each other which drives the method to fail.

We perform the following series of experiments: we corrupt the original data by a fixed number μ of outbursts for two different designs. In one design, the outliers are equidistantly placed. In the other one, outliers are randomly distributed. We run then our outlier finding procedure for different choices of μ. In Tables 1 through 4, we give the percentages of successful outlier detection and false detection for different choices of τ and a different number of outliers.

TABLE I
OUTLIER DETECTION PERCENTAGE IN EQUIDISTANT DESIGN.

τ / μ	20	50	100	200	300	500
0.1	100.0	100.0	98.0	42.0	4.3	0.2
0.2	100.0	100.0	98.0	38.5	1.0	0.2
0.3	100.0	100.0	97.0	17.5	0.3	0.0
0.4	100.0	100.0	96.0	3.0	0.0	0.0
0.5	100.0	96.0	96.0	0.5	0.0	0.0

TABLE II
OUTLIER DETECTION PERCENTAGE IN RANDOM DESIGN.

τ / μ	10	50	100	200	300	$\mu = 500$
0.1	80.0	82.0	70.0	45.5	46.0	16.4
0.2	70.0	72.0	61.0	35.5	30.3	8.2
0.3	60.0	64.0	55.0	26.0	21.3	3.2
0.4	60.0	60.0	47.0	20.0	15.7	2.2
0.5	60.0	52.0	41.0	16.5	11.3	1.8

The results are quite expectable. In the equidistant case the outliers are located quite well when they are distanced (low values of μ), up to the critical distance in which every outlier suffers the influence of two neighbors and the method

TABLE III

FALSE DETECTION PERCENTAGE IN EQUIDISTANT DESIGN.

τ / μ	10	50	100	200	300	500
0.1	0.3	1.2	0.0	2.7	18.0	0.0
0.2	0.0	0.0	0.0	0.9	11.7	0.0
0.3	0.0	0.0	0.0	0.0	6.8	0.0
0.4	0.0	0.0	0.0	0.0	5.5	0.0
0.5	0.0	0.0	0.0	0.0	4.3	0.0

TABLE IV

FALSE DETECTION PERCENTAGE IN RANDOM DESIGN.

τ / μ	20	50	100	200	300	500
0.1	0.7	1.4	1.5	3.1	3.4	6.4
0.2	0.0	0.2	0.6	1.0	1.4	3.0
0.3	0.0	0.1	0.3	0.3	0.5	1.4
0.4	0.0	0.1	0.2	0.2	0.3	0.7
0.5	0.0	0.1	0.1	0.1	0.1	0.3

collapses abruptly. In the random design, the outlier detection rate is not so successful for a small concentration of outliers. This is caused by the number of outliers which result to occur close to each other, in spite of a low total number of outliers. In compensation, the method attains a higher detection rate when the total number of outliers is larger, as a number of outliers occurs isolated from the others.

VIII. ENERGY CRITERIA

The three outlier detection methods explained until now are based in the construction on a couple of functions (one that sees the whole action of the possible point, and one that mollifies it) and its comparison. The reason for these criteria to work well is provided by the Riesz Basis property of wavelets which yields the characterization of the norm of a wide range of spaces in terms of the wavelet coefficients, as exploited in definition (IV.3). In the criteria used so far (global, local and local corrected) we compared $e_{\text{local}}(f)$ and $e_{\text{local}}(f^{[i]})$, changing only the definition of $f^{[i]}$. We call these *direct criteria*.

However, we could use the same argument provided by the Riesz Basis property and propose to employ $e_{\text{local}}(f - f^{[i]})$ as an outlier finding criteria. This is also a natural choice which we call *residual criterion*. Some differences to the previous strategy are the following:

- Variable order of magnitude of adequate thresholding parameters. In direct methods the order of magnitude of the parameter τ appears to be quite intuitive: the energy change must be numerically noticeable. In the residual methods, on the contrary, one usually finds appropriate values for τ, but they are obviously very sensitive to the data and the underlying function.
- Different performance. If the addition of an outlier enforces a *redistribution* of local energy rather that an *increase* of it, see Section VII for situations in which this may occur, the direct methods will fail to detect it, as explained above, but residual methods still have a chance. The prize for this is that the methods are more expensive to compute.

Consider for instance the function on the left of Figure 17. We add an overall background noise and 5% of outliers of diverse amplitude and randomly distributed, as plotted in the center of the figure. Our LSW method gives the reconstruction at the right of the figure. If we compute our full set of criteria on our reconstruction, we get the successful detection as well as the wrong elimination percentages given in Table 5 for the direct methods and in Table 6 for the residual method. The numbers confirm our expectations, see the two reconstructions given in Figure 18. Direct methods fail to find the full set of outliers. Residual methods filter more outliers but cannot avoid throwing away more data points and possibly relevant information; compare the reconstruction of the high energetic feature located at $x = 0.8$, which appears much more damaged in the right plot than in the left.

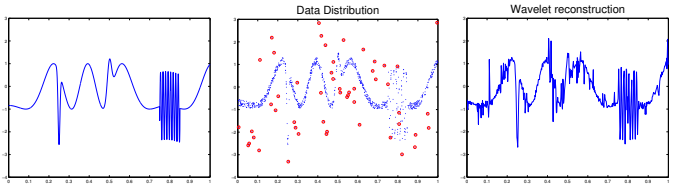

Fig. 17. Data Set 2. Original function (left), irregular sampling with noise and outliers (middle) and wavelet reconstruction (right).

TABLE V

PERFORMANCE OF DIRECT METHODS IN THE ANALYSIS OF THE OUTLIER–CORRUPTED DATA FROM FIGURE 17.

% Detected Outliers

τ	glob	local	local+
0.00	75.0	73.1	75.0
0.05	69.2	67.3	67.3
0.10	65.4	59.6	63.5
0.20	63.5	51.9	51.9
0.30	57.7	46.2	44.2
0.40	42.3	40.4	36.5

% Eliminated Data

τ	glob	local	local+
0.00	34.7	35.7	35.0
0.05	11.1	9.6	9.4
0.10	5.9	5.6	5.6
0.20	2.5	3.1	3.3
0.30	1.7	2.0	1.6
0.40	1.2	1.3	1.0

TABLE VI

PERFORMANCE OF RESIDUAL METHODS IN THE ANALYSIS OF THE OUTLIER–CORRUPTED DATA FROM FIGURE 17.

% Detected Outliers

τ	glob	loc	loc+
1.0e+05	100	100	100
5.0e+06	96.2	98.1	98.1
1.0e+07	92.3	90.4	92.3
1.5e+07	90.4	88.5	88.5
5.0e+07	76.9	65.4	69.2
1.0e+08	65.4	57.7	61.5
1.5e+08	57.7	46.2	50.0
2.0e+08	42.3	36.5	38.5

% Eliminated Data

τ	glob	loc	loc+
1.0e+05	75.7	76.9	77.5
5.0e+06	28.5	29.3	29.5
1.0e+06	22.4	23.1	23.6
5.0e+07	19.5	19.2	19.4
1.0e+07	10.6	10.5	10.8
5.0e+08	6.7	6.3	7.3
1.5e+08	5.3	4.7	5.5
2.0e+08	4.6	3.9	4.8

As a final remark, note that the use of $f - f^{[i]}$, that is, residual methods, would allow us to also use B–Splines as ansatz functions, as the measure of $f - f^{[i]}$ in L_2 could be reasonably understood as an indicator for outlier presence. In contrast, direct methods are only meaningful in a wavelet ansatz.

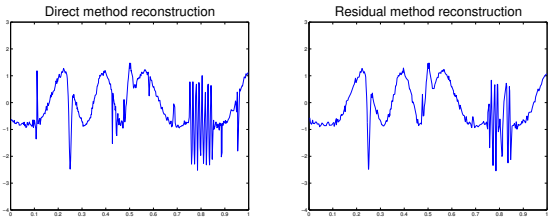

Fig. 18. Reconstructions of example Data Set 2, after removal of points marked as outliers. Data cleaned by a direct method (left) and a residual method (right).

TABLE VII
PERFORMANCE OF DIRECT CRITERIA IN THE ANALYSIS OF THE OUTLIER–CORRUPTED DATA FROM FIGURE 19.

% Detected Outliers

τ	glob	loc	loc+
0.000	90.0	90.0	90.0
0.025	90.0	90.0	90.0
0.050	85.0	85.0	85.0
0.075	85.0	85.0	85.0
0.100	75.0	80.0	80.0
0.125	75.0	75.0	80.0
0.150	70.0	75.0	80.0
0.175	70.0	70.0	70.0
0.200	65.0	70.0	70.0

% Eliminated Data

τ	glob	loc	loc+
0.000	34.2	35.1	34.5
0.025	1.2	2.0	2.0
0.050	0.5	0.4	0.4
0.075	0.3	0.2	0.2
0.100	0.1	0.1	0.1
0.125	0.1	0.1	0.1
0.150	0.1	0.0	0.0
0.175	0.0	0.0	0.0
0.200	0.0	0.0	0.0

IX. HIGHER SPATIAL DIMENSIONS

Our methods can naturally be extended to higher dimensions, choosing again the semi-norm (IV.3) as the relevant energy norm for the case $p = r = 2$. We see an example of the procedure with the data in Figure 19. In the plot on the left we see a view of the well known Franke function. We want to reconstruct it from the 2000 scattered randomly chosen sampling points given in the central plot. To the 20 points marked in red we add a constant value, creating in this way a random distribution of outliers in the original data. We see in the plot on the right a wavelet reconstruction found by our standard LSW method. Here we can observe how the presence of the outliers creates undesired local oscillations in the surface in the proximity of each outlier. The parameter q steering the candidates for tree growth in the adaptive wavelet procedure is in the 2D examples always set to $q = 100$, and $j_{\text{cut}} = 7$.

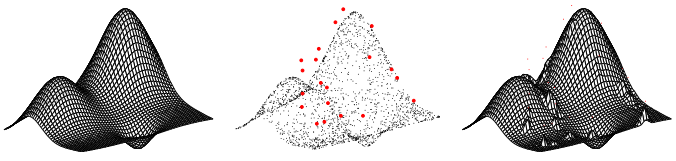

Fig. 19. Outlier distribution in a scattered sampling with 2000 points of the Franke function (left). Sampling geometry with outliers (red points; middle). Wavelet reconstruction with 20 points (right).

To assert the performance of the method for this data set we compute the percentage of the outliers found and false detections for the several criteria we have discussed. In Table 7 we see the results obtained by the direct methods.

TABLE VIII
PERFORMANCE OF RESIDUAL CRITERIA IN THE ANALYSIS OF THE OUTLIER–CORRUPTED DATA FROM FIGURE 19.

% Detected Outliers

τ	glob	loc	loc+
0.000	100.0	100.0	100.0
0.500	100.0	95.0	100.0
1.000	95.0	90.0	95.0
2.000	85.0	80.0	90.0
3.000	75.0	70.0	80.0
4.000	75.0	60.0	75.0
5.000	65.0	55.0	70.0

% Eliminated Data

τ	glob	loc	loc+
0.000	99.0	99.0	99.0
0.500	3.1	1.1	1.9
1.000	0.9	0.4	0.8
2.000	0.6	0.2	0.4
3.000	0.5	0.1	0.4
4.000	0.4	0.1	0.4
5.000	0.4	0.0	0.4

The results assert the likeliness of the three criteria. All three of them give a successful rate of outlier finding with minor losses of non–corrupted data points. Note, however, that the method does not provide the complete detection of outliers, as the outlier interaction effect described in Section VII attains to hide some of them. A second run of this method on the data (after removal of the outliers detected in the first run) detects successfully the remaining outliers. According to our argumentation in Section VIII, the residual methods can disentangle better this interaction and, consequently, detect all the outliers in just one run. The price of this method is a slightly higher rate of false detections.

A further example is provided in the analysis of a geophysical data set [PR]. The set includes 18634 points ordered in a square grid, plotted in the left of Figure 20. We add 1000 randomly distributed outliers to this data, yielding the data on the right of the same Figure. In the left of Figure 21 we can see the performance of our algorithm with one run. A first run 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, an iteration on the cleaned data, offers 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 masked by neighboring ones have now been successfully detected.

X. CONCLUSION

We have presented in this article some robust regression techniques to handle outliers within a coarse-to-fine data fitting algorithm based on adaptive wavelets. Different criteria which are based on measuring the energy of reconstructions with and without the outlier based on weighted wavelet coefficient norms have been developed and tested numerically. Our adaptive wavelet scheme yields a numerically fast and reliable way to detect outliers which can amount up to 5% of the total amount of data.

ACKNOWLEDGMENT

We thank two anonymous referees for their constructive remarks.

REFERENCES

[Ca] D. Castaño. Adaptive Scattered Data Fitting with Tensor Product Spline–Wavelets. Dissertation, Math.—Nat. Fakultät, Universität Bonn, October 2004.

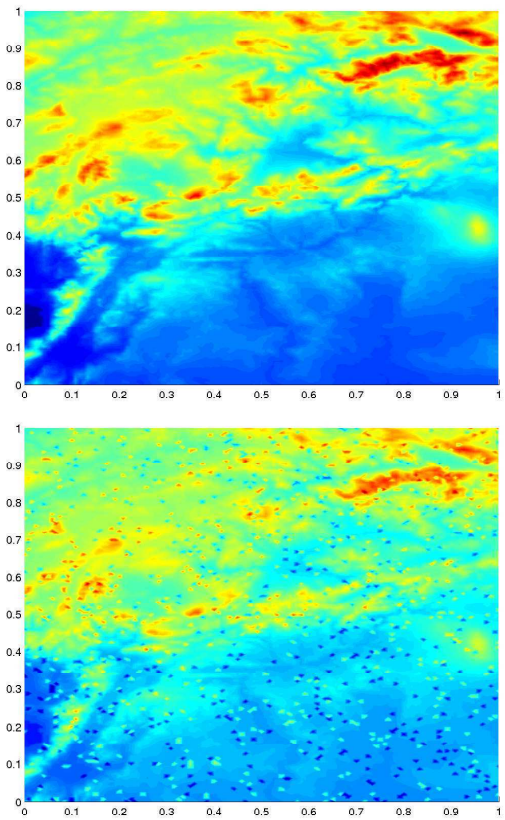

Fig. 20. Geophysical data set with $N = 18605$ data points. Vertical view (top); data set corrupted with 1000 outliers (% 5.6 of the data, bottom).

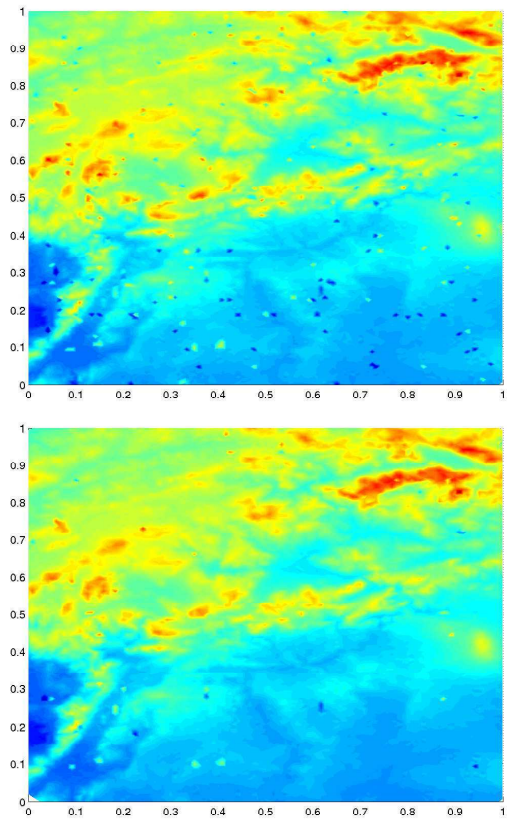

Fig. 21. Reconstruction of the geophysical data from Figure 20 after robust cleaning. First run (top); second run (bottom).

[CK1] D. Castaño and A. Kunoth. Adaptive fitting of scattered data by spline-wavelets. In: Curve and Surface Fitting, A. Cohen, J.-L. Merrien and L.L. Schumaker (ed.), Nashboro Press, 2003, 65–78.

[CK2] D. Castaño and A. Kunoth. Multilevel regularization of wavelet based fitting of scattered data — Some experiments. May 2004, 14 p., to appear in Numer. Algor.

[CDLL] A. Chambolle, R.A. DeVore, N.–Y Lee, B.J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. IEEE Trans. Image Proc. 7 (3), 1998, 319–335.

[Ch] C.K. Chui, An Introduction to Wavelets, Vol. 1. Academic Press, Inc., Boston, 1992.

[Co] R.L. Coldwell. Robust fitting of spectra to splines with variable knots. AIP Conf. Proc. 475(1), 1990, 604 ff.

[D1] W. Dahmen, Wavelet and multiscale methods for operator equations, Acta Numerica (1997), 55–228.

[DKU] W. Dahmen, A. Kunoth and K. Urban. Biorthogonal spline-wavelets on the interval – Stability and moment conditions. Appl. Comput. Harm. Anal., 6 (1999), 132–196.

[DV] R. A. DeVore, Nonlinear approximation, Acta Numerica, 1998, 51–150.

[GG] J. Garcke and M. Griebel, Data mining with sparse grids using simplicial basis functions, Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 26–29, 2001, San Francisco, California, 87–96.

[GHK] Th. Gerstner, H.–P. Helfrich and A. Kunoth. Wavelet analysis of geoscientific data. In: Dynamics of Multiscale Earth Systems, H.J. Neugebauer, C. Simmer (eds.), Lecture Notes in Earth Sciences, Springer, 2003, pp. 69–88.

[GHJ] B.F. Gregorski, B. Hamann, and K.I. Joy. Reconstruction of B-spline surfaces from scattered data points, in: Magnenat-Thalmann, N. and Thalmann, D., eds., Proceedings of Computer Graphics International 2000, 163–170.

[GH] G. Greiner and K. Hormann. Interpolating and approximating scattered 3D-data with hierarchical tensor product splines, in: Surface Fitting and Multiresolution Methods, A. Le Mehaute, C. Rabut and L.L. Schumaker (eds.), Vanderbilt University Press, Nashville, 1996, 163–172.

[HKPT] W. Härdle, G. Keryacharian, D. Picard, A. Tsybakoc. Wavelets, Approximation, and Statistical Applications. Springer, Lecture Notes in Statistics, 1998.

[He] M. Hegland. Adaptive sparse grids. ANZIAM J. 44(E), 2003, C335–C353.

[HPMM] J. Hofierka, J. Parajka, H. Mitasova and L. Mitas, Multivariate interpolation of precipitation using regularized spline with tension, Transactions in GIS 6(2), 2002, 135–150.

[HR] K. Höllig, U. Reif, Nonuniform web-splines, Computer Aided Geometric Design 20(5), 2003, 277–294.

[Hu] P.J. Huber. Robust Statistics, John Wiley & Sons, New York, 1981.

[LW] R.L. Lauer and G. N. Wilkinson (eds.), Robustness in Statistics, 1979.

[LWS] S. Lee, G. Wolberg, and S.Y. Shin, Scattered data interpolation with multilevel B-splines, IEEE Trans. Visualization and Computer Graphics 3(3), 1997, 228–244.

[PS] V. Pereyra and G. Scherer, Large scale least squares scattered data fitting, Appl. Numer. Maths. 44(1–2), 2002, 73–86.

[PR] The Puerto Rico Tsunami Warning and Mitigation Program. Data obtainable at http://poseidon.uprm.edu

[SHLS] V. Scheib, J. Haber, M.C. Lin, and H.P. Seidel, Efficient fitting and rendering of large scattered data sets using subdivision surfaces, Computer Graphics Forum (Proc. Eurographics 2002), 2-6 September 2002, 353–362.

[Sch] L.L. Schumaker, Fitting surfaces to scattered data, in: Approximation Theory II, G.G. Lorentz, C.K. Chui and L.L. Schumaker (eds.), Academic Press, New York, 1976, 203–268.

[SDS] E. J. Stollnitz, T. D. DeRose and D. H. Salesin, Wavelets for Computer Graphics, Morgan Kaufmann Publishers, 2000.

[Z] F. Zeilfelder, Scattered data fitting with bivariate splines, in: Tutorials on Multiresolution in Geometric Modelling, Mathematics and Visualization, A. Iske, E. Quak, and M.S. Floater (eds.), Springer-Verlag, Heidelberg, 2002.