Scattered data smoothing by empirical Bayesian shrinkage of second generation wavelet coefficients

Maarten Jansen[1], Guy Nason[2] and Bernard Silverman[2]

[1]ECE department, Rice University
[2]School of mathematics, University of Bristol

ABSTRACT

We propose a novel approach for scattered data smoothing based on second generation wavelets. This wavelet transform automatically adapts to the irregularity of the grid. Our implementation also pays attention to numerical stability, a crucial property in estimation procedures. The wavelet coefficients are shrunk either with simple soft-thresholding or with an empirical Bayesian estimation.

1. INTRODUCTION

Until recently, the application of wavelets in smoothing data on irregular grids was limited by the fact classical wavelet theory is built on a dyadic multiresolution analysis of equidistant samples. In order to proceed with non-uniform samples, most approaches included a sort of pre- or post-processing of the given data in order to reduce the problem to the equidistant setting.[18,21,14,6,2,24,23,4,3] Interpolation is a typical example of such a data correction method.

The lifting scheme[30] provides a natural way to extend[29,9,10] the classical wavelets to non-equidistant grids, and offers the perspective of smooth functions defined by irregular point sets. In smoothing applications, this obviously leads to a more direct and hence natural approach. As far as we know, smoothing algorithms using second generation wavelets are limited to one-dimensional data.[33,12,13] On the other hand, due to its close connection to subdivision, a common tool for surface generation in computer graphics, lifting is popular in that research area.[16]

At least four issues are important when dealing with second generation wavelets, that is, convergence of the subdivision scheme, smoothness of the limit function, numerical condition of the wavelet transform and approximation property(vanishing moments). Experience from computer graphics once more illustrates that working in multiple dimensions is essentially different from the one-dimensional case. Moreover, specific problems arise for applications in statistical estimation. Those include the computation of the correlation matrix of the wavelet coefficients. This computation, although linear for short term dependent noise, may be time consuming, and it may effect the performance of the overall algorithm, especially in terms of numerical condition. A bad numerical condition means that the transform is 'far from orthogonal', i.e. no Riesz-basis is guaranteed. This turns out to be a challenging problem in applications of smoothing.[33]

The whole of these considerations leads to a multiscale triangulation of the scattered data points. This paper describes the construction of such a triangulation, in which several measures are taken to stabilize the algorithm from the numerical point of view. We also explain how to compute the correlation matrix of the noise with linear complexity. This matrix can be used to renormalize the wavelet coefficients so that we have homoscedastic noise. Simple (soft-)thresholding is then possible to reduce the noise in the data.

Because of the irregularity of the data and the wavelet basis functions, the equilibrium between 'important' and 'noisy' wavelet coefficients is more delicate than in the regular grid case: all wavelets have a unique scale and adjacent coefficients may correspond to basis functions with considerable difference in support. Even if all coefficients are normalized according to their variance, thresholding may therefore not be sufficiently data-adaptive. Inspired by scale-dependent thresholding in the equidistant setting, this paper aims at an algorithm that takes the support area of the basis function into account.

This is done explicitly through a Bayesian argument. The idea is that the wider the support of a basis function, the more likely it is going to interfere with a data singularity, and hence, the more likely it is that the associated coefficient is important and should be preserved for reconstruction. This is filled in with an empirical approach, i.e.

Correspondence: Email: maarten.jansen@cs.kuleuven.ac.be; WWW: http://www.cs.kuleuven.ac.be/~maarten/; Telephone: ++32 16 32 7080; Fax: ++32 16 32 7996;

applying a marginal maximum likelihood estimator, the data themselves provide values for the parameters of the prior probability functions. We use a mixture prior model[1] with a mass probability at zero and a heavy tail.

This paper is built on the combination of two components: Section 2 explains second generation wavelets for irregularly spaced data, and Section 3 has the details of the Bayesian procedure. We end with experimental illustrations, conclusions and suggestions for further research.

2. WAVELET TRANSFORMS ON IRREGULAR GRIDS

Classical wavelet[26,22] transforms are constructed as repeated filter banks. The filter operations in these algorithms are time- and scale-invariant, i.e. all input data are processed the same way, regardless of their position in the input signal. This is because the approximation properties of the corresponding wavelet basis functions are designed through a Fourier analysis and the whole classical wavelet theory heavily relies on notions like convolutions and so on.

This approach no longer holds when the input data are on irregular grids. Correlations between close neighbours are different from those between data in regions with few observations. The filters should adapt themselves to the varying sample intervals. It is not trivial to extend the filter bank idea to the case of scattered data in a straightforward way. The extension is based on the notion of lifting, as explained below.

2.1. The Lifting scheme

In the first instance, the lifting scheme[30,27,28] is a decomposition of a filter bank into a set of elementary steps. The first step splits the input into even and odd coefficients. This corresponds to subsampling in the classical filter bank. Each of the following steps of this building process adds more properties — typically vanishing moments — to the corresponding wavelet basis functions, which explains the name *lifting scheme*. All classical wavelet filter banks can be factored into a sequence of lifting steps.[11]

As illustrated in Figure 1, lifting steps appear in two alternating types, called primal and dual lifting: the first affects the primal (synthesis) wavelet functions, the second type changes the dual (analysis) wavelet function. Primal lifting is also known as *update*, in contrast to *prediction* for dual lifting. This nomenclature comes from an

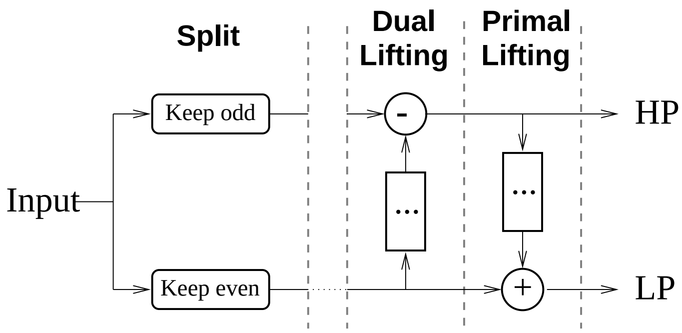


Figure 1. Decomposition of a filter bank into lifting steps. The first type of lifting is called *dual lifting* or a *prediction step*. The other type is *primal lifting* or *update*.

interpretation of the effect of the first dual lifting step: after the input splitting, the data in both branches are highly correlated. The idea is to *predict* data in one branch with a combination of data from the other one and to code the difference. These differences are small for smooth data, and will serve as wavelet coefficients. If, for instance, the prediction operator is an interpolating polynomial, these wavelet coefficients are exactly zero if the input comes from a polynomial of lower or equal degree. This illustrates that prediction indeed fixes the number of vanishing moments at the analysis, thereby motivating its name dual lifting step.

2.2. Second generation wavelets

The attractiveness the lifting scheme lies in its extendibility to non-classical settings. For instance, interpolating prediction has no need for equidistant data. Papers by Daubechies e.a.[10,9] (the second being an overview of the field) have interesting results on the smoothness of the (primal) basis functions associated with cubic interpolating prediction. The same papers also illustrate that ignoring the sample grid leads to a solution which is smooth on an equidistant grid. After *remapping* to the actual grid, however, the output may become arbitrarily non-smooth, since all the grid irregularities are reflected in the basis functions.

2.3. Second generation wavelets in two dimensions

Far less is known about smoothness in two dimensions. Efforts have been made to generalise the idea of commutation and (second order) divided differences,[17] but it is not clear what smoothness characterizes wavelets from general subdivision and lifting schemes.

Secondly, in 2-d, there is no obvious split of data into 'odds' and 'evens'. Neighbourhoods can no longer be defined in terms of 'previous' and 'next' point(s), since scattered data in 2-d cannot be ordered, unless in special cases, as for the tensor product wavelets in image processing.

A third problem in 2-d is the decreasing power of wavelet bases in sparse data representation. Wavelets are an excellent tool for dealing with point singularities in 1-d as well as in 2-d, but 2-d data typically contain line singularities, such as edges in images. Classical wavelet bases are not really able to catch these types of singularities. Unlike point singularities, lines cause an increasing number of large coefficients at finer scales. The development of better image representations is an important subject of ongoing research.[15,7,19]

This paper focuses on the second aspect. Although a general 2-d setting does not allow for a meaningful ordering of the data points, we can define a neighbourhood system using a triangulation of the input. Instead of splitting the input into two (nearly) equal groups, each step of our algorithm predicts only one input point and updates its immediate neighbours. At each step, the algorithm checks which input point corresponds to the scaling basis function with the smallest support and this point is selected for prediction: we want to process fine scale details first before proceeding to coarse scale phenomena. The prediction is a least square plane through the point's adjacent vertices and the update ensures one primal vanishing moment, i.e. it guarantees that the scaling coefficients at the next, coarser scale have the same average as the input scaling coefficients. Since most vertices have more neighbouring vertices than strictly required to fulfill this objective, we choose for a minimum norm solution. A general lifting scheme may add more dual and primal steps. Our algorithm is limited to one dual and one primal step. The motivation for these particular choices is further discussed in Section 2.5.

2.4. The wavelet transform used in our experiments

We now provide further details on our algorithm, i.e. we fill in the filter coefficients. Suppose at step (or 'scale' j), we decide to predict a scaling coefficient $s_{j+1,0}$ by a combination of its neighbours:

$$d_j = s_{j+1,0} - \sum_{k \in \partial_{j+1} 0} \beta_{j,k} s_{j+1,k}.$$

In this expression, $\beta_{j,k}$ are the prediction coefficients and $\partial_{j+1} 0$ stands for the set of points, adjacent to vertex 0 at scale $j+1$ (i.e. connected through an edge). Our approach computes the coefficients $\beta_{j,k}$ according to a linear least squares prediction.

This prediction also defines the scaling functions $\varphi_{j,k}$ in the multiresolution analysis. The two-scale equation, associated with the prediction can be found using a technique called *subdivision*.[32] In our case, this leads to

$$\varphi_{j,k} = \varphi_{j+1,k} + \beta_{j,k} \varphi_{j+1,0},$$

and obviously, the same relation holds for the integrals $I_{j,k}$ of those functions:

$$I_{j,k} = I_{j+1,k} + \beta_{j,k} I_{j+1,0}.$$

To ensure wavelet functions ψ_j with zero integral, we update the scaling coefficients $s_{j+1,k}$ in the neighbourhood of the now predicted vertex:

$$s_{j,k} = s_{j+1,k} + \alpha_{j,k} d_j, \qquad k \in \partial_{j+1} 0.$$

Again subdivision can be used to verify that the effect of this operation in terms of basis functions can be written as:

$$\psi_j = \varphi_{j+1,0} - \sum_{k \in \partial_{j+1} 0} \alpha_{j,k} \varphi_{j,k}.$$

At first sight, it may look a bit bizarre that the update step affects the wavelet functions by changing not the wavelet coefficients, but rather the scaling coefficients. The integral of the wavelet basis function is zero if:

$$I_{j+1,0} = \sum_{k \in \partial_{j+1} 0} \alpha_{j,k} I_{j,k}.$$

This condition leaves us with some degree of freedom for the update coefficients $\alpha_{j,k}$. Choosing a minimum norm solution leads to:

$$\alpha_{j,k} = \frac{I_{j,k} I_{j+1,0}}{\sum_{l \in \partial_{j+1} 0} I_{j,l}^2}.$$

2.5. A note on stability

The reason for the least squares prediction and minimum norm update is numerical condition. Second generation wavelet transforms are reported to show major problems with numerical stability when executed without additional precautions.[25,33,32] A high condition number means that a small error on the wavelet coefficients may lead to a serious error after reconstruction. Since regression algorithms manipulate or select wavelet coefficients, it becomes hard to predict what effect this operation has on the output. Small changes to coefficients may cause a considerable bias in the output and more specific, it is no longer guaranteed that small coefficients are nearly meaningless.

Numerical instability originates when large update coefficients create wavelet functions ψ_j that are close to neighbouring scaling functions $\phi_{j,k}$ at the same scale.[25,32] If all degrees of freedom for the update step are used to add more vanishing moments, the magnitude of these update coefficients heavily depends on the scaling function integrals and higher moments. As discussed before, those scaling functions in their turn are determined by the prediction (dual lifting) step. A further analysis[32] reveals that 'unbalanced' scaling functions (i.e. scaling functions with heavy side lobes) are particularly unfavourable. Such 'unbalanced' basis functions are created by a prediction operator which acts on different scales at one time. The grid in Figure 2(a) for instance has a wide 'gap' between two regions of dense samples. If the prediction operator uses four subsequent points with even index, it involves close neighbours near the gap together with larger scale phenomena at the gap itself. A simple linear polynomial prediction in 1-d avoids this scale mixing, since it uses only two immediate neighbours next to the predicted point. The span between those two neighbours is characterized by one scale only.

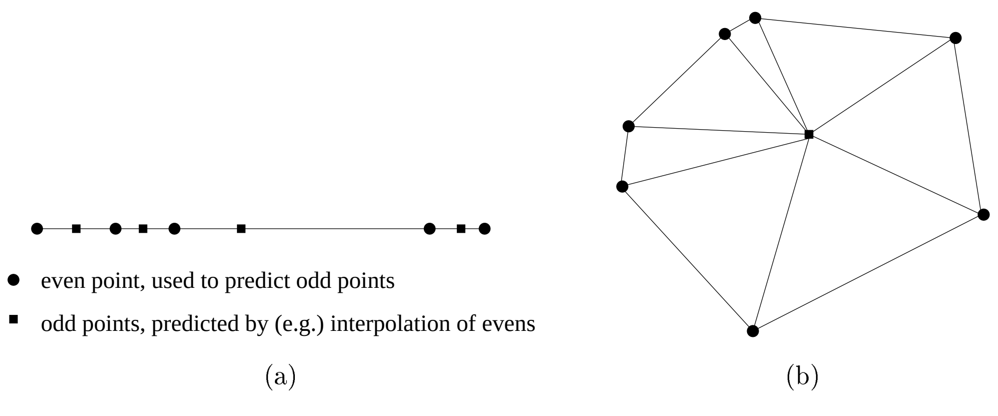


Figure 2. Sample configurations that could give rise to unstable estimation. If a single prediction step involves both a group of close samples and a wide gap, the transform becomes unstable. A prediction that uses immediate neighbours only is safe in 1-d (a), but not in 2-d (b).

In 2-d, the situation becomes less favourable. Even if the prediction of a given vertex only involves its immediate neighbours (i.e. vertices connected to the predicted point by a single edge), some of the neighbours might form a cluster of close samples as in Figure 2(b). Taking a least square plane as prediction operation reduces the unbalancing effect of close samples.

2.6. The wavelet coefficient variances

An additional problem, specific for smoothing algorithms comes from the noise itself: smoothing parameters — typically thresholds — depend on the noise variance. In a classical, stable setting, all coefficients within a scale correspond to basis functions with constant norm. Renormalizing one of the basis functions would be compensated by a change in coefficient variance. The higher the norm, the smaller the variance. In the possibly unstable second generation case, high basis norms and high variances may go together. Normalizing the basis functions towards homoscedastic noise then even further unbalances the importance of different coefficients. This is an *a priori* coefficient importance: it does not depend on the actual value, but on the basis function behind it. A threshold cannot take this background into account, it can only operate on the coefficient magnitude. Stable transforms correspond to a basis with more or less equally important elements.

Computation of coefficient variances can be done in linear complexity if the input correlation matrix Q is band limited, such as for uncorrelated noise. If $\tilde{W}$ is the forward wavelet transform matrix, then the correlation matrix S in wavelet domain satisfies:

$$S = \tilde{W} Q \tilde{W}^T.$$

This expression reads as a wavelet transform on all rows of Q, followed by a transform on all columns. For classical wavelets, this would correspond to a rectangular 2-d wavelet transform. More on the computation of the wavelet correlation matrix can be found in Refs. 21,31.

3. BAYESIAN SHRINKAGE

Now consider the following model of observations subject to noise:

$$Z_i = f(x_i, y_i) + \epsilon_i,$$

where the noise ϵ_i are independent $N(0, \sigma^2)$ random variables. The grid locations are irregular but fixed, i.e. they are not an instance from a probabilistic distribution. Wavelet based smoothing algorithms estimate $f(x, y)$ by taking an appropriate wavelet transform, modifying the coefficients in order to reduce noise, and finally inverse transforming the updated coefficients. The coefficient modification step is typically a shrinking procedure. Soft- and hard-thresholding are the best know examples but more sophisticated shrinking may follow (among others) from a Bayesian analysis of the noisy coefficients. Bayesian procedures permit priors where, for instance, the degree of sparsity may be specified.

3.1. Prior model and posterior density

We use the prior model described by Johnstone and Silverman.[20] Just like in other papers,[8,1] the idea of sparsity is expressed by a mixture prior model for the uncorrupted coefficients V_j:

$$V_j \sim (1 - p_j)\delta_0 + p_j \gamma_\sigma.$$

So, with probability $(1-p_j)$, $V_j = 0$. Unlike in other papers, the model we use, has a heavy tail in γ_σ. Johnstone and Silverman suggest (the not so heavy tail of) a double exponential (Laplacian) density, or the following continuous mixture density, which we use throughout our experiments: if $U \sim \gamma_\sigma$, then:

$$U|\Xi = \xi \sim N(0, (\xi^{-1} - 1)\sigma^2) \text{ with } \Xi \sim \text{Beta}(\tfrac{1}{2}, 1).$$

This is:

$$\gamma_\sigma(u) = \frac{|u|}{\sigma^2} e^{u^2/2\sigma^2} \left[\Phi_\sigma(|u|) - 1\right] + \frac{1}{\sqrt{2\pi}\sigma},$$

where ϕ_σ and Φ_σ are the zero mean normal density and distribution function respectively with standard deviation σ. Most important feature of this density $\gamma_\sigma(u)$ is its heavy, Cauchy-like tail:

$$\gamma_\sigma(u) \sim \frac{1}{\sqrt{2\pi}\sigma u^2}, u \to \infty.$$

The parameter σ is convenient for the computation of the noisy coefficient density: if $W_j = V_j + \omega_j$ and $\omega_j \sim N(0, \sigma^2)$, then

$$W_j \sim (1 - p_j)\phi_\sigma + p_j g_\sigma,$$

where

$$g_\sigma(w) = f_{W|V}(w|0)$$

is the convolution of γ_σ with a normal density:

$$g_\sigma(w) = (\gamma_\sigma * \phi_\sigma)(w) = \frac{\sigma}{\sqrt{2\pi}} \frac{1 - \mathrm{e}^{-w^2/2\sigma^2}}{w^2},$$

and

$$g_\sigma(0) = \frac{1}{2\sqrt{2\pi}\sigma}.$$

The posterior density for the unknown, uncorrupted wavelet coefficient is then:

$$\begin{aligned} f_{V|W}(v|w) &= f_{V|W}(v|w, V = 0)\, P(V = 0|W = w) \\ &\quad + f_{V|W}(v|w, V \neq 0)\, P(V \neq 0|W = w), \end{aligned}$$

with:

$$f_{V|W}(v|w, V = 0) = \delta_0(v),$$

obviously, and, according to Bayes' rule:

$$f_{V|W}(v|w, V \neq 0) = \frac{\gamma_\sigma(v)\phi_\sigma(w - v)}{g_\sigma(w)}.$$

Bayes' rule also completes the expression:

$$p^* := P(V \neq 0|W = w) = \frac{P(V \neq 0) f_W(w|V \neq 0)}{f_W(w)} = \frac{p g_\sigma(w)}{p g_\sigma(w) + (1 - p)\phi_\sigma(w)}.$$

3.2. Bayesian decision rule: posterior median

Once we have the expression for the posterior density $f_{V|W}$, we have to decide what to do with this information in order to obtain an estimate of the unobserved value of V. The posterior mean

$$\hat{v} = \mathrm{E}(V|W = w)$$

is a popular decision rule but it lacks the thresholding property: all coefficients in the output will be non-zero, thereby leaving noise in the reconstruction. An alternative is the posterior median, satisfying

$$F_{V|W}(\hat{v}) = \frac{1}{2}.$$

This requires an expression for the posterior (cumulative) distribution

$$F_{V|W}(v|w) = (1 - p^*) I(v > 0) + p^* \frac{\int_{-\infty}^{v} \gamma_\sigma(u)\phi_\sigma(w - u) du}{g_\sigma(u)}.$$

Call the integral in this expression:

$$h(v, w) = \int_{-\infty}^{v} \gamma_\sigma(u)\phi_\sigma(w - u) du.$$

Working this out for $\sigma = 1$, leads to:

1. for $v < 0$ and $w \neq 0$:

$$h(v,w) = \frac{1}{\sqrt{2\pi}} \left[\left(\frac{v}{w} - \frac{1}{w^2} \right) \mathrm{e}^{wv} \Phi_1(v) \mathrm{e}^{-w^2/2} + \frac{\Phi_1(v-w)}{w^2} + \frac{\phi_1(v-w)}{w} \right],$$

2. for $v > 0$ and $w \neq 0$:

$$h(v,w) = g(w) - h(-v,-w),$$

3. and for $w = 0$:

$$h(v,0) = \frac{1}{2\sqrt{2\pi}} \left[v^2 \Phi(v) + v\phi(v) + \Phi(v) \right].$$

3.3. Estimating the parameters (MLE)

The expression for the posterior density still has an unknown factor: $p_j = P(V_j = 0)$ the prior probability that an uncorrupted coefficient in our model is zero. This is typically a scale dependent parameter: the wider the basis function, the larger the probability that it hits one of the points of discontinuity, thereby giving rise to a non-zero coefficient.

In the case of a classical wavelet transform for equally spaced data, this p_j at scale j could relatively easily be estimated from the observed wavelet coefficients $w_{j,k}$ at scale j by maximising the *marginal log likelihood function*:

$$L(p) = \sum_{k=1}^{2^j} \log \left[(1-p)\phi_\sigma(w_{j,k}) + pg_\sigma(w_{j,k}) \right].$$

This likelihood maximisation only makes sense if we have enough data within scale j to perform a reliable estimation. Since the irregularly spaced data setting leads to basis functions with all different scales, we first have to group functions with nearby support into artificial resolution levels, and then estimate 'average' sparsity degrees p for these levels. We denote by $l(j) \in \mathbb{Z}$ the (artificial) level ψ_j belongs to and by $\mathcal{L}(j) = \{i | l(i) = l(j)\}$ the set of all coefficients in that level.

The goal is to do this in such a way that application of our procedure to classical wavelet coefficients reduces to the algorithm as described in Ref. 20. We define the finest level as the set of the first $N/2$ coefficients, where N is the number of observations. From then on, we start a new level and a new sparsity estimation every time the corresponding scaling function integral exceeds twice the initial integral of that level. If $\hat{p}_{l(j)}$ denotes this average estimator for the level coefficient w_j belongs to, we could estimate the p_j as:

$$\hat{p}_j = \hat{p}_{l(j)} \frac{I_{j+1,0}}{\overline{I}_{l(j)}},$$

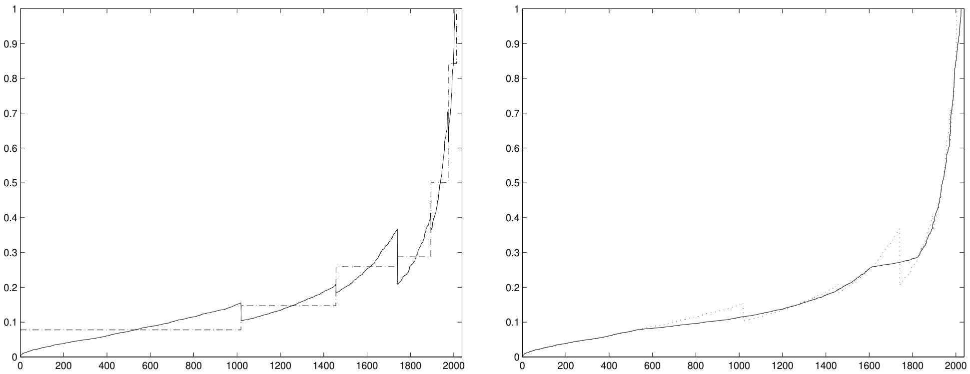
where $I_{j+1,0}$ is, as before, the integral of the scaling function in the last step before the corresponding vertex was eliminated, and

$$\overline{I}_{l(j)} = \frac{1}{\#\mathcal{L}(j)} \sum_{i \in \mathcal{L}(j)} I_{i+1,0},$$

the average value of these integrals over all coefficients at the given level. As mentioned before, this integral is a good measure for the corresponding wavelet support and hence for the probability of interference with any of the singularities in the underlying, noise-free data. Figure 3(a) however illustrates that this leads to unwanted, artificial discontinuities at the transitions between the self-defined resolution levels. To get rid of them, we interpolate between two successive average integrals and their corresponding average prior probabilities to obtain an estimate for each intermediate prior probability according to the corresponding scaling function integral, i.e. if $\overline{I}_{l(j)} \leq I_{j+1,0} \leq \overline{I}_{l(j)+1}$, then

$$\hat{p}_j = \frac{\left(I_{j+1,0} - \overline{I}_{l(j)}\right) \hat{p}_{l(j)+1} + \left(\overline{I}_{l(j)+1} - I_{j+1,0}\right) \hat{p}_{l(j)}}{\overline{I}_{l(j)+1} - \overline{I}_{l(j)}}.$$

Figure 3 (b) shows the effect. The p-values in this figure correspond to the wavelet coefficients in Figure 5.



(a) (b)

Figure 3. Estimation of the prior probability $p_j = P(V_j \neq 0)$ that the uncorrupted coefficient is non-zero, as a function of scale j (finest scales at the left). This estimation is based on an artificial construction of resolution levels. Those allow for a reliable estimation of the average p_j-value (a) Dotted line: average p_j-values. Solid line: correction for each coefficient separately, taking into account the support of its basis function. (b) Solid line: avoiding artificial discontinuities by interpolating on this support between the two neighbouring average supports.

4. RESULTS, ILLUSTRATIONS

We illustrate our algorithm with the example in Figure 4. The grid consists of (x, y) values that were randomly created on $[0,1] \times [0,1]$, using a uniform random generator and the function was constructed with the Matlab-command

```
f = (2*x+y).*((3*x-y)<1)+(10-x).*((3*x-y)>=1);
```

The corresponding wavelet coefficients appear in Figure 5. The estimated prior probabilities $\hat{p}_j$ are those in Figure 3. It is clear that the estimation of these prior probabilities must be based on a group of coefficients, a single coefficient is not sufficient to estimate its *a priori* importance. Figure 6 has the output both of a threshold algorithm, using a minimum mean squared error threshold, and of the proposed Bayesian algorithm. Thanks to the stabilizing operations in prediction and update, the simple threshold works fine as a first estimator, but the Bayesian procedure is clearly more subtle.

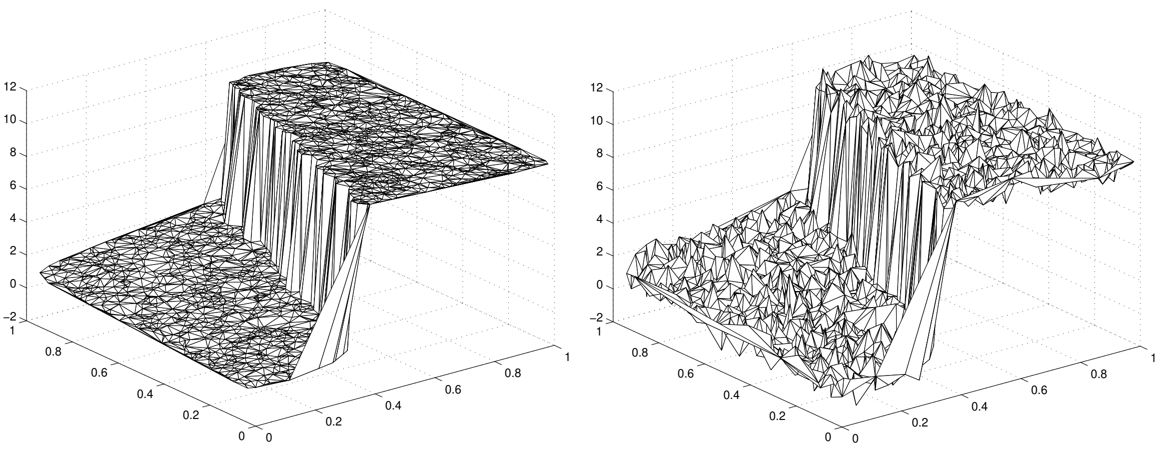
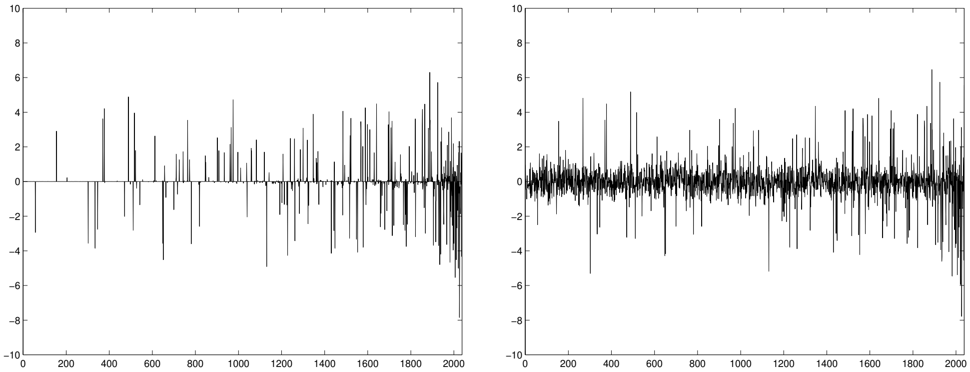


Figure 4. Test data in irregular grid with Delaunay-triangulation. (top) uncorrupted data. (bottom) Data with artificial uncorrelated Gaussian noise.



(a) (b)

Figure 5. (a)Wavelet coefficients (finest scales at the left) of data in Figure 4(a). (b) Wavelet coefficients (finest scales at the left) of data in Figure 4(b).

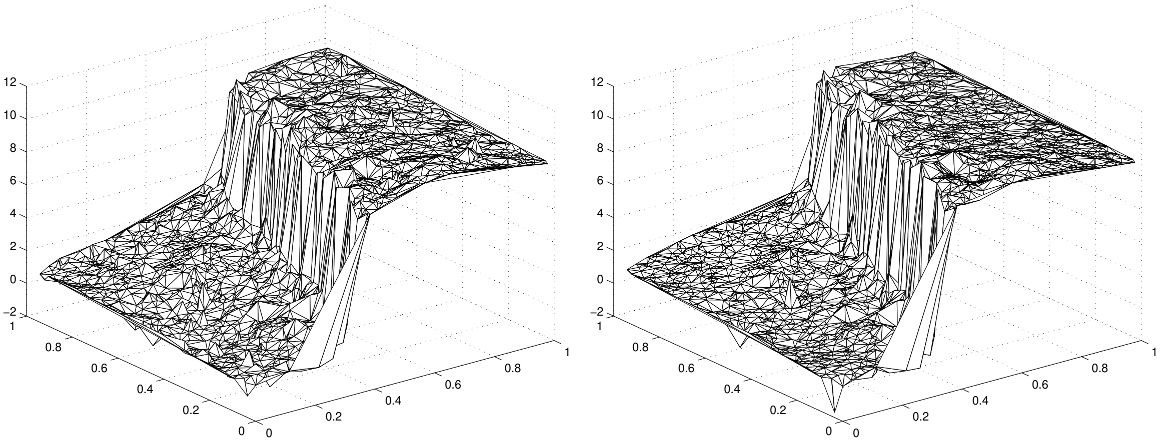


Figure 6. Reconstruction from shrunk wavelet coefficients. (top) Soft-thresholding, using a minimum mean squared error threshold. (bottom) proposed Bayesian model.

5. CONCLUSIONS AND FURTHER RESEARCH

This paper has investigated the application of second generation wavelets in scattered data smoothing. Second generation wavelets automatically take care of the sample positions. Experiments in this paper have illustrated that the second generation approach finds the appropriate answer to different structures of sample points. Equipped with a prior mixture model for uncorrupted wavelet coefficients and using marginal maximum likelihood estimation, a Bayesian framework provides a fully automatic shrinking algorithm which adapts itself to the data (i.e. to the noise and the sparsity of the uncorrupted coefficients).

Many questions in this application are still open, most of them implying theoretical problems. Second generation wavelets are built on the concept of subdivision. The question whether a subdivision scheme on a given grid converges and to which basis function is of course of primordial importance. Second, the smoothness of the limit function (the basis function) limits the smoothness of each data reconstruction based on this particular subdivision scheme. We conjecture that the basis functions in the scheme presented in this paper are continuous everywhere with possible discontinuities in the derivative at any grid point. That means that on an infinitely fine, irregular grid, the basis functions would be continuous everywhere but nowhere differentiable. The question arises how important this is from a practical point of view: scattered data live on a finite, fixed grid.

In computer graphics, subdivision is widely used as a method for surface generation.[16] This application is different in the sense that the resolution is specified by the user, not by the data. Surfaces should be smooth up to arbitrarily fine scales. The least square plane prediction from this paper is certainly not sufficient for those applications. In order to obtain C^1-smoothness, the prediction should include wider neighbourhoods than just the adjacent vertices as proposed in this paper. The prediction coefficients then typically follow from a variational approach, which minimizes some 'energy' (distortion) measure, for instance, second order divided differences.[17,10,9] These predictions however require more computations and especially in computing the coefficient covariance matrix, this might become a crucial consideration.

Another problem could be numerical condition. The remaining 'blips' in Figure 6 may be due to an instability in the transform. Again from the computer graphics world it is known that critical downsampling, such as in the classical, fast wavelet transform, gives no protection against high condition numbers. It is worth investigating what the use of alternative, overcomplete data representations could be in smoothing. An example is the generalization of Burt-Adelson pyramids,[5] as proposed by Daubechies e.a..[9,10]

ACKNOWLEDGMENTS

Maarten Jansen is postdoctoral research fellow of the Fund of Scientific Research Flanders (Belgium) (FWO). The main part of the research presented in this paper was done while he was with the University of Bristol. The authors gratefully acknowledge support of EPSRC grant GR/M10229 and Guy Nason gratefully acknowledges the support of EPSRC Advanced Fellowship AF/001664.

REFERENCES

1. F. Abramovich, F. Sapatinas, and B. W. Silverman. Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society, Series B*, 60:725–749, 1998.
2. A. Antoniadis, G. Grégoire, and W McKeague. Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.*, 89:1340–1353, 1994.
3. A. Antoniadis, G. Grégoire, and P. Vial. Random design wavelet curve smoothing. *Statistics and Probability Letters*, 35:225–232, 1997.
4. A. Antoniadis and D.-T. Pham. Wavelet regression for random or irregular design. *Computational Statistics and data analysis*, 28(4):333–369, 1998.
5. P. J. Burt and E. H. Adelson. Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, 31(4):532–540, 1983.
6. T. Cai and L.D. Brown. Wavelet shrinkage for nonequispaced samples. *Annals of Statistics*, 26(5):1783–1799, 1998.
7. E. J. Candes and D. L. Donoho. Curvelets - a surprisingly effective nonadaptive representation for objects with edges. Technical report, Department of Statistics, Stanford University, 2000.
8. M. Clyde, G. Parmigiani, and B. Vidakovic. Multiple shrinkage and subset selection in wavelets. *Biometrika*, 85:391–401, 1998.
9. I. Daubechies, I. Guskov, P. Schröder, and W. Sweldens. Wavelets on irregular point sets. *Phil. Trans. R. Soc. Lond. A*, To be published.
10. I. Daubechies, I. Guskov, and W. Sweldens. Regularity of irregular subdivision. *Constructive Approximation*, 15:381–426, 1999.
11. I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.*, 4(3):245–267, 1998.
12. V. Delouille, J. Franke, and R. von Sachs. Nonparametric stochastic regression with design-adapted wavelets. Technical report, Université Catholique de Louvain, 2000.
13. V. Delouille, J. Simoens, and R. von Sachs. Smooth design-adapted wavelets for nonparametric stochastic regression. Technical report, Université Catholique de Louvain, 2001.
14. B. Delyon and A. Juditsky. On the computation of wavelet coefficients. *J. of Approx. Theory*, 88:47–79, 1997.
15. D. L. Donoho. Wedgelets: Nearly minimax estimation of edges. *Annals of Statistics*, 27(3):859–897, 1999.
16. M. Eck, T. DeRose, T. Duchamp, H. Hoppe, M. Lounsbery, and W. Stuetzle. Multiresolution analysis of arbitrary meshes. In *Siggraph 95 Conference Proceedings*, pages 173–182, 1995.
17. I. Guskov. Multivariate subdivision schemes and divided differences. Technical report, Princeton University, 1998.

18. P. Hall and B. A. Turlach. Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Annals of Statistics*, 25(5):1912 – 1925, 1997.
19. M. Jansen, H. Choi, Sridhar L., and R. Baraniuk. Multiscale image processing using normal triangulated meshes. In *Proc. IEEE Int. Conf. on Image Proc. — ICIP '01*, Thessaloniki, Greece, 2001.
20. I. M. Johnstone and B. W. Silverman. Empirical bayes approaches to mixture problems and wavelet regression. Technical report, Stanford University, Bristol University, 2001.
21. A. Kovac and B. W. Silverman. Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Amer. Statist. Assoc.*, 95:172–183, 2000.
22. S. G. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 525 B Street, Suite 1900, San Diego, CA, 92101-4495, USA, 1998.
23. M. Pensky and B. Vidakovic. On non-equally spaced wavelet regression. Preprint, Duke University, Durham, NC, 1998.
24. S. Sardy, D.B. Percival, A.G. Bruce, H-Y. Gao, and W. Stuetzle. Wavelet de-noising for unequally spaced data. *Statistics and Computing*, 9:65–75, 1999.
25. J. Simoens and S. Vandewalle. On the stability of wavelet bases in the lifting scheme. TW Report 306, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, 2000.
26. G. Strang and T. Nguyen. *Wavelets and filter banks.* Wellesley-Cambridge Press, Box 812060, Wellesley MA 02181, fax 617-253-4358, 1996.
27. W. Sweldens. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In A. F. Laine and M. Unser, editors, *Wavelet Applications in Signal and Image Processing III*, pages 68–79. Proc. SPIE 2569, 1995.
28. W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Appl. Comput. Harmon. Anal.*, 3(2):186–200, 1996.
29. W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2):511–546, 1997.
30. W. Sweldens and P. Schröder. Building your own wavelets at home. In *Wavelets in Computer Graphics*, ACM SIGGRAPH Course Notes, pages 15–87. ACM, 1996.
31. M. Vannucci and F. Corradi. Covariance structure of wavelet coefficients: Theory and models in a bayesian perspective. *Journal of the Royal Statistical Society, Series B*, 61(4):971–986, 1999.
32. E. Vanraes, M. Jansen, and A. Bultheel. Stabilized lifting steps in noise reduction for non-equispaced samples. In M. A. Unser, A. Aldroubi, and Laine A. F., editors, *Wavelet Applications in Signal and Image Processing VII*, volume 4478 of *SPIE Proceedings*, July 2001.
33. E. Vanraes, M. Jansen, and A. Bultheel. Stabilizing wavelet transforms for non-equispaced data smoothing. TW Report 319, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, Feb 2001.