Wavelet shrinkage of itch response data

Robert C. Morgan
Unilever Research US Limited
45 River Road
Edgewater
NJ 07020
USA

and
G. P. Nason
Department of Mathematics
University of Bristol
University Walk, BRISTOL BS8 1TW England

May 25, 1999

Abstract

This article addresses the problem of denoising piecewise constant functions by using both an jump-tolerant moving average and Haar wavelet shrinkage. The piecewise constant functions are contaminated with Poisson-like noise and are measurements of perceived itch by human subjects in an experiment to relate perceived itch to bloodflow in response to histamine iontophoresis. We show that the translation-invariant wavelet transform with universal wavelet shrinkage provides a fast and automatic method for producing good estimates of the underlying perceived itch and that the method is superior to the jump-tolerant moving average and standard wavelet shrinkage. The article then looks in more detail at the noise structure of the signals and highlights some new wavelet techniques that might be useful.

Key words: wavelet shrinkage, Poisson noise, iontophoresis

Cet article adresse le problème de denoising des fonctions par morceaux constantes en utilisant un rétrécissement brancher-tolrant de moyenne mobile et de ondelette de Haar. Les fonctions par morceaux constantes sont souillées avec Poisson-comme le bruit et sont des mesures de perçu démangent par les sujets humains dans une expérience pour associer perçu

démangent au bloodflow en réponse à l'iontophoresis d'histamine. Nous prouvons que le ondelette traduction-invariable transforment avec le rétrécissement universel de ondelette fournit un rapide et la méthode automatique pour produire de bonnes évaluations du fondamental perçu démangent et que la méthode est supérieure à la moyenne mobile brancher-tolérante et au rétrécissement standard de ondelette. L'article regarde alors en plus détail la structure de bruit des signaux et met en valeur quelques nouvelles techniques de wavelet qui pourraient être utiles.

1 Introduction

This article addresses the problem of denoising piecewise constant functions. The piecewise constant functions in this article are recorded signals where the signal level registers a subject's perception of itch. Simultaneous measurements of bloodflow are taken and the eventual aim of the experimentation is to measure the extent to which perceived itch level is related to bloodflow. Once the relation is quantified future itch measurements can be more reliably and easily measured using bloodflow measurements.

However, the piecewise constant functions which register the perceived level of itch (controlled by a subject) are themselves measured in the presence of noise. The functions need to be denoised before they can be related to bloodflow measurements and this article addresses only this problem. Section 2 describes the experimental issues in more detail.

There are many ways in which piecewise constant functions can be denoised see for example Spokoiny [14] and references contained therein. We describe two here: one is based on adapting a moving average; the other is based on wavelet shrinkage using Haar wavelets. Section 3 shows that the jump-tolerant moving average method performs well, but only if its parameters are chosen manually and differently for each signal. Section 4 describes wavelet shrinkage using the discrete wavelet transform and shows how it performs well except for a noticeable Gibb's effect near jumps. Using Coifman and Donoho's TI-transform [2] the over- and undershooting effects almost completely disappear resulting in extremely good estimates. Section 5 looks in more detail at the noise that is present with the itch response data and makes suggestions about more advanced shrinkage methods that could be used.

To keep this article self-contained we have included a brief description of wavelets and wavelet shrinkage in Section 4. However, ours is a sparse description and more comprehensive expositions may be found in Donoho *et al.* [6] or in Nason and Silverman [13]. For more comprehensive discussions of wavelets see Daubechies [3], Meyer [12] or Burrus and Gopinath [1].

2 Itch response data

2.1 Introduction and background to data

In attempting to design personal products which give pleasing sensations to the skin, it is important to develop a detailed understanding of the underlying physiological and neurophysiological mechanisms which take place. One area of research that aims to address this is the measurement of blood flow in the skin, in the presence of certain chemicals. The chemicals can be used to produce various common sensations such as itching, stinging and burning, so that links may be drawn between the occurrence of these sensations and the behaviour of the blood-flow. The use of wavelets in helping to analyse the results of these experiments was investigated for the particular sensation of itch, but the procedure would be similar for any sensation.

2.2 Experimental details

The sensation of itch may be stimulated by the presence of histamine. In order that this is not confused with other sensations such as pain, it is necessary that the histamine is brought under the skin 'non-invasively'. In other words it should not be injected. The method used is known as *iontophoresis*. A chamber filled with a histamine solution is placed in contact with the skin. Inside the chamber are two electrodes, which are also in contact with the skin. When a voltage is applied across the electrodes, an electric current will pass through the skin, drawing histamine with it. Using this method it is possible to control the exact dose of histamine which enters the skin by altering the voltage. The chamber is approximately 3cm in diameter. As it is possible that the action of attaching the chamber may itself cause itch, a period of several minutes is allowed to elapse between attachment and application of voltage, in order for any such itch to settle down. It is accepted that after this period the static presence of the chamber on the skin will not cause itch. Further details about the methodology and applications of iontophoresis are given in Harris [7].

These experiments test the hypothesis that the level of itch which a subject feels is linked to the increase in blood flow around the skin area where the iontophoresis takes place. To test this hypothesis, the blood flow and subjective itch are measured simultaneously. The blood flow response signal or *flux* is measured using laser Doppler apparatus.

Subjective itch is measured by the subject moving a slider along a potentiometer. Two scales are used:

- Visual Intensity Scale (VIS) — a purely quantitative zero to one scale, on which zero represents no itch, and one represents the most intense itch imaginable by the subject. The subject is trained in the use of this scale and operates the potentiometer continuously themself.

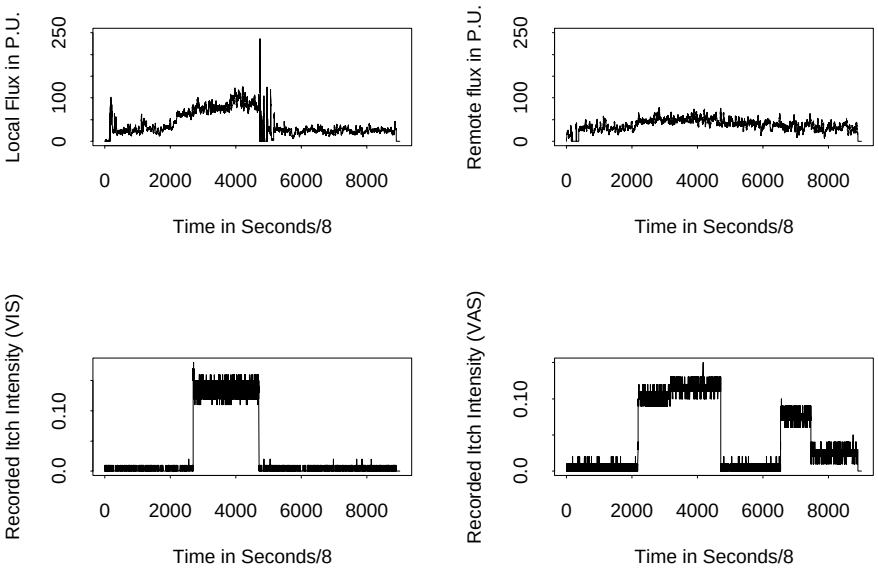

Figure 1: Complete output for subject 11.2. Top left: blood flow at iontophoresis site. Top right: blood flow at remote site. Bottom left: recorded itch intensity VIS scale. Bottom right: recorded itch intensity VAS scale. The scale is "Time in seconds/8" which means that $t = 800$ on the plot corresponds to 100 seconds from the beginning of the experiment. These units are chosen by the recording apparatus.

- Visual Affective Scale (VAS) — a five-point qualitative scale, with the five points representing descriptors such as annoying, unpleasant, painful. The investigator operates the potentiometer, updating the reading every minute by asking the panellist which of the descriptors best describes their present feeling. For presentation these descriptors are ordered and translated to points on a zero to one scale.

2.3 An example data set

Figure 1 shows an actual output from an iontophoresis experiment where the iontophoresis is begun at $t = 2000$. Each experiment produces a set of four curves all measured over time. Typically the flux (top line in figure) and subjective curves (bottom line) are very different. The flux usually increases steadily after iontophoresis, with a small amount of noise around the signal. The subjective curves consist of almost vertical jumps, separated by horizontal plateaux — they are approximately piecewise constant. There is a lot of very spiky noise around the signal of the subjective data, mostly due to the recording apparatus. One overall and eventual goal of our analyses is to link the flux data to the subjective data: before this can be done successfully the VIS and VAS curves need to be denoised. The next section considers two methods that might be appropriate.

3 Methods for curve estimation

There are very many methods that can be used to estimate curves within the presence of noise. Only a small proportion of these are suitable for curves where it is known *a priori* that the curves contain discontinuities, such as with our VAS and VIS curves above. To be of any use in our practical context our eventual chosen method has to be completely automatic and fast. Many methods were considered but since the curves seemed to be piecewise constant we ended up considering two methods: one based on an jump-tolerant moving average; the other based on wavelet shrinkage with Haar wavelets. Note there are still many other competitors that we could have used such as local polynomials.

3.1 Adapted moving averages

We devised the following moving average method which was designed to preserve large jumps and smooth out small ones. The moving average routine depends on two user-defined parameters: *smoothing width*, m, and *minimum jump size*, r, which are chosen subjectively. First define

$$U_{i,r} = \min_j\{j :\mid y_i - y_{i+j} \mid > r, j > 0\} - 1 \quad (1)$$

$$L_{i,r} = \min_j\{j :\mid y_i - y_{i-j} \mid > r, j > 0\} + 1. \quad (2)$$

Then our jump-tolerant moving average a_i is defined as

$$a_i = \frac{1}{[\min(m, U_{i,r}) - \max(-m, L_{i,r}) + 1]} \sum_{j=\max(-m,L_{i,r})}^{\min(m,U_{i,r})} y_{i+j}. \quad (3)$$

The jump-tolerant moving average is best described with the aid of diagrams.

Suppose $m = 5$. Then for a given y_i there are three possible ways in which a_i will be calculated. In general, the point y_i itself, the previous 5 points and the next 5 points are all used to compute a_i. If all of these points are inside the range $y_i \pm r$ (Case 1: see Figure 2) then they are all included in the average. If, however, y_{i+3} say, were to fall outside the range $y_i \pm r$ (Case 2: see Figure 3) then it would not be included in the average, and neither would any of the later points. These later points are excluded even if some of them fall inside the range $y_i \pm r$ (Case 3: see final point in Figure 3). (Note: If y_{i-3} was outside the range, then it would be excluded from the average along with all *earlier* points).

The smoothing width, m, and the minimum jump size, r, are chosen subjectively to obtain the best fit. The value of m determines the degree of smoothing along the plateaux, while r determines the minimum jump between two consecutive points that is considered as a step change rather than noise. When optimising m, the assumption that the underlying signal is approximately piecewise constant

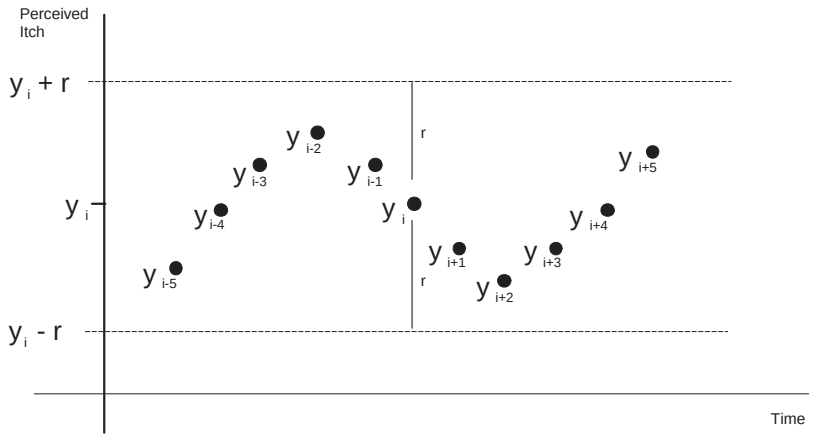

Figure 2: Moving average calculation. Case 1 (All points used to compute average)

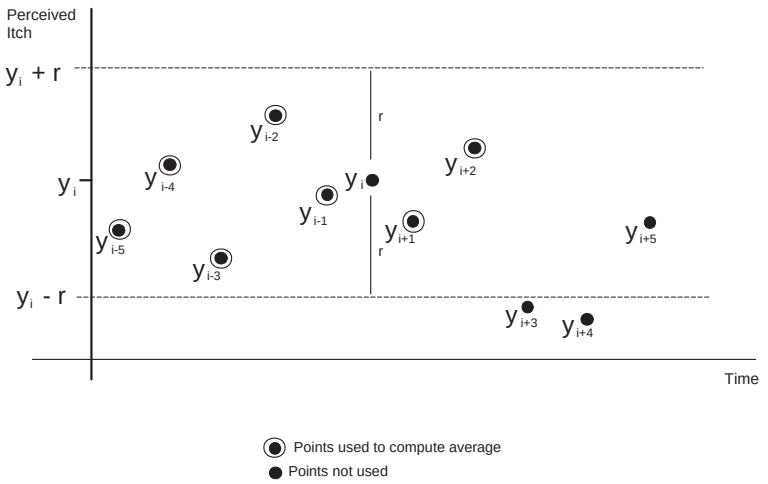

Figure 3: Moving average calculation. Cases 2 and 3

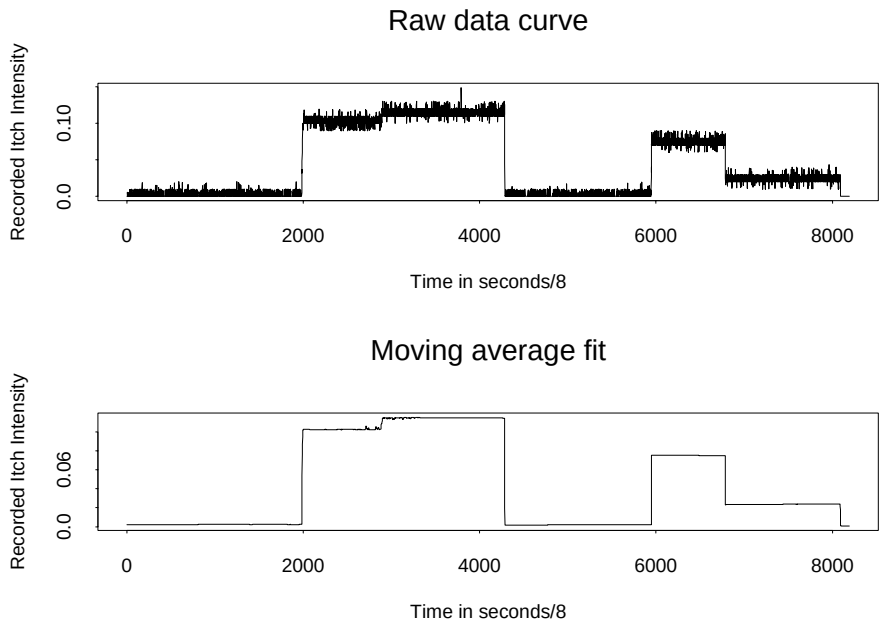

Figure 4: VIS curve for subject 11.2: Raw data and moving average

means that within each plateau it is impossible to over-smooth the data, suggesting that m may be chosen arbitrarily large. However in cases where the size of the noise is close to the size of discontinuities in the signal, a smaller value of m is advantageous as it reduces the smoothing out of these discontinuities. Also, the algorithm becomes slow if large values of m are chosen. Choosing r can be difficult: if it is too large then the routine smooths out discontinuities in the signal, whereas if it is too small then the routine interprets noise as real discontinuities. For a given signal it is usually possible to choose a suitable value of r by hand, by manually assessing the size of the smallest discontinuity and estimating the amplitude of the noise, then choosing r to be between these values. When the size of the noise is close to the size of the discontinuities it is not possible to choose an ideal value of r, and the effect of over-smoothing needs to be reduced by choosing a small value of m as mentioned above.

The moving average routine proved successful in smoothing the noise from the VIS curves whilst retaining the discontinuities in the signal. For example, see Figure 4.

By choosing separate values of m and r for each subject it would have been possible to achieve a very good fit in each case. However, this approach seemed too complex and subjective, so an attempt was made to choose values of m and r which could be used for all subjects. It was possible to choose values of m and r which gave reasonably good results for all subjects, although it took some work to find these. It would therefore be useful if the choice of these parameters could be made automatic which is not easy.

Furthermore, in cases where the amplitude of the noise was as large as the smallest discontinuity (determined by observing where the signal exhibited a step change rather than a short spike) the method suffered from the problem that there was no ideal value of r, and even reducing the value of m did not solve this entirely. To summarise, there were two problems with the moving average routine:

- not automatic — necessary to choose a suitable value of m and r for the sample of curves, and in order to achieve a very good fit, necessary to choose m and r separately for each curve. Would be advantageous to have an automatic method, i.e. one which did not require any subjectively chosen parameters.
- poor performance when the noise amplitude is greater than or equal to the smallest discontinuity in signal.

These two problems with the moving average method indicated that it would be worthwhile seeking an alternative "smoothing" method. The main requirement of this new method would be that it could perform as well automatically as the moving average method performed with manual parameter selection. If possible the method should be able to perform as well as the moving average routine with parameters chosen separately for each curve, but certainly it should be able to perform as well as the moving average routine with one set of parameters chosen for all the curves in the sample. It would be a bonus if the method was able to reconstruct signals in the presence of noise whose amplitude was of a similar size to the step changes.

4 Wavelet shrinkage

4.1 Introduction

Wavelet shrinkage is a beautifully simple method: the noisy signal is represented as a wavelet expansion. The wavelet coefficients are then shrunk and the resulting expansion forms an estimate of the true, underlying signal. We only give the barest details here. However, there are many papers describing the technique notably the seminal Donoho, Johnstone, Picard and Kerkyacharian [6] but see also Nason and Silverman [13] for a simple introduction.

In mathematical terms the wavelet shrinkage procedure goes as follows. We assume that noisy VIS or VAS signals are generated on the interval $[0, 1]$. Suppose that the true underlying function is given by $f(t)$ and that the noisy signal, $\{y_i\}_{i=1}^n$ is given by the signal plus noise model

$$y_i = f\left(\frac{i}{n}\right) + \epsilon_i \tag{4}$$

where $\{\epsilon_i\}_{i=1}^n$ is a independent and identically distributed sequence of Gaussian random variables with mean zero and common variance of σ^2. The basic wavelet shrinkage method also assumes that $n = 2^J$ for some integer J, in other words the number of data points has to be dyadic. Note also that model (4) only includes data that are equally-spaced on a fixed design. More recent research work has relaxed the restrictive design, number of data points and error distribution assumptions and some of these are described in section 5.

Wavelets. As mentioned above the underlying function $f(t)$ has a wavelet expansion. This is similar to a Fourier expansion or an expansion in terms of orthogonal polynomials. The main difference is that the basis functions are wavelets (small waves) and not sine/cosine or polynomials. The two basic building blocks in our expansion for the VIS/VAS curves are the Haar scaling function given by

$$\phi(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

and the Haar mother wavelet given by

$$\psi(x) = \begin{cases} 1 & x \in [0, \frac{1}{2}), \\ -1 & x \in [\frac{1}{2}, 1], \\ 0 & \text{otherwise} \end{cases}$$

The mother wavelet and scaling function are subjected to translation and scaling by factors of 2 (also called dilation). For example, we can simultaneously dilate the mother wavelet by a factor of 2^j and translate it by $2^{-j}k$ for integers j and k by forming

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k).$$

Functions can be built out of scaled and translated mother wavelets and translated scaling functions at a particular scale j_0. The following expansion is used:

$$f(x) = \sum_k c_k \phi_{j_0 k}(x) + \sum_{j \geq j_0} \sum_k d_{jk} \psi_{jk}(x), \tag{5}$$

where

$$\phi_{j_0 k}(x) = 2^{j_0/2}\phi(2^{j_0} x - k).$$

The equation is very suggestive and shows how the function $f(x)$ can be constructed out of: shifted versions of the scaling functions at j_0 — roughly speaking representing the overall level of the function over x; and scaled and shifted versions of the mother wavelets — roughly speaking representing the detail of the function at different scales. Expansion (5) is not restricted to Haar wavelets it works for many different classes of wavelets which sometimes have very special and useful properties: such as vanishing moments, symmetry, orthogonality. However, for our simple VAS/VIS functions Haar wavelets are ideal because they exactly emulate the underlying piecewise constant nature of the curves.

In all that follows assume that $j_0 = 0$. The quantity j_0 is often called the primary resolution as it determines the fineness on which the expansion is based. In real problems it is a parameter that needs to be chosen.

The discussion above describes the wavelet transform of a function $f(x)$. However, our data, given by model (4) is actually discrete. Fortunately, there is a simple discretized version of (5). The noisy data can be transformed into discrete wavelet coefficients d by

$$d = Wy \tag{6}$$

where $y = (y_1, y_2, \ldots, y_n)^T$ are our data, and d is a $n \times 1$ vector of wavelet coefficients (discretized versions of the c_k and d_{jk} given in (5)). The discrete wavelet coefficients in d have a special pyramidal structure: there is only one $c_.$ coefficient, one $d_0.$ coefficient, two $d_1.$, and so on. This means that the coefficients can be written as a pyramid as

$$\begin{pmatrix} c_0 \\ d_{00} \\ d_{10}, d_{11}, \\ d_{20}, d_{21}, d_{22}, d_{23} \\ \cdots \end{pmatrix} \tag{7}$$

Indeed, there is a fast algorithm due to Mallat [11] that computes the discrete wavelet transform using a "pyramid algorithm". The pyramid algorithm only requires $\mathcal{O}(n)$ operations in contrast to the matrix multiplication in (6) which requires $\mathcal{O}(n^2)$. The Haar wavelets are a family of orthonormal wavelets and consequently the transform matrix in (6) is orthogonal and hence the inverse matrix is merely W^T. Further, there is a fast, $\mathcal{O}(n)$ inverse algorithm as well.

Wavelet shrinkage The discrete wavelet transform of our $y = f + \epsilon$ noise model in (4) is

$$d = \theta + \epsilon,$$

where $\theta = Wf$. The noise ϵ in the wavelet domain is also iid Gaussian with the same variance as in the data domain because of the orthonormality of the transform W. The main reason for using wavelet shrinkage arises from the following heuristics:

1. the signal, f, (in this case the VIS and VAS curves) can be represented efficiently by few wavelet coefficients, θ, (in this case the VAS/VIS curves can be sparsely represented by Haar wavelets)

2. since the noise is iid it will be statistically the same for each coefficient. In other words, the noise does not show any preference for any particular coefficient or set of coefficients.

Therefore if the noise level is small enough then the wavelet transform of our VIS/VAS curves will consist of a few large coefficients (which come from the

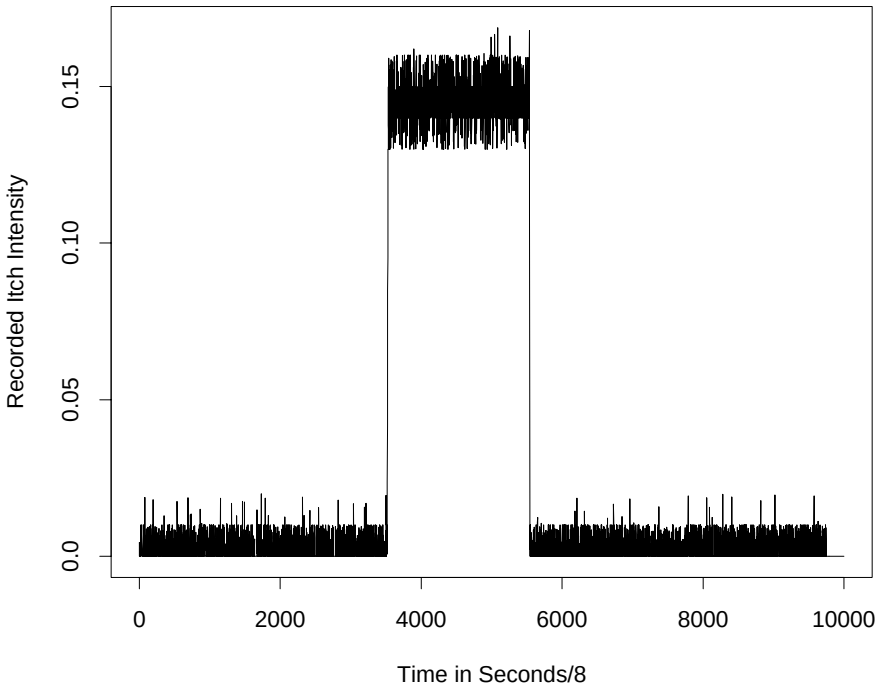

Figure 5: Raw VIS curve for subject 09.

signal plus noise) and many smaller coefficients (which correspond to just noise). An obvious way of recovering the true signal is to keep the large coefficients and remove any small ones. This technique is known as *hard thresholding* but there are many variants. The concepts of large and small in this case are taken with respect to the noise level σ which is estimated. In these examples we taken the finest scale wavelet coefficients and use the mean absolute deviation (MAD) estimator of σ divided by 0.6745 (this constant ensures a consistent estimator for Gaussian data). Many methods have been created to choose a threshold value at which smaller coefficients are removed. In this paper we only consider two: universal thresholding (or *VisuShrink*) and *SUREshrink*. Detailed descriptions of these two techniques can be found in Donoho and Johnstone [4] and [5] respectively.

After shrinkage the transform can be inverted producing an estimate of the true VIS or VAS curve.

4.2 A simple example

As an example of wavelet shrinkage we apply the technique to the data presented in Figure 5 which contains the VIS curve for subject 09. VIS curve for subject 09 is simple as it clearly only contains one hump. The curve contains 8192 points and was recorded over about 20 minutes. Curve 09's simplicity is useful because if our methods work well on this simple curve then they should also work well on more complicated examples.

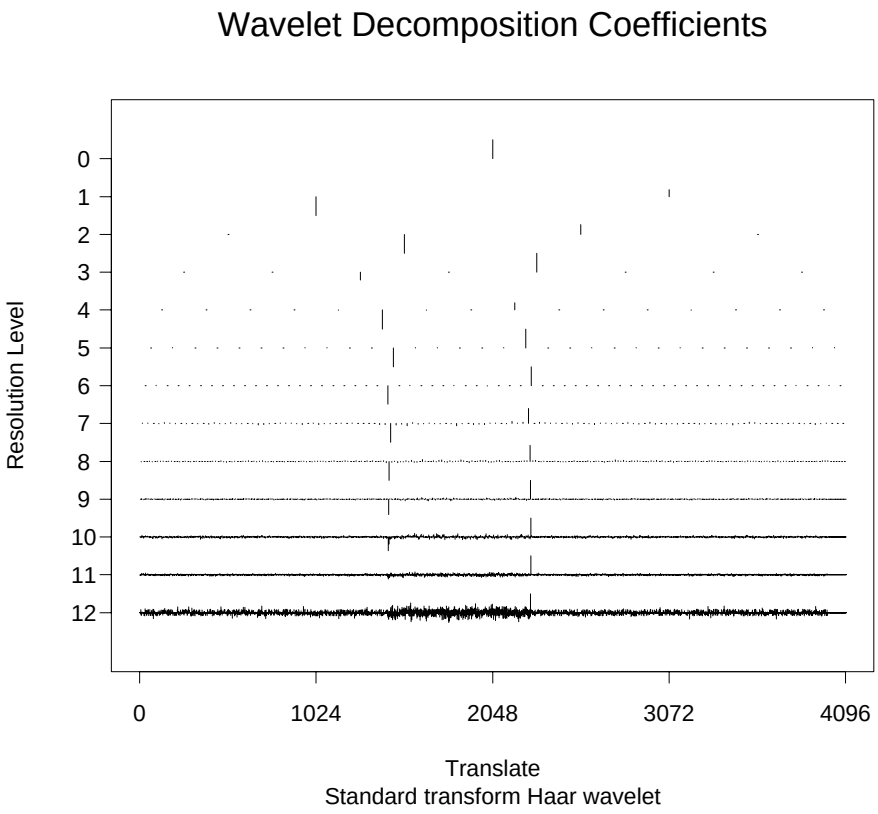

Figure 6: Haar wavelet coefficients for subject 09's VIS curve. Each level in this plot has been magnified to its maximum extent so that you can see the structure. If all coefficients were to be plotted to the same scale then you would not see the higher resolution coefficients — they would be too small.

Figure 6 shows a plot of the coefficients from a discrete wavelet transform of subject 09's VIS curve. The plot shows the pyramidal structure of the wavelet coefficients: each small vertical line in the plot corresponds to a discrete wavelet coefficient. The horizontal axis in this plot corresponds exactly to the horizontal time axis in Figure 5 (although they are labelled differently). The coefficient at the top (resolution level 0) is d_{00}, those at resolution level 1 are d_{10} and d_{11}, and so on just like the pyramid in (7), except we have omitted to plot c_0. The coefficients at the top are coefficients of large-scale wavelets and those at the bottom are coefficients of the small-scale wavelets. Notice how there are many more small-scale wavelets — basically because you can fit more in at that resolution level and retain orthogonality between wavelets.

Our heuristics above are illustrated very nicely by Figure 6. The underlying signal (the large hump) is exhibited in the large coefficients at horizontal position 1450 and 2270 at resolution levels from 2 upwards. The "hump" is sparsely represented by these coefficients. The noise has actually spread out over all

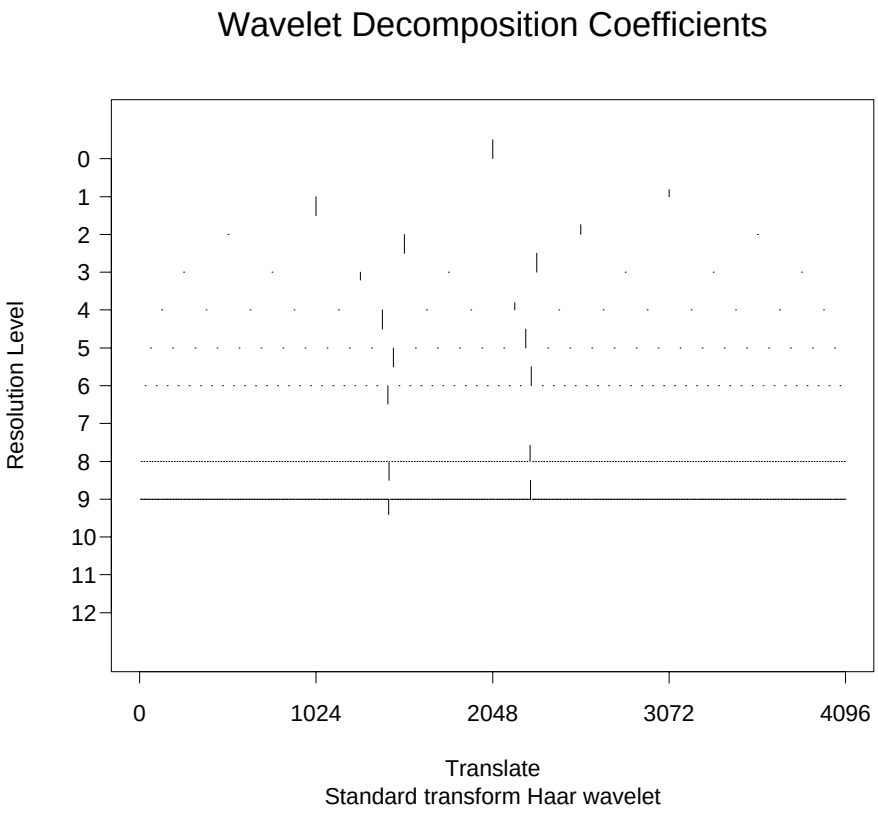

Figure 7: Haar wavelet coefficients after hard thresholding using a universal threshold. Each level in this plot has been magnified to its maximum extent so that you can see the structure.

coefficients fairly evenly. However, since the higher resolution levels have been magnified (so that you can see them) the noise is clearest at resolution level 12).

Now we apply hard thresholding estimating the noise by using the robust MAD-based estimate. Using universal thresholding we obtain a threshold value of approximately 0.136. Any coefficient smaller than this value in absolute magnitude is removed — this results in a tableau of coefficients depicted in Figure 7. Notice how the procedure has successfully removed most of the noise. Finally, to obtain our estimate we invert the hard thresholded wavelet coefficients: the estimate is shown in Figure 8. Note that the estimate is very good apart from around the jump: here the estimate overshoots both at the jump up and jump down. The next section describes how we used a more sophisticated wavelet algorithm to produce a better estimate.

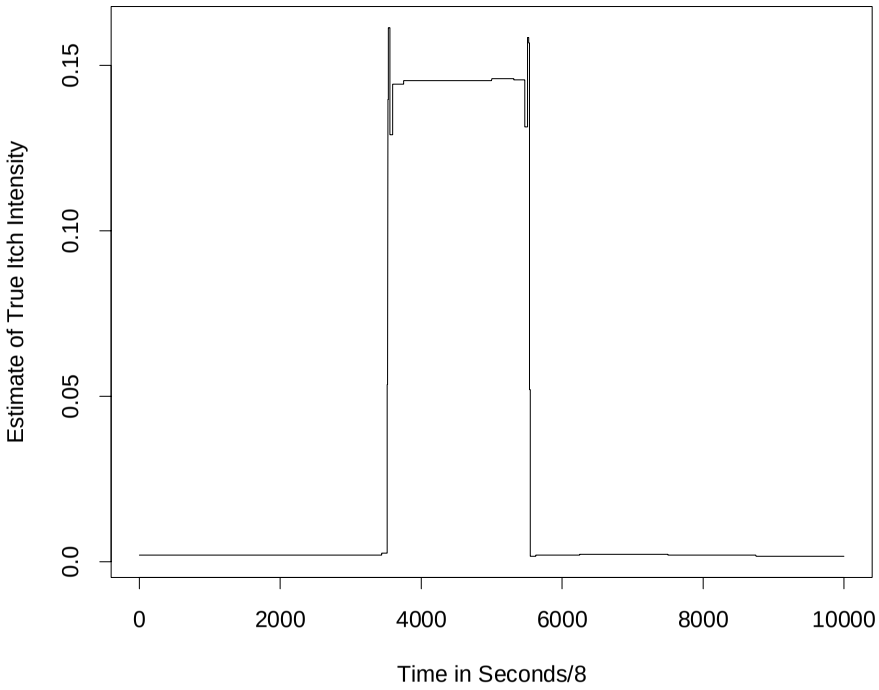

Figure 8: Estimate produced by inversion of hard thresholded coefficients in Figure 7.

4.3 Using the TI-transform

The overshoot illustrated in Figure 8 is a good practical example of Gibb's effect at work. A number of things were tried to improve the estimation:

- we tried to change the underlying wavelet from a Haar wavelet. However, none of these wavelets gave an estimate that was as good as with Haar thus confirming our initial guess that Haar wavelets, since they match the underlying structure of the signal, are appropriate for the itch data.
- we tried changing the threshold value manually. Occasionally we obtained a "perfect" reconstruction without overshoot, although most of the time we either obtained a bad estimate and/or overshoot. However, this manual fiddling is not automatic and thus, for large collections of itch data would be too time consuming.
- we tried changing the method of thresholding from hard thresholding. If anything, the results were worse.

However, since the problem seemed to be a Gibb's effect (or under/overshooting on discontinuities) we turned to the work of Coifman and Donoho [2] who explained that Gibbs effects were probably due to the discontinuities being located between wavelets (remember wavelets are aligned to an nested dyadic grid). They

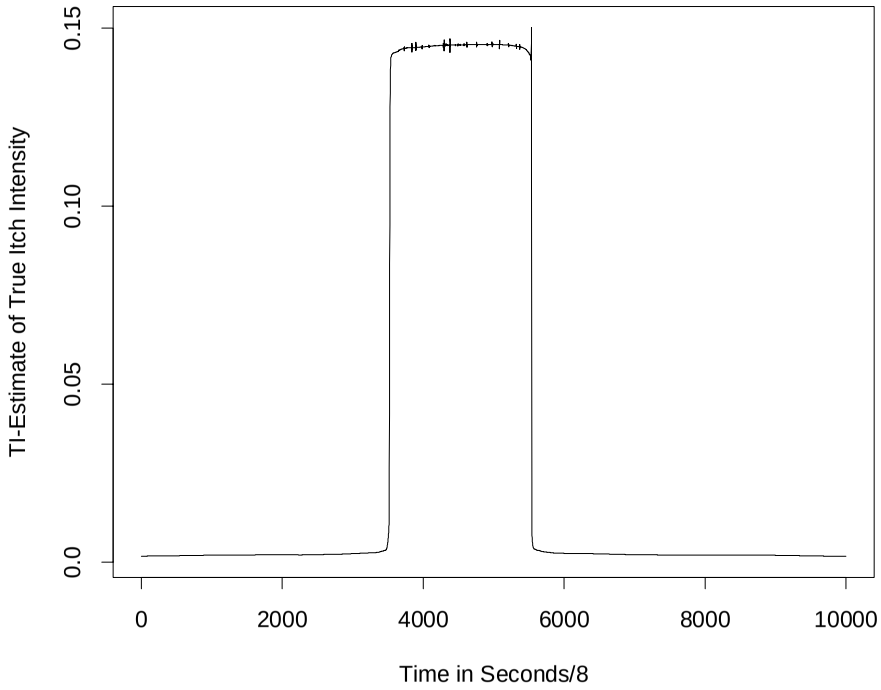

Figure 9: Estimate of the true itch intensity of subject 09 produced by the TI-transform.

suggested the simple procedure of *CycleSpinning* which produces wavelet shrinkage estimates by producing an average estimate of several shifted versions of the above wavelet shrunk estimate. The idea behind this is that some of the shifted versions will be perfectly aligned with the discontinuities producing a better overall estimate. An extension of this idea is to produce all possible shifted estimates to produce a translation-invariant estimate. The good news is that the wavelet algorithm underlying translation-invariant estimate, called the TI-transform, can be produced rapidly in $\mathcal{O}(n \log n)$ operations — hardly much more than the ordinary discrete wavelet transform.

The curve estimation procedure with the *TI-transform* is similar to that of the wavelet shrinkage method as above: transform the noisy data using the TI-transform, shrink the coefficients, and then perform an "inverse" transform. There are actually many different "inverse" transforms: one can invert any particular shifted set of coefficients, or invert an average of some or all of them. We adopt the "average all shifts" inversion here. The thresholding methods described above can still be used here. Figure 9 shows an estimate produced by the TI-transform method. It can be seen that the overshooting effect is considerably reduced.

Figure 9 illustrates another interesting feature about this set of itch data: notice that the TI-estimate on the "plateau" is still a bit noisy whereas the estimate at the "zero-level" is very smooth. All the wavelet methods above were

somewhat inappropriately used because the noise is not homogeneous and the thresholding methods above assume stationary noise. The noise is composed of two different types and wavelet methods that can be applied are described in the next section.

5 A closer look at the noise

The noise structure was examined in closer detail by looking at small sections of the VIS curve from subject 09. These are shown in Figure 10 and suggest that the noise behaves in two distinct patterns depending on whether the underlying signal is zero or positive.

First of all the data with zero signal was investigated. A plot of the first thousand points is shown at the top of Figure 10. The noise appears to exist at three distinct amplitudes:

- 0.01 itch units - these are very frequent
- 0.02 - around 1 of these spikes for every 100 data points
- 0.03 - only 1 in the whole data set

The data with positive signal was then investigated. On viewing a sequence of 1000 points from this section of the signal (bottom plot in Figure 10) it appeared that the signal occured at a level of 0.145, and that there was noise at the following amplitudes:

- 0.005 - very frequently.
- 0.015
- 0.025

On closer inspection it was revealed that the situation was slightly more complex. The signal seemed to oscillate between the values 0.14 and 0.15. As well as this oscillation there was noise of amplitude 0.01 in both directions and also noise of amplitude 0.02, but this did not occur frequently and only occured in the positive direction.

One possible explanation for the various amplitudes was that they were due to human induced noise (i.e. hand shaking while operating the potentiometer) and machine induced noise. However, human noise is unlikely to have a constant amplitude so it could be due to different types of machine induced noise, for example different amplitudes of feedback in the electrical circuitry. These possible explanations were investigated further by recording the VIS curve without a subject operating the potentiometer and it was found that machine noise was responsible.

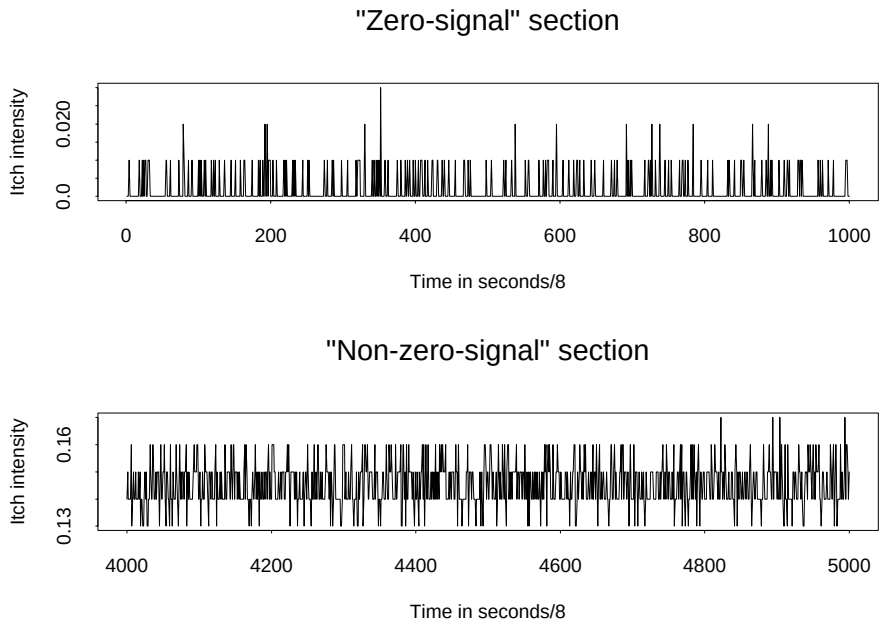

Figure 10: V09 Non-interpolated: Two different sections of 1000 points.

The added noise is clearly discrete in nature and it is possible that the Poisson distribution (or some close variant) might be suitable for modelling it. Indeed, most wavelet shrinkage models (including the ones we used above) assume additive normally distributed errors which is certainly not the case here (and the models do not assume a discrete process whose intensity might be piecewise constant). Kolaczyk [9] provides a more detailed description of what to do in certain cases of Poisson noise, although this extension is beyond the scope of the present paper. Indeed, empirically the variance of the noise does not increase with the mean so strict Poisson noise is unlikely. However, since we know the amplitude and form of the noise it is quite easy to choose a threshold value. Since we are using Haar wavelets we can easily compute the size of the largest Haar wavelet coefficient due purely to the noise (since we know the size of the smallest noise spike). Choosing the threshold like this results in a threshold of 0.022. This produces an acceptable estimate, although a few noise spikes are erroneously retained. Figure 11 shows a denoised version of the first 8192 observations of VAS recording in the bottom right of Figure 1 — the denoising was carried out using the TI-transform, with universal thresholding and using the MAD-based estimate of noise level estimation. This shows that although the noise structure is not normally distributed the TI-method for normal noise is reasonably robust.

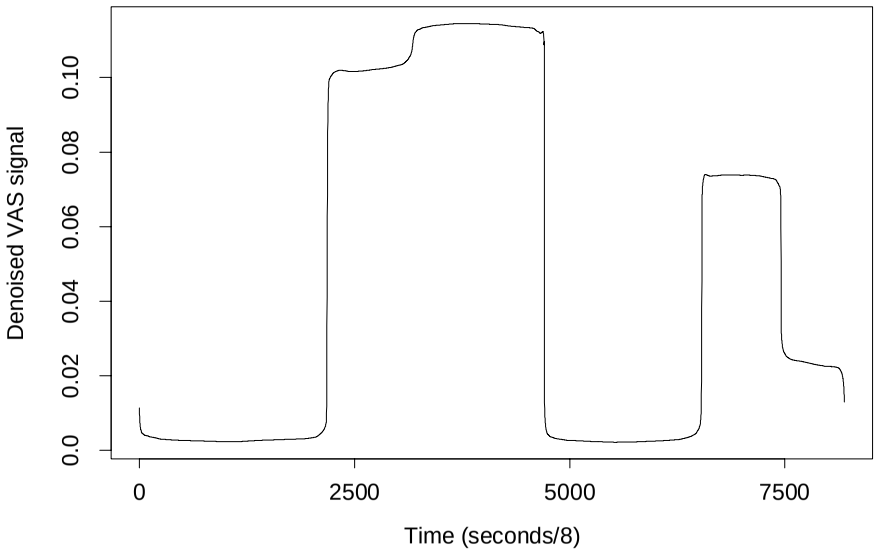

Figure 11: Denoised version of Subject 11's VAS recording. The TI-transform with universal thresholding using a MAD-based estimate of the noise level was used. The resulting threshold was 0.031.

6 Conclusions and further work

In this paper we have demonstrated how itch sensitivity data may be denoised. We have used two different methods: jump-tolerant moving averages and Haar wavelet shrinkage. Our main objective was to find a fast automatic method to denoise itch sensitivity curves — use of the TI-transform with universal thresholding worked extremely well and just as well as the jump-tolerant moving averages with a manual choice of moving average parameters.

Further investigation of the noise structure revealed that it was discrete in nature and that classical wavelet shrinkage methods are, in theory, inappropriate (although, the one we finally selected appears to work well). It may be the case that models using Poisson distributed noise may be just as, if not more, successful.

Our analyses above assume that the noise is i.i.d. It may be appropriate to extend our analyses to to incorporate Poisson-like error distributions and also more advanced thresholding techniques such as: choosing different thresholds for different resolution levels (for correlated data), see Johnstone and Silverman [8]; applying the methods of Kovac [10] which extend the wavelet shrinkage method to datasets with arbitrary design and number of points whilst retaining the speed of computation.

7 Acknowledgments

We would like to thank Francis McGlone for supplying the data used in this article. We would like to thank three referees and the editorial team for helpful comments on this article.

References

[1] S.L. Burrus and R.A. Gopinath. *Wavelets: introduction to wavelets and wavelet transforms.* Prentice Hall, Englewood Cliffs, NJ, 1997.

[2] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics: Proceedings of the XVth Rencontres Franco-Belges des Statisticiens.*, volume 103 of *Lecture Notes in Statistics*. Springer-Verlag, 1995.

[3] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.

[4] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–55, 1994.

[5] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, 90:1200–1224, 1995.

[6] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc. B*, 57:301–337, 1995.

[7] R. Harris. Iontophoresis. In S. Licht, editor, *Therapeutic Electricity and Ultraviolet Radiation*, pages 156–178. Elizabeth Licht, New Haven, 1967.

[8] I.M. Johnstone and B.W. Silverman. Wavelet threshold estimators for data with correlated noise. *J. R. Statist. Soc. B*, 59:319–351, 1997.

[9] E. Kolaczyk. Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds, 1998. (submitted for publication).

[10] A. Kovac. *Wavelet thresholding for unequally spaced data.* PhD thesis, Department of Mathematics, University of Bristol, Bristol, UK, 1998.

[11] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and MAchine Intelligence*, 11(7):674–693, 1989.

[12] Y. Meyer. *Wavelets and Operators.* Cambridge University Press, Cambridge, 1992.

[13] G. P. Nason and B. W. Silverman. The discrete wavelet transform in s. *Journal of Computational and Graphical Statistics*, 3:163–91, 1994.

[14] V.G. Spokoiny. Image denoising: pointwise adaptive approach. Technical report, Weierstrass Institute for Applied Analysis and Stochastics, 1998.