myjournal manuscript No.
(will be inserted by the editor)

Robust nonparametric regression and modality

A. Kovac[1]*

Universität Essen, Fachbereich Mathematik, 45117 Essen, Germany

Received: date / Revised version: date

Abstract The paper considers the problem of nonparametric regression with emphasis on controlling the number of local extremes and on resistance against patches of outliers. The robust taut string method is introduced and robustness properties are discussed. An automatic procedure is described.

1 Non-parametric regression and modality

Non-parametric regression is a topic of much current interest in statistics. The usual description of this problem is to estimate a function f on the basis of observations $y_1, \ldots, y_n$ at time points $t_1, \ldots, t_n$ where

$$y_i = f(t_i) + \varepsilon_i \tag{1}$$

and the $\varepsilon_1, \ldots, \varepsilon_n$ are noise. We describe our task somewhat differently and aim to specify a simple function $\tilde{f}$ such that the residuals $r_i = y_i - \tilde{f}_i$ look like noise.

Send offprint requests to:

* Research supported in part by Sonderforschungsbereich 475, University of Dortmund

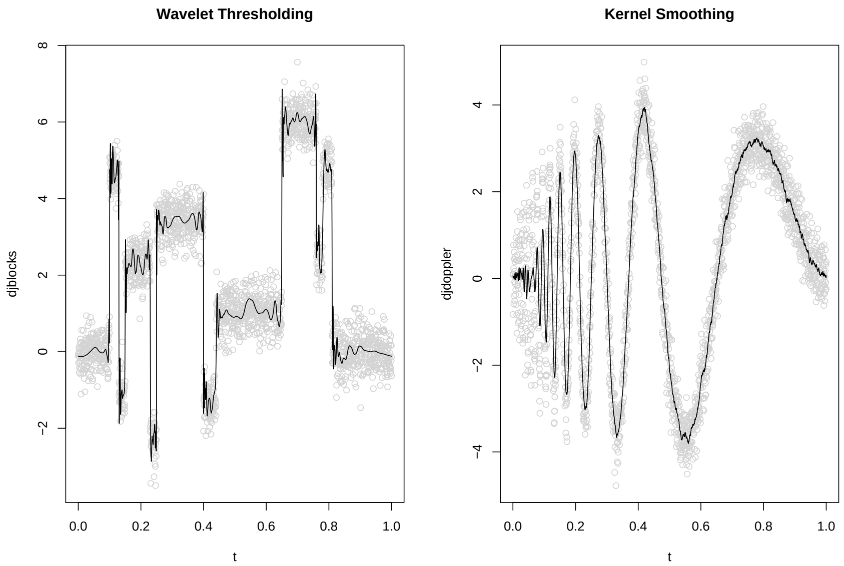

Fig. 1 Application of two classical smoothing methods to simulated data. The left panel shows Gibbs effects that occur often when wavelet thresholding is applied to functions with discontinuities. In the right panel a kernel estimator does not adapt locally to different degrees of smoothness.

Throughout this paper simplicity will be measured by the number of local extreme values whereas noise will be characterised by sums of the signs of the residuals r_i on different scales and positions.

Numerous methods have been suggested including kernel estimators (Nadaraya, 1964, Watson, 1964), spline smoothing (Silverman, 1985), local polynomial fitting (Fan and Gijbels, 1996) and wavelet thresholding (Donoho et al., 1995). However, none of these methods controls the number of local extremes, and as a result they all tend to produce superfluous extremes or exhibit artifacts like Gibbs effects near discontinuities (see Figure 1). Several procedures have been proposed

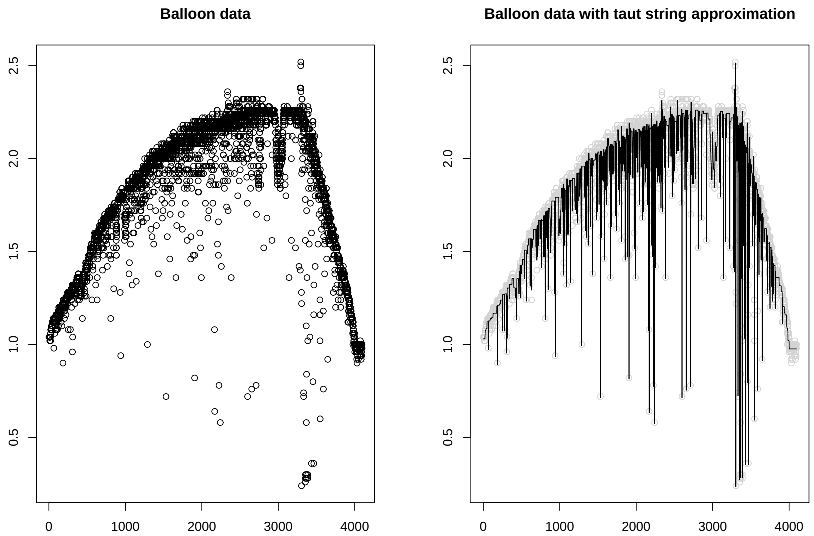

Fig. 2 Radiation measurements of a weather balloon. The large outlier patches were caused by a rope which cut the measuring device occasionally from the direct sun light. The left panel shows the data set, the right panel the data set and the usual taut string approximation.

to decide for each local extreme whether it comes from the data or is an artifact (Chaudhuri and Marron, 1997).

Davies (1995) and Mammen and van de Geer (1997) introduce another method which employs a taut string constrained to lie in a tube centered around the integrated data. Its derivative is used as a piecewise constant approximation for the given data and the value on each constant interval between local extrema is the mean of the observations that fall in this interval.

Davies and Kovac (2001b) developed the taut string method further by using an automatic procedure to determine a locally adaptive tube width. They show that their procedure attains asymptotically the correct modality and a similar result holds for the related problem of density estimation (Davies and Kovac, 2001a). The taut string approximation to the data can be regarded as the solution of the minimization problem

$$\sum_{i=1}^{n}(y_i - f(t_i))^2 + \sum_{i=2}^{n} \lambda_i |f(t_i) - f(t_{i-1})| = \min . \tag{2}$$

with data-driven λ_i.

2 Robust smoothing under control of modality

2.1 The taut string method and robustness

Figure 2 shows in its left panel a data set where all the methods discussed so far fail. The data are described by Davies and Gather (1993) and they are taken from a balloon which took measurements of radiation from the sun. It happened occasionally that the measurement device was cut off from the sun causing large outlier patches, the longest ranging over nearly 40 observations. The right panel shows the taut string approximation which is very sensitive to the outliers in the data. The taut string method is able to detect and retain very small sharp peaks which is a useful property for certain kind of data, for example data from NMR spectroscopy. In other settings like the radiation data a greater resistance against outliers is desirable.

2.2 A robust taut string version

We consider and discuss below a method similar to the taut string method, but where the L^2-norm in (2) is replaced by the L^1-norm and where we restrict ourselves to a global parameter λ:

$$\sum_{i=1}^{n} |y_i - f(t_i)| + \lambda \sum_{i=2}^{n} |f(t_i) - f(t_{i-1})| = \min . \tag{3}$$

Solutions of this minimization problem are again piecewise constant. On intervals which are not local extremes the value of the function is the median of the observations in that interval. If a solution takes a local extremum on an interval $[t_k, t_l]$ with $1 < k < l < n$, then it is the $\max(0.5 - \lambda/(l-k+1), 0)$-quantile in case of a local maximum and the $\min(0.5+\lambda/(l-k+1), 1)$-quantile in the case of a local minimum. At the left and right edges of the data set the solutions take either the $\max(0.5 - \lambda/2(l-k+1), 0)$-quantile or the $\min(0.5 + \lambda/2(l-k+1), 1)$-quantile, depending again on whether it is a local maximum or minimum.

Theoretically a solution of the minimization problem could be computed using L^1-programming. However, using standard routines (Barrodale and Roberts, 1973). we were not able to calculate a L^1-solution for data sets larger than 1200 because of large memory consumptions.

Davies and Kovac (2001b) give an algorithm of order $O(n)$ that calculates the solution of the minimization problem (2). Their algorithm can be adapted to the L^1-functional (3). The necessary calculations can be carried out in a way such that the computational complexity of the method is of order $O(n \log(n)^2)$.

For residuals r related to some approximation $\tilde{f}$ and $1 \leq k \leq n$ we investigate the sum of the signs of the first k residuals

$$\mathrm{SS}(k; r) = \sum_{i=1}^{k} \mathrm{sign}(r_i).$$

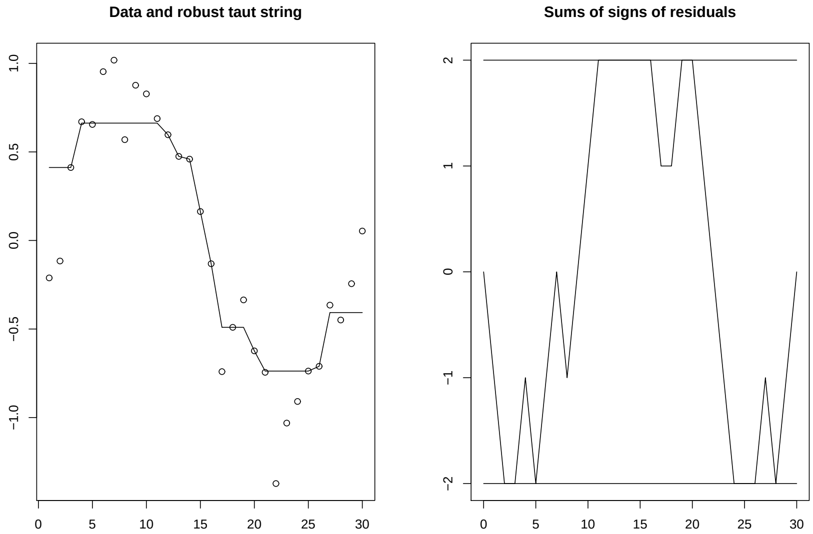

Fig. 3 The robust taut string method and the cumulative sums of the signs of the residuals. The left panel shows 30 points of a sine curve disturbed by normal noise and the robust taut string approximation with $\lambda = 2$. The right panel shows the cumulative sums of the signs of the residuals.

Maximizing $|\,\mathrm{SS}(k;r)|$ over k yields a distance measure that may be used to define a λ-neighborhood $N(y;\lambda)$ of the data:

$$N(y;\lambda) = \{g \in \mathbb{R}^n : \max_{1\leq k\leq n} |\,\mathrm{SS}(k; y-g)| \leq \lambda\}$$

The procedure described above can be carried out in a way such that $\tilde{f}$ satisfies the following properties:

Theorem 1 *For given data $y_1, \ldots, y_n$ and given integer λ there exists a solution $\tilde{f}_\lambda$ of the minimization problem (3) such that:*

1. For all $k \in \{1, \ldots, n\}$:

$$|\,\mathrm{SS}(k; y - \tilde{f}_\lambda)| \leq \lambda$$

2. If $k < n$ *such that* $\tilde{f}_\lambda(t_k) < \tilde{f}_\lambda(t_{k+1})$*, then*

$$\mathrm{SS}(k; y - \tilde{f}) = \lambda.$$

3. If $k < n$ *such that* $\tilde{f}_\lambda(t_k) > \tilde{f}_\lambda(t_{k+1})$*, then*

$$\mathrm{SS}(k; y - \tilde{f}) = -\lambda.$$

4. At the right edge of the data

$$\mathrm{SS}(k; y - \tilde{f}) = 0.$$

5. $\tilde{f}$ *minimizes the modality among all* $g \in N(y; \lambda)$.

We refer to this particular solution as the *robust taut string approximation.* The first four statements of Theorem 1 are illustrated in Figure 3 where 30 points of a sine curve were perturbated with Gaussian white noise. The left panel shows the data with the robust taut string approximation using $\lambda = 2$. The cumulative sums of the signs of the residuals are plotted in the right panel.

2.3 Robustness

Another important property of the robust taut string method is robustness against outliers. Since nonparametric regression methods have to be able to identify local phenomena and structures the usual global concept of breakdown from linear regression must be replaced by a localized version. The robust taut string

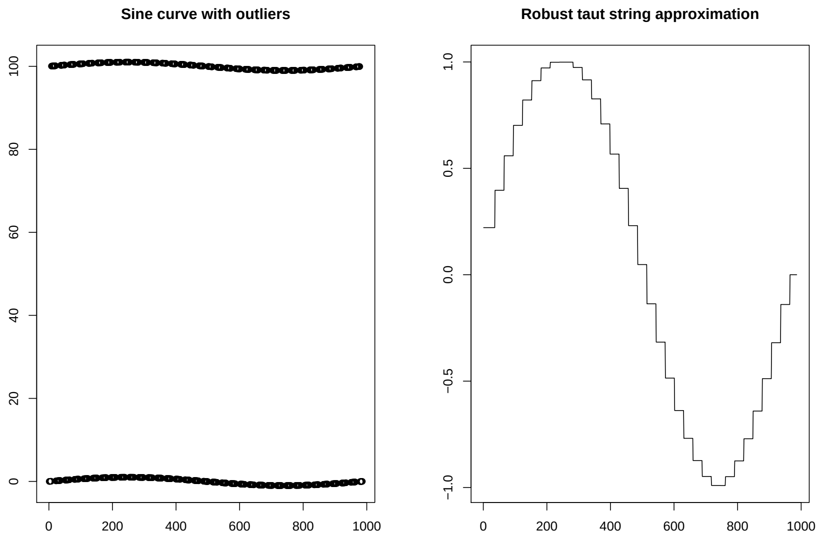

Fig. 4 Sine curve, outliers and the robust taut string method. The left panel shows 986 points of a sine curve disturbed by 476 outliers that come in patches of 14 outliers each. The right panel shows the reconstruction with the robust taut string method using $\lambda = 7$.

method is resistant against outliers in a sense that it can withstand patches of outliers consisting of up to 2λ measurements. More precisely, consider for given $\tau_1 < \tau_2 < \tau_3 < \tau_4$ the family $D(\tau)$ of all data sets that are equal to y on $[t_{\tau_1}, t_{\tau_2}]$ and $[t_{\tau_3}, t_{\tau_4}]$:

$$D(\tau; y) = \{\tilde{y} : \tilde{y}_i = y_i \text{ for all } i \in [t_{\tau_1}, t_{\tau_2}] \cup [t_{\tau_3}, t_{\tau_4}]\}.$$

Then a simple consequence from Theorem 1 is the following theorem:

Theorem 2 *For a given data set* $(t_i, y_i), i = 1, \ldots, n$ *and a given parameter* λ *consider*

$$b(\tau; y) = \sup_{\tau_2 < i < \tau_3, \tilde{y} \in D(\tau; y)} |\tilde{f}_\lambda(t_i) - \tilde{f}_\lambda(t_i)|.$$

Then

1. *if* $\tau_3 - \tau_2 \geq 2\lambda + 2$,

$$b(\tau; y) = \infty$$

2. *if* $\tau_3 - \tau_2 \leq 2\lambda + 1$ *and* $\min(\tau_2 - \tau_1, \tau_4 - \tau_3) \geq 2\lambda$,

$$b(\tau; y) < \infty$$

Thus, a patch of $2\lambda + 1$ outliers will always let the robust taut string approximate breakdown while smaller patches are do not influence the approximation if they are surrounded by $2\lambda + 1$ not outlying observations on each side. Figure 4 shows the robust taut string method applied to 986 points of a sine curve. The data shown in the left panel were disturbed by 34 outlier patches each of length 14 with gaps of size 15 between them, so there were 476 outlier in total. The right panel shows the approximation using $\lambda = 7$.

For another example we return to the balloon data. Since the longest outlier patch is somewhat smaller than 40, it makes sense to choose $\lambda = 20$. The resulting robust taut string approximation is indeed not very strongly influenced by the outliers as can be seen from Figure 5.

2.4 Modality

We now turn to the connection between the robust taut string method and local extreme values. Let $k(y; \lambda)$ denote the number of local extreme values of the robust taut string approximation using the tradeoff parameter λ. We have already seen in Theorem 1 that $\tilde{f}(y; \lambda)$ minimizes the modality among all vectors in $N(y; \lambda)$.

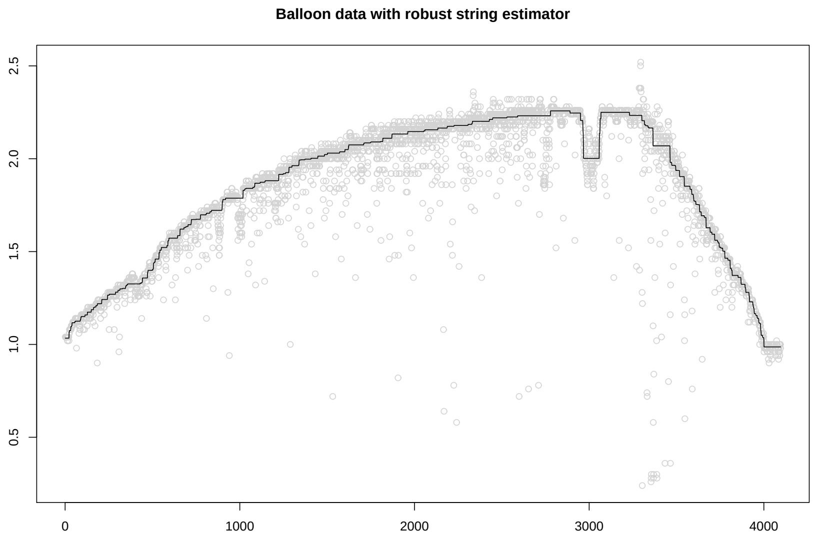

Fig. 5 Radiation measurements of a weather balloon and the robust taut string approximation with $\lambda = 20$.

Thus, if the data y_i are generated from the model (1) and we choose λ large enough such that $f \in N(y;\lambda)$ we will end up with at most as many local extreme values as the function f and avoid possible artifacts. Consequently, we require that with high probability

$$\max_{1\leq k\leq n} |\sum_{i=1}^{k} \mathrm{sign}(\varepsilon_i)| \leq \lambda.$$

If the ε_i have median 0, then $\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\mathrm{sign}(\varepsilon_i)$ is a random a walk which will converge to a Brownian motion as n tends to ∞. In this case on choosing $\lambda = C(\alpha)\sqrt{n}$ where $C(\alpha)$ is the α-quantile of the distribution of the maximum of a Brownian motion on $[0,1]$ we conclude that

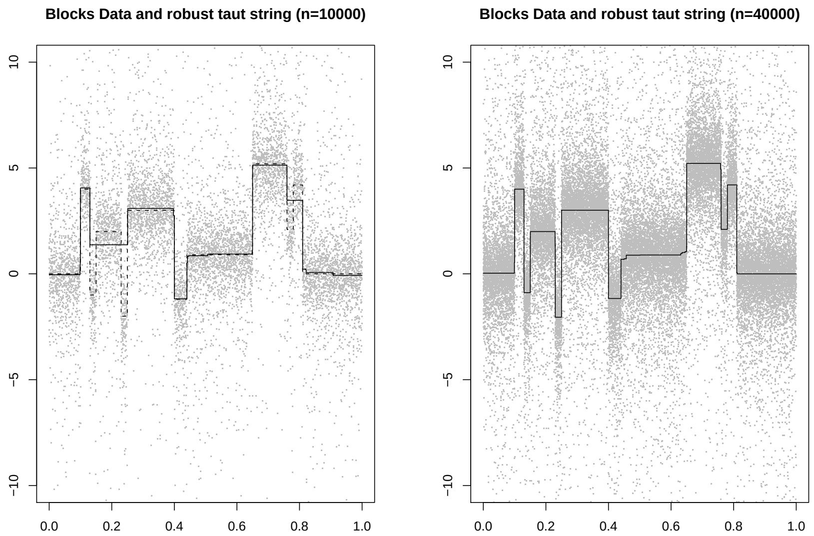

Fig. 6 Blocks data with Cauchy noise. The left panel shows 10000 points of the Blocks signal disturbed by Cauchy noise and an approximation using the robust taut string method and $\lambda = C(\alpha)\sqrt{n}$. The dashed line is the original Blocks signal. The same procedure was repeated in the right panel with $n = 35000$.

$$\lim_{n\to\infty} \mathbb{P}(k(y; C(\alpha)\sqrt{n}) \leq k) \geq \alpha \tag{4}$$

Unfortunately, the inequality (4) is not sharp. For practical purposes the value $C(\alpha)\sqrt{n}$ for the tradeoff parameter λ turns out to be much too large. This is illustrated in Figure 6 where 10000 points from the Blocks signal from Donoho et al. (1995) were disturbed by Cauchy noise and where some features are missing in the reconstruction using $\lambda = C(0.5)\sqrt{n}$. The sample size must be increased to 35000 to achieve the correct modality with come confidence. We introduce an automatic data-driven choice for λ in the section that improves the detection rate considerably.

2.5 Multiresolution of signs

We now give an approximation of white noise which adapts on the local behaviour of the observations. Our goal is to decide whether residuals related to some approximation of the data look like noise or not. Davies and Kovac (2001b) characterize Gaussian white noise by considering a multiresolution scheme of the residuals. We adapt their method, but consider sums of the signs of the residuals rather than the residuals themselves. This definition has previously been used by Dümbgen and Johns (2000) and Dümbgen (2001) to construct confidence bands for isotonic and convex median curves.

Assume for the moment that n is a power of two:

$$n = 2^J.$$

For $r \in \mathbb{R}^n, j \in \{0, \ldots, J\}$ and $k \in \{0, \frac{n}{2^j} - 1\}$ we define *multiresolution sums of signs*

$$w_{j,k}(r) = \frac{1}{2^{j/2}} \sum_{i=k2^j+1}^{(k+1)2^j} \operatorname{sign}(r_i).$$

If $\varepsilon_1, \ldots, \varepsilon_n$ are independent noise with $\mathbb{P}(\varepsilon > 0) = \mathbb{P}(\varepsilon < 0) = \frac{1}{2}$, then

$$\frac{1}{2} \sum_{i=k2^j+1}^{(k+1)2^j} \operatorname{sign}(\varepsilon_i)$$

will follow a binomial $\mathfrak{b}(2^j, 0.5)$-distribution. Therefore we regard the residuals as noise if the corresponding multiresolution sums of sign satisfy for all j and k the

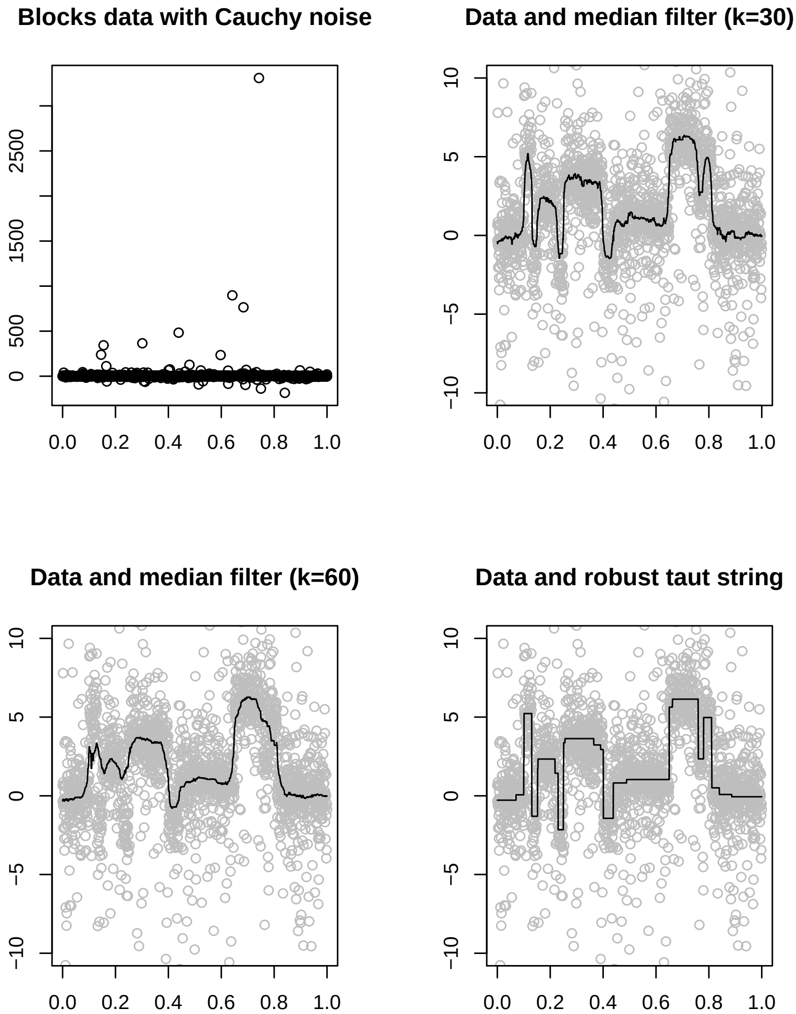

Fig. 7 Blocks data with Cauchy noise. The upper left panel shows 2048 points of the Blocks signal disturbed by Cauchy noise. The upper right and the lower left panel show reconstructions using a median fi lter with two different window widthes. The lower right panel shows the reconstruction with the robust taut string method using the multiresolution criterion to choose the tradeoff-parameter λ.

conditions

$$|w_{j,k}(r)| \leq \sqrt{2\log(n)}. \tag{5}$$

This leads to

Problem 1 (The multiresolution problem) Determine the smallest integer k for which there exists a function $\tilde{f}$ on $[0,1]$ with k local extreme values such that the residuals r satisfy the multiresolution condition (5).

Since there is no obvious connection between the modality and the multiresolution sums $w_{j,k}$ we adapt a similar technique as in Davies and Kovac (2001b) and use the robust taut string method to produce candidate functions with increasing modality and choose the first function that satisfies the conditions (5) as an approximation to the data.

This approach is illustrated in Figure 7 where 2048 points from the Blocks signal from Donoho et al. (1995) were disturbed by Cauchy noise and where the multiresolution method was able to retain the correct modality. Also shown are two approximations from a median filter using two different window widthes $k = 30$ and $k = 60$. The median filter tends to oversmooth the discontinuities and introduces spurious local extreme values while the automatic robust taut string method nearly attains the original Blocks signal.

References

1. BARRODALE, I., AND ROBERTS, F. An improved algorithm for discrete l_1 linear approximation. *SIAM Journal of Numerical Analysis 10* (1973), 839–848.

2. CHAUDHURI, P., AND MARRON, J. S. Sizer for exploration of structures in curves. *Journal of the American Statistical Association 94* (1999), 807–823.

3. DAVIES, L., AND GATHER, U. The identification of multiple outliers (with discussion). *Journal of the American Statistical Association 88* (1993), 782–801.

4. DAVIES, P. L. Data features. *Statistica Neerlandica 49* (1995), 185–245.

5. DAVIES, P. L., AND KOVAC, A. Densities, spectral densities and modality. *Submitted for publication* (2001).

6. DAVIES, P. L., AND KOVAC, A. Local extremes, runs, strings and multiresolution. *Annals of Statistics to appear* (2001).

7. DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G., AND PICARD, D. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society 57* (1995), 371–394.

8. DÜMBGEN, L. Confidence bands for convex median curves using sign-tests. *Preprint A-01-05* (2001).

9. DÜMBGEN, L., AND JOHNS, R. Confidence bands for isotonic median curves using sign-tests. *Preprint A-00-24* (2000).

10. FAN, J., AND GIJBELS, I. *Local Polynomial Modelling and its Applications*. Chapman and Hall, London, 1996.

11. MAMMEN, E., AND VAN DE GEER, S. Locally adaptive regression splines. *Annals of Statistics 25* (1997), 387–413.

12. NADARAYA, E. A. On estimating regression. *Theory of Probability and its Applications 10* (1964), 186–190.

13. SILVERMAN, B. W. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society 47* (1985), 1–52.

14. WATSON, G. S. Smooth regression analysis. *Sankhyā 26* (1964), 101–116.