

Smooth functions and local extreme values

A. Kovac¹

*Department of Mathematics
University of Bristol*

Abstract

Given a sample of n observations y_1, \dots, y_n at time points t_1, \dots, t_n we consider the problem of specifying a function \tilde{f} such that \tilde{f}

- is smooth,
- fits the data in the sense that the residuals $y_i - \tilde{f}(t_i)$ satisfy the multiresolution criterion

$$\left| \frac{1}{\sqrt{k-j+1}} \sum_{i=j}^k y_i - \tilde{f}(t_i) \right| < \sqrt{2 \log(n)} \sigma \quad 1 \leq j \leq k \leq n, \quad (1)$$

- is as simple as possible so that \tilde{f} exhibits the minimum number of local extreme values.

We analyse in particular a fast method which is based on minimising

$$\sum_{i=1}^n (y_i - f(t_i))^2 + \sum_{i=1}^{n-1} \lambda_i \sqrt{(f_{i+1} - f_i)^2 + (t_{i+1} - t_i)^2}$$

where the λ_i are chosen automatically. The new method can also be applied to density estimation.

Key words: Nonparametric regression, modality, smoothness, total variation.

¹ Research supported in part by Sonderforschungsbereich 475, University of Dortmund.

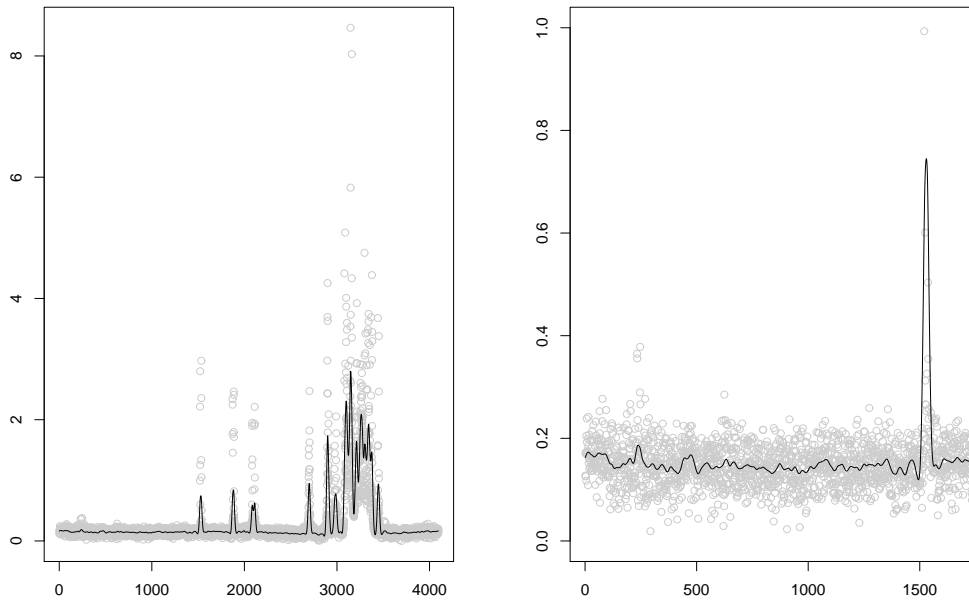


Fig. 1. Data from Spectroscopy taken in the analysis of a gall stone and an approximation obtained from spline smoothing. Left panel: entire data set, right panel: excerpt.

1 Nonparametric regression and modality

1.1 Introduction

Local extreme values often play an important rôle in non-parametric statistics. Figure 1 shows data from spectroscopy and consists of measurements that were taken in the analysis of a gall stone. Each peak corresponds to a certain type of structure that is present in the substance being analysed and the power of each peak measures how many structures are contained in the substance. The chemists that were analysing this data set want an automatic procedure that removes all the noise and keeps exactly the true peaks of the signal.

Figure 1 also displays an approximation to this data set obtained from spline smoothing. The spline does not approximate the underlying signal very well because the height of each peak in the spline estimate is much smaller than the data points near the peak. At the same time the estimate contains a lot of local extreme values that do not appear to be related to true features of the signal. Even worse is that some of these artificial extreme values are even more pronounced than some of the true peaks. This can be seen in the excerpt in the right panel where the peak near $t = 500$ or the one just left to the true

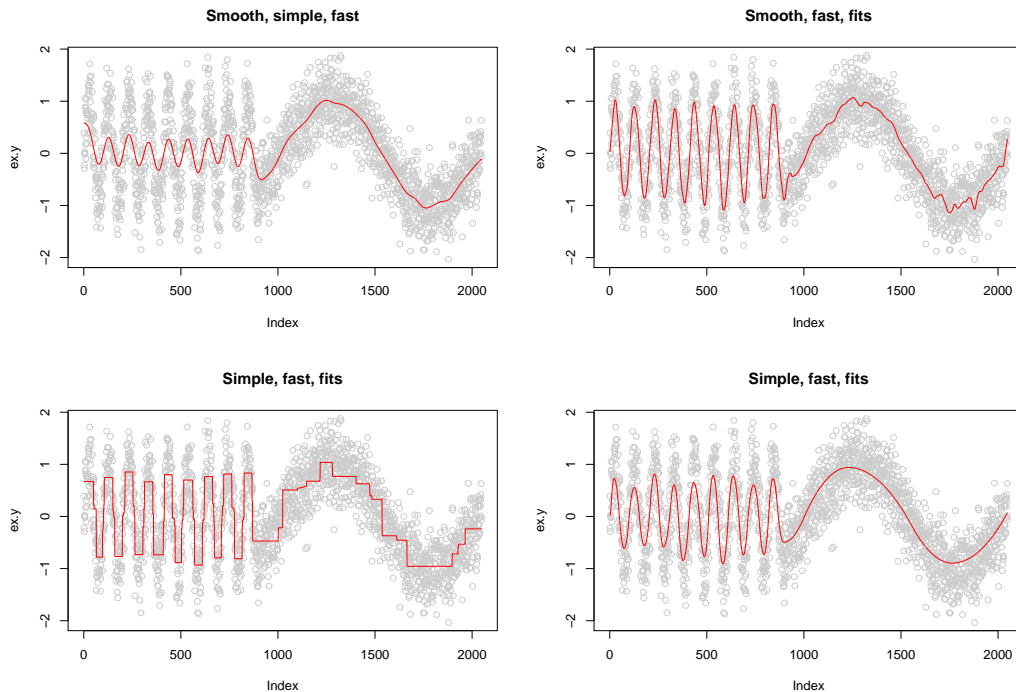


Fig. 2. Four smoothing methods applied to the same sample. Top left: kernel estimate using large bandwidth, top right: kernel estimate using local plug-in bandwidths, bottom left: taut string method with local squeezing, bottom right: minimisation of total variation under multiresolution constraints

peak at $t = 1500$ look larger than the true peak at $t = 250$. This makes it hard to distinguish true peaks from artificial peaks.

We consider in this paper the problem of non-parametric regression and our goal is to specify a function f that fits the data and is simple in the sense that it does not contain any spurious local extreme values. We also require our approximations to be smooth because although even piecewise constant approximations can be useful for exploring the structure, there exists huge interest in smooth approximations among people who actually employ non-parametric methods in applications. Finally we are seeking a fast algorithm for the computation of a particular approximation.

Figure 2 shows typical problems with existing methods in nonparametric regression. Four different methods were applied to the same simulated data set. A kernel estimate using a relatively large bandwidth can be seen in the top left panel. The computation time was very short and the approximation is definitely smooth and simple, however it does not fit the data in the left half. The top right panel shows another kernel estimate, this time using automatic local plug-in bandwidths (Seifert et al., 1994). This approximation fits the data very well and again was fast to compute. Since a Gaussian kernel was used the function is infinitely differentiable and hence smooth. On the other

hand there are a couple of spurious local extreme values and therefore the estimate is not simple. The approximation in the bottom left corner is the result of the taut string method (Davies and Kovac, 2001) which is a very fast method producing simple approximations that fit the data, however they are piecewise constant and therefore not smooth. Finally the last panel shows a smooth and simple function that just fits the data. It was produced by minimising the total variation of the second derivative subject to multiresolution conditions (Kovac and Meise, 2006) which gives rise to a huge linear program that for complicated data sets can not be solved quickly.

Our approach towards approximations of the data without spurious local extreme values relies on a suitable definition of adequacy and methods that produce simple approximations. A measure of adequacy gives rise to a set of adequate functions \mathcal{F} , each of them representing a plausible model for the data in the sense that the data look like a “typical” sample from the model. In the context of nonparametric regression this set of adequacy will contain all functions such that sums of the residuals at different scales and locations are not larger than what we would expect from white noise. For a more precise definition we refer to Davies (1995).

Having specified the set of adequate functions our next step is to look for an adequate function $\tilde{f}^1 \in \mathcal{F}$ with minimal modality. Since it is difficult to obtain an exact solution to this problem the taut string method (Davies and Kovac, 2001) is used instead to generate a sequence of functions with increasing complexity, ie increasing number of local extreme values. Then \tilde{f}^1 is chosen as the first adequate function, ie the function with the smallest number of local extreme values within this sequence. In a second step a smooth function $\tilde{f}^2 \in \mathcal{F}$ with the same monotonicity behaviour as \tilde{f}^1 is calculated. This is achieved by minimising a functional similar to the taut string functional, but which leads to smooth rather than piecewise constant approximations. Additional monotonicity constraints guarantee the same monotonicity behaviour as the taut string approximation.

1.2 Previous work

Much work has been done on the problem of nonparametric regression. Of the different approaches we mention kernel estimation (Nadaraya 1964; Watson 1964), spline smoothing (Silverman 1985; Green and Silverman 1994), wavelets (Donoho, Johnstone, Kerkyacharian and Picard 1995) and local polynomials (Fan and Gijbels 1995, 1996). This paper is particularly concerned with penalised regression estimators. Again the literature is vast and still growing. As a good starting point we recommend Huang (2003), Antoniadis and Fan (2001) and van de Geer (2001).

None of these methods is directly concerned with local extremes. One of the few methods where positions of local extreme values play a crucial rôle is the taut string method (Davies and Kovac, 2001). See also the references therein for more work in this area.

An alternative approach to minimising the modality among all functions that satisfy the multiresolution criterion is to minimise the total variation instead (Kovac and Meise, 2006). This leads to a linear program that can be solved by standard techniques. By minimising the total variation of the second derivative and the introduction of additional monotonicity and convexity constraints it is possible to obtain smooth and simple approximations, but the computational burden can sometimes be large.

The remainder of this paper is organised as follows: In Section 2 we introduce the smooth taut string functional and discuss a fast method for calculating a minimiser of this functional. We also discuss an automatic choice of the local smoothing parameters involved and the combination with monotonicity constraints. In Section 3 we compare the new method with five other automatic smoothing methods with respect to number of local extreme values and mean squared error and discuss an application to the data from Spectroscopy. We briefly discuss an application of the new methodology to density estimation and two-dimensional curves in Section 4. Some short proofs are deferred to Section 5.

2 The smooth taut string method

2.1 Nonparametric regression and the multiresolution criterion

We study the nonparametric regression problem in more detail. Given observations $y_1, \dots, y_n \in \mathbb{R}$ at time points $t_1 < t_2 < \dots < t_n$ we consider the problem of specifying a function f such that the residuals $y_i - f_i$ look like noise and the function is as simple as possible.

More specifically we consider a set \mathcal{F} of adequate functions

$$\mathcal{F} = \{f \in \mathbb{R}^n : |\sum_{i \in I} (y_i - f_i)| \leq \gamma_I \text{ for all } I \in \mathcal{I}\}.$$

Here \mathcal{I} is a collection of index sets on which we check if the sum of the residuals is as small as we would expect from noise. A typical choice is to take all possible subintervals

$$\mathcal{I}_1 = \{ \{j, j+1, \dots, k\} \text{ for all } 1 \leq j \leq k \leq n \}.$$

Computational complexity can be reduced by considering a smaller collection like all intervals with dyadic end points

$$\mathcal{I}_2 = \{ \{2^j k + 1, \dots, 2^j(k + 1)\} \text{ for all } 0 \leq j \leq \lfloor \log_2(n) \rfloor, k = 0, 1, \dots, \lceil \frac{n}{2^j} \rceil \}.$$

The constants γ_I will usually depend only on $|I|$. In the case of iid normal noise a suitable choice is to set

$$\gamma_I = \sqrt{2|I| \log(n)} \tilde{\sigma}$$

where $\tilde{\sigma}$ is an estimate of the standard deviation derived from the data. Throughout this article σ will be estimated by

$$\tilde{\sigma} = \frac{1.48}{\sqrt{2}} \text{Median}\{|y(t_2) - y(t_1)|, \dots, |y(t_n) - y(t_{n-1})|\}$$

(Davies and Kovac, 2001).

The aim of this paper is to find a smooth function $\tilde{f} \in \mathcal{F}$ with as few local extreme values as possible as well as a fast algorithm to calculate \tilde{f} .

2.2 The smooth taut string functional

The taut string method is known to produce functions with small modality. It can be viewed as a fast algorithm for minimising a functional where the L_2 -norm of the residuals is penalised by the total variation of f :

$$T^{TS}(f) = \sum_{i=1}^n (y_i - f_i)^2 + \lambda \sum_{i=1}^{n-1} |f_{i+1} - f_i| = \min.$$

The penalty term looks at fluctuations in y -direction and therefore two functions have the same total variation if their local extreme values are identical. This property of the total variation norm leads to piecewise approximations when $T(f)$ is minimised. A straightforward idea to obtain smoother approximations is to replace the differences between subsequent values of f by their Euclidean distances and consider the new functional

$$T^{SM}(f) = \sum_{i=1}^n (y_i - f_i)^2 + \lambda \sum_{i=1}^{n-1} \sqrt{(t_{i+1} - t_i)^2 + (f_{i+1} - f_i)^2}$$

or more general a version that allows for localised penalty terms

$$T^{SM}(f) = \sum_{i=1}^n (y_i - f_i)^2 + \sum_{i=1}^{n-1} \lambda_i \sqrt{(t_{i+1} - t_i)^2 + (f_{i+1} - f_i)^2}. \quad (2)$$

Unlike the original taut string functional the smooth taut string functional is differentiable, so minimisation is theoretically possible using standard techniques like steepest descent or Newton-Raphson method. Unfortunately the speed of convergence decays rapidly and although the value of the functional will not change a lot after some initial iterations the changes in f can be significant.

2.3 An Algorithm

We develop in this section a fast algorithm to find the minimiser of the smooth taut string functional T^{SM} in 2. To make notation easier we consider the more general functional

$$T(f) = \sum_{i=1}^n (y_i - f_i)^2 + \sum_{i=1}^{n-1} \lambda_i g_i(f_{i+1} - f_i) \quad (3)$$

where we assume that each of the functions g_i is convex, continuously differentiable and takes a global minimum in $t = 0$.

Before we move on to a description of the algorithm we first summarise some important properties about solutions of (3) in a lemma:

Lemma 1 *There is a unique minimiser \tilde{f} of the functional (3) and it satisfies:*

(i) *For all $1 \leq j \leq k \leq n$*

$$2 \sum_{i=j}^k (f_i - y_i) = \begin{cases} \lambda_k g'(f_{k+1} - f_k) - \lambda_{j-1} g'(f_j - f_{j-1}) & \text{if } 1 < j \leq k < n \\ \lambda_k g'(f_{k+1} - f_k) & \text{if } 1 = j \leq k < n \\ -\lambda_{j-1} g'(f_j - f_{j-1}) & \text{if } 1 < j \leq k = n \\ 0 & \text{if } 1 = j, k = n. \end{cases}$$

(ii) *For all $k < n$*

$$\tilde{f}_{k+1} = \tilde{f}_k + (g'_k)^{-1} \left(\frac{2}{\lambda_k} \sum_{i=1}^k (\tilde{f}_i - y_i) \right). \quad (4)$$

(iii) *Let f be some vector that satisfies the equations (4) for all $k = 2, \dots, n$. If $f_1 > \tilde{f}_1$, then $f_k > \tilde{f}_k$ for all $k > 2$ and vice versa in the case where $f_1 < \tilde{f}_1$.*

The idea behind the algorithm is that once we know \tilde{f}_1 we can at least theoretically calculate the remaining components of \tilde{f} quite easily. Let us assume for the moment that the derivatives of the g'_k are all unbounded. If \tilde{f}_1 was known

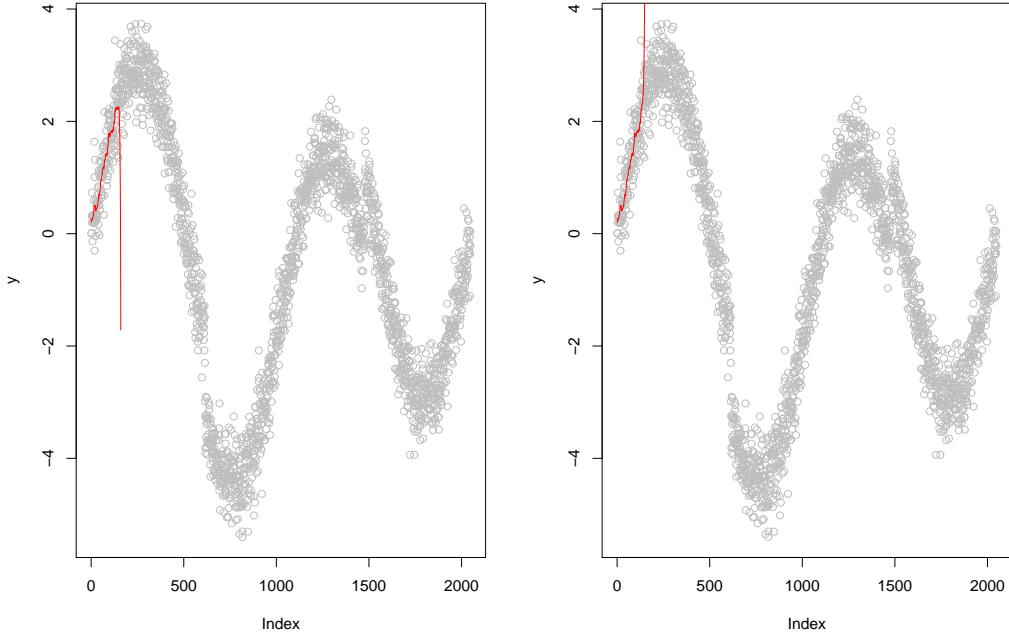


Fig. 3. The solutions of (4) using $f_1 = 0.2017878903881263$ and $f_1 = 0.2017878903881264$.

we could use the recursive formula (4) to calculate successively $\tilde{f}_2, \tilde{f}_3, \dots, \tilde{f}_n$. Of course, \tilde{f}_1 is not usually known, but Lemma 1 makes it possible to find it. In particular the lemma shows that $\sum_{i=1}^n (f_i - y_i) > 0$ if $f_1 > \tilde{f}_1$ and $\sum_{i=1}^n (f_i - y_i) < 0$ if $f_1 < \tilde{f}_1$. So we can use nested intervals to approximate \tilde{f}_1 to arbitrary precision ε by using the following algorithm:

- (I) Let l and u be crude initial lower and upper bounds for \tilde{f}_1 like $l = y_1 - 5\sigma$ and $u = y_1 + 5\sigma$ or even $l = \min(y_1, \dots, y_n) - 5\sigma$ and $u = \max(y_1, \dots, y_n) + 5\sigma$.
- (II) Determine recursively a solution f of (4) with $f_1 = (l + u)/2$.
- (III) If $\sum_{i=1}^n (f_i - y_i) > 0$, set $u := f_1$, otherwise set $l := f_1$.
- (IV) If $u - l > \varepsilon$ go to step (II).

Theoretically this algorithm should provide a minimiser of the functional (3) in $O(n)$ operations provided that the functions g_j do not depend on n or at least the complexity of the evaluation of $(g'_{k-1})^{-1}$ is independent of n .

If some of the derivatives of the g'_k are bounded like in the example (2) above, it is possible that the recursive formula (4) used in step (2) of the algorithm does not work because the term $D := \frac{2}{\lambda_{k-1}} \sum_{i=1}^{k-1} (f_i - y_i)$ exceeds either an upper or lower bound of g'_{k-1} . If it is too large, then $\sum_{i=1}^{k-1} f_i$ and hence f_1 is too large, so we need to set $u := f_1$. Similarly we set $l := f_1$ if D becomes too small. With this additional step the algorithm described above should enable

us to find the minimiser of the functional (2). In practice, however numerical problems occur. Figure 3 shows the solution of (4) when applied to data from a scaled version of the Heavisine signal (Donoho et al., 1995) with added Gaussian noise for $f_1 = 0.2017878903881263$ and $f_1 = 0.2017878903881264$ where $g_i(x) = \sqrt{1/n + x^2}$ and $\lambda_i = 30$ for all $i = 1, \dots, n-1$ and $n = 2048$. It can clearly be seen that the recursion formula (4) is extremely sensitive with respect to f_1 and that even extremely small errors accumulate very fast.

Therefore we use simultaneous nested intervals for all data points. The idea is that once $u - l$ is smaller than some given precision and thus f_1 is sufficiently close to \tilde{f}_1 , we move to \tilde{f}_2 and use nested intervals for this data point as well and so on. Computational complexity is reduced by updating lower and upper bounds for each data point at each step:

- (I) Let l and u be vectors of crude initial lower and upper bounds for $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n$ like $l_i = y_i - 5\sigma$ and $u_i = y_i + 5\sigma$ or even $l_i = \min(y_1, \dots, y_n) - 5\sigma$ and $u_i = \max(y_1, \dots, y_n) + 5\sigma$ for all $i = 1, \dots, n$.
- (II) Let $k := 1$.
- (III) Set $f_j = (d_j + u_j)/2$ for all $j = 1, \dots, k$. Then use the recursion formula (4) to calculate f_{k+1}, \dots, f_n . If the evaluation of some point f_l with $k < l \leq n$ is not possible because the argument of $(g'_{l-1})^{-1}$ is too large, then set $f_j := \infty$ for all $j > k$ and similar $f_j := -\infty$ if the argument is too small.
- (IV) If $\sum_{i=1}^n (f_i - y_i) > 0$, set $u_i := \min(f_i, u_i)$ for all $i \geq k$. Otherwise set $l_i := \max(f_i, l_i)$ for all $i \geq k$.
- (V) If $u_k - d_k \leq \varepsilon$ set $k := k + 1$.
- (VI) If $k \leq n$ go to step (II).

This algorithm was used to produce Figure 4 where the new methods was applied to the same data set as in Figure 3. Computationally the algorithm works very fast as shown by the following lemma.

Lemma 2 *Assume that g_1, \dots, g_{n-1} are bounded and that $(g'_k)^{-1}$ is differentiable for $k = 1, \dots, n-1$. Denote $\mu_n = \min_{k,x} ((g'_k)^{-1})'(x)$. Then the computational complexity is of order $O(n/\log(1 + \mu_n))$.*

2.4 Deriving an adequate approximation – adaptive choice of λ_i using local squeezing

The smooth taut string functional (2) has the potential to provide simple and smooth approximations to data. Now we discuss how to find suitable values for the parameters λ_i such that the resulting approximation satisfies the multiresolution criterion (1) and is as simple as possible. We adapt the local squeezing approach by Davies and Kovac (2001) and use the functional

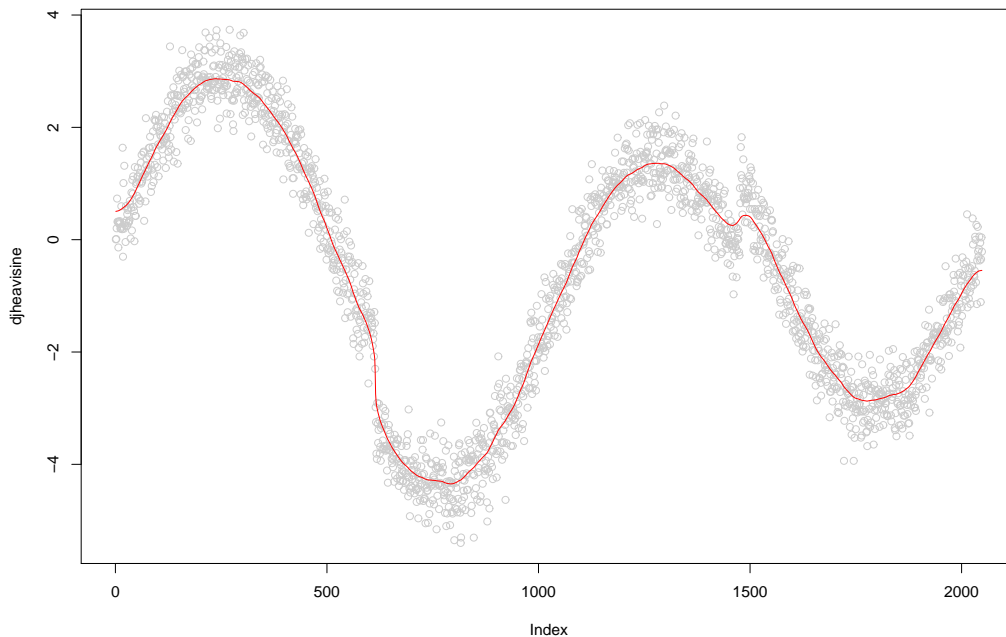


Fig. 4. Simultaneous nested intervals applied to the same problem as in Figure 3.

(2) to generate a sequence of functions with increasing complexity, ie increasing number of local extreme values. The first adequate function, ie the function with the smallest number of local extreme values within this sequence is used as an approximation to the data.

More precisely we start with some tuning vector $\lambda^{(1)}$ having constant components $\lambda_1^1 = \lambda_2^1 = \dots = \lambda_n^1$ where λ_1 is so large that the corresponding solution of (2) is too simple and oversmooths the data. Suppose that we have already chosen tuning vectors $\lambda^{(1)}, \dots, \lambda^{(s)}$, and the corresponding fits are denoted by $\tilde{f}^{(1)}, \dots, \tilde{f}^{(s)}$. If $\tilde{f}^{(s)}$ is still inadequate for the data, we define $\mathcal{J}^{(s)}$ to be the union of all intervals $\{j-1, j, \dots, k\}$ such that $\{j, \dots, k\}$ is an interval in \mathcal{I} violating (1) with $\tilde{f} = \tilde{f}^{(s)}$. Then for some fixed $\gamma \in (0, 1)$, e.g. $\gamma = 0.9$, we define

$$\lambda_j^{(s+1)} = \begin{cases} \gamma \lambda_j^{(s)} & \text{if } j \in \mathcal{J}^{(s)} \\ \lambda_j^{(s)} & \text{if } j \notin \mathcal{J}^{(s)} \end{cases}$$

From (i) in Lemma 1 we can derive that for sufficiently large s the fit $\tilde{f} = \tilde{f}^{(s)}$ does satisfy (1) for all $I \in \mathcal{I}$.

The left panel of Figure 5 shows the application of this local squeezing approach to the simulated data set that we have analysed before. In this case the new procedure performs exactly as it is supposed to do and produces a smooth and simple function that is an adequate approximation to the data.

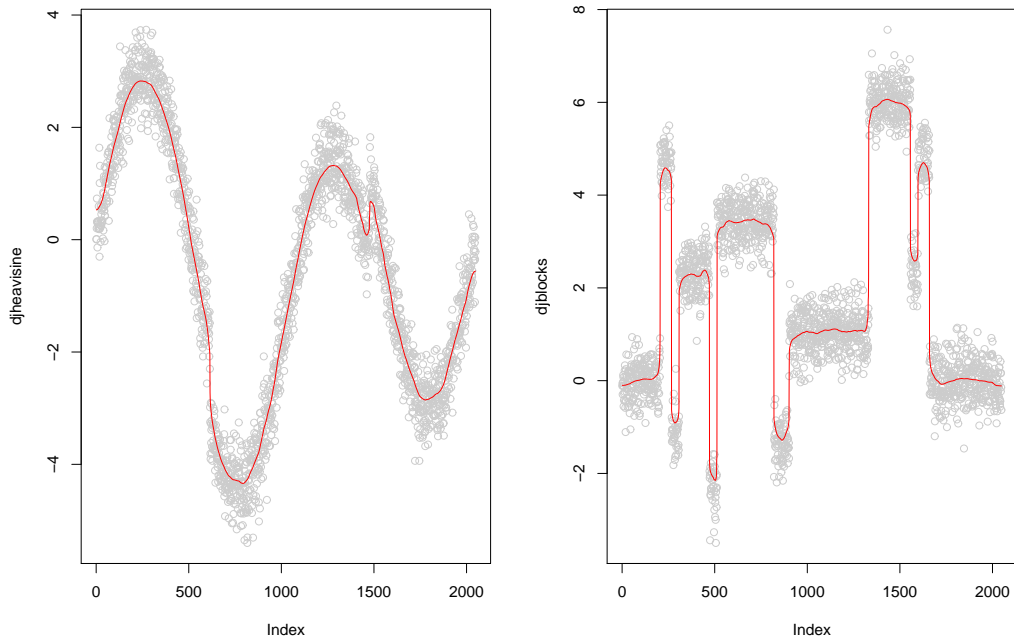


Fig. 5. The smooth taut string method applied to two samples from the Heavisine and Blocks signal. Bandwidths were obtained automatically by local squeezing.

2.5 Monotonicity constraints

The right panel of Figure 5 shows the application to a sample from the Blocks signal which is piecewise constant. The approximation produced by the new method exhibits several small spurious local extreme values. To get rid of these we propose to introduce monotonicity constraints. For numbers $\tau_1, \tau_2, \dots, \tau_{n-1} \in \{-1, 1\}$ we consider the problem of minimising $T(f)$ in (3) among all f such that $(f_{i+1} - f_i)\tau_i \geq 0$. The constraints can be derived from the usual taut string method which is known to have the correct modality asymptotically or alternatively from minimising the total variation subject to multiresolution constraints (Kovac and Meise, 2006). In both cases the existence of a solution to this constrained optimisation problem is guaranteed.

The monotonicity constraints can easily be incorporated into the algorithm developed above. The idea is based on the following property:

Lemma 3 *There is a unique minimiser \tilde{f} of the functional (3) subject to multiresolution constraints $(f_{i+1} - f_i)\tau_i \geq 0$ and it satisfies:*

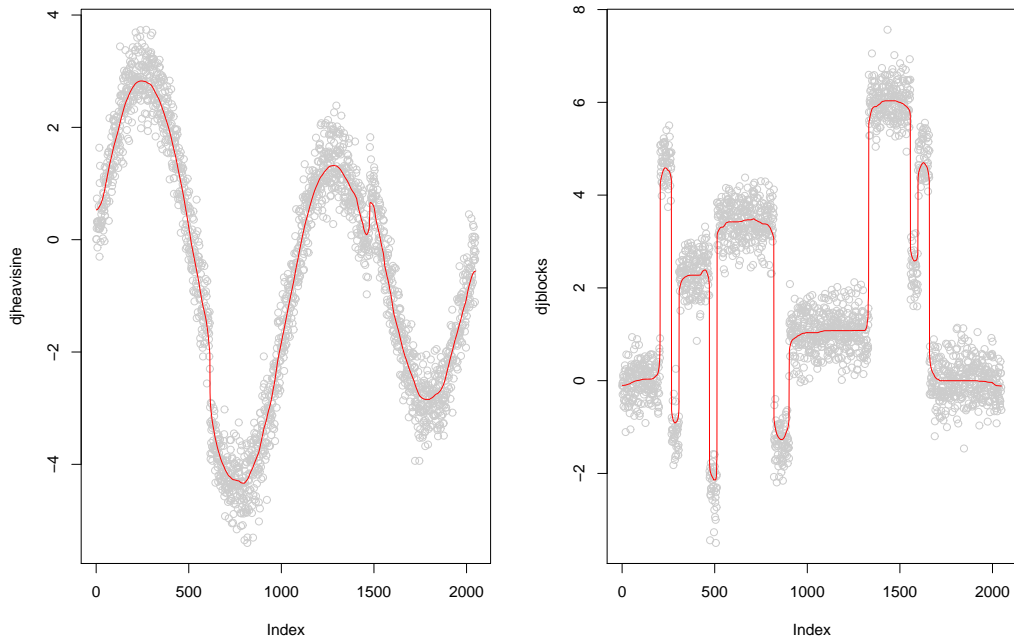


Fig. 6. The smooth taut string method using local squeezing and monotonicity constraints applied to two samples from the Heavisine and Blocks signal.

(i) For all $k < n$ either

$$\tilde{f}_{k+1} = \begin{cases} \tilde{f}_k + (g'_k)^{-1} \left(\frac{2}{\lambda_k} \sum_{i=1}^k (\tilde{f}_i - y_i) \right) & \text{if } \tau_i \sum_{i=1}^k (\tilde{f}_i - y_i) > 0 \\ \tilde{f}_k & \text{if } \tau_i \sum_{i=1}^k (\tilde{f}_i - y_i) \leq 0 \end{cases} \quad (5)$$

(ii) Let f be some vector that satisfies the equations (5) for all $k = 2, \dots, n$. If $f_1 > \tilde{f}_1$, then $f_k > \tilde{f}_k$ for all $k > 2$ and vice versa in the case where $f_1 < \tilde{f}_1$.

Therefore we can calculate a solution of the constrained minimisation problem exactly as set out in Section 2.3 for the unconstrained problem. Moreover, the local squeezing technique described above can be used to supply an adequate approximation for noisy data with the specified monotonicity behaviour. Figure 6 shows an application of this approach to the Heavisine and Blocks signal. Whereas the approximation for the Heavisine signal is not much different from the approximation obtained above in Figure 5, in the case of the piecewise constant Blocks signal the approximation exhibits much less noise and in particular no spurious local extreme values.

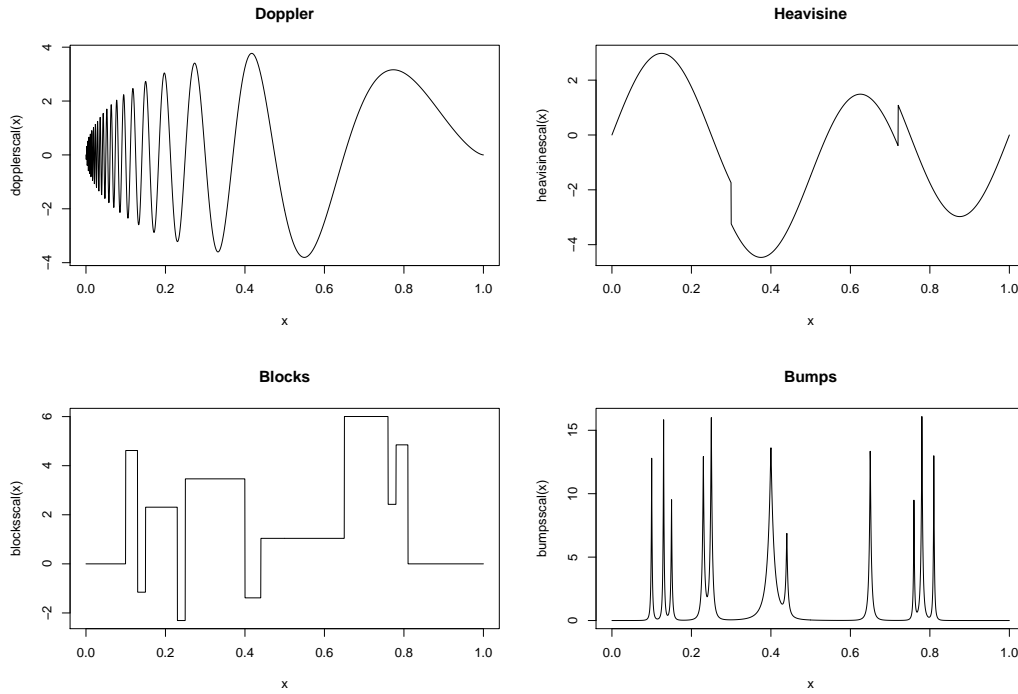


Fig. 7. Rescaled versions of standard test signals by Donoho and Johnstone.

3 Numerical examples

3.1 A simulation study

A simulation study was carried out to compare the smooth taut string method with five other automatic smoothing methods with respect to number of local extreme values and mean squared error. The five competitors were

- the taut string method (Davies and Kovac, 2001) using their local squeezing technique to determine the local tube widths,
- wavelet thresholding (Donoho et al, 1995) using universal $\sqrt{2 \log(n)}$ hard thresholds and Daubechies' least asymmetric wavelets with five vanishing moments,
- smoothing splines (Silverman, 1985) using all time points as knots and cross validation to determine the smoothing parameter,
- kernel estimator with local plugin-bandwidths (Brockmann et al, 1993) evaluated at all time points,
- local polynomial estimator with local bandwidths determined by a direct plug-in method (Ruppert et al, 1995).

Rescaled versions of four standard test signals by Donoho (1993) and Donoho and Johnstone (1994) were used which can be seen in Figure 7. For each test

Method	n	Doppler	Heavisine	Blocks	Bumps
		true mod: 40	true mod: 6	true mod: 9	true mod: 21
Taut string	256	17	4	9	21
Smooth string		17	4	9	21
Wavelets		26	6	60	68
Splines		57	6	71	94
Kernel estimator		41	8	57	49
Local polynomials		26	6	45	33
Taut string	512	21	6	9	21
Smooth string		21	6	9	21
Wavelets		28	10	91	90
Splines		115	8	141	212
Kernel estimator		62	10	104	95
Local polynomials		70	6	87	88
Taut string	1024	25	6	9	21
Smooth string		25	6	9	21
Wavelets		32	16	136	120
Splines		181	11	240	375
Kernel estimator		80	12	182	192
Local polynomials		124	6	172	199
Taut string	2048	28	6	9	21
Smooth string		29	6	9	21
Wavelets		38	20	172	132
Splines		244	16	391	543
Kernel estimator		96	14	306	343
Local polynomials		178	8	211	393
Taut string	4096	32	6	9	21
Smooth string		32	6	9	21
Wavelets		41	26	224	136
Splines		304	28	615	702
Kernel estimator		113	18	503	522
Local polynomials		231	24	415	487

Table 1

Median modality for six different smoothing methods, four test signals and five different sample sizes.

signal and three different sample sizes ($n = 256, 1024, 4096$) 1000 samples were generated by adding Gaussian noise with mean 0 and standard deviation 0.4.

The results are reported in Tables 1 and 2 and interesting conclusions can be drawn from them. With respect to modality the median number of local extreme values was determined for each signal, method and sample size. The taut string method and the smooth taut string method attained on average the correct number in most situations except for small samples of the Heavisine signal where the second discontinuity was not detected and for the Doppler data where even for larger samples some very small oscillations at

Method	n	Doppler	Heavisine	Blocks	Bumps
Taut string	256	0.315 (0.00121)	0.116 (0.00047)	0.019 (0.00031)	0.142 (0.00093)
Smooth string		0.338 (0.00121)	0.078 (0.00041)	0.106 (0.00071)	0.225 (0.00112)
Wavelets		0.118 (0.00073)	0.048 (0.00028)	0.312 (0.00148)	0.266 (0.00137)
Splines		0.152 (0.00005)	0.038 (0.00014)	0.200 (0.00062)	0.464 (0.00164)
Kernel estimator		0.131 (0.00030)	0.028 (0.00015)	0.186 (0.00044)	0.744 (0.00041)
Local polynomials		0.293 (0.00052)	0.030 (0.00015)	0.355 (0.00139)	1.896 (0.00267)
Taut string	512	0.200 (0.00061)	0.082 (0.00032)	0.009 (0.00016)	0.118 (0.00046)
Smooth string		0.178 (0.00053)	0.043 (0.00021)	0.047 (0.00026)	0.134 (0.00051)
Wavelets		0.065 (0.00031)	0.029 (0.00019)	0.165 (0.00066)	0.175 (0.00065)
Splines		0.081 (0.00020)	0.208 (0.00008)	0.139 (0.00036)	0.165 (0.00047)
Kernel estimator		0.074 (0.00018)	0.019 (0.00008)	0.134 (0.00030)	0.199 (0.00041)
Local polynomials		0.107 (0.00021)	0.021 (0.00008)	0.206 (0.00034)	0.764 (0.00063)
Taut string	1024	0.132 (0.00032)	0.052 (0.00021)	0.005 (0.00010)	0.120 (0.00040)
Smooth string		0.099 (0.00024)	0.023 (0.00012)	0.027 (0.00013)	0.082 (0.00023)
Wavelets		0.039 (0.00016)	0.017 (0.00012)	0.081 (0.00027)	0.101 (0.00034)
Splines		0.051 (0.00010)	0.014 (0.00005)	0.097 (0.00016)	0.109 (0.00017)
Kernel estimator		0.052 (0.00011)	0.013 (0.00005)	0.088 (0.00016)	0.080 (0.00014)
Local polynomials		0.059 (0.00012)	0.015 (0.00005)	0.123 (0.00029)	0.261 (0.00025)
Taut string	2048	0.087 (0.00018)	0.034 (0.00012)	0.003 (0.00006)	0.089 (0.00024)
Smooth string		0.054 (0.00012)	0.012 (0.00006)	0.020 (0.00008)	0.049 (0.00015)
Wavelets		0.023 (0.00008)	0.010 (0.00006)	0.049 (0.00014)	0.054 (0.00016)
Splines		0.031 (0.00006)	0.010 (0.00003)	0.069 (0.00008)	0.065 (0.00008)
Kernel estimator		0.036 (0.00006)	0.009 (0.00003)	0.059 (0.00008)	0.051 (0.00012)
Local polynomials		0.034 (0.00007)	0.011 (0.00004)	0.098 (0.00014)	0.089 (0.00011)
Taut string	4096	0.056 (0.00011)	0.020 (0.00006)	0.002 (0.00003)	0.065 (0.00013)
Smooth string		0.029 (0.00006)	0.005 (0.00003)	0.014 (0.00005)	0.028 (0.00005)
Wavelets		0.011 (0.00003)	0.006 (0.00003)	0.031 (0.00007)	0.027 (0.00007)
Splines		0.018 (0.00003)	0.007 (0.00002)	0.047 (0.00004)	0.039 (0.00004)
Kernel estimator		0.025 (0.00004)	0.006 (0.00002)	0.038 (0.00005)	0.030 (0.00011)
Local polynomials		0.020 (0.00004)	0.007 (0.00002)	0.056 (0.00004)	0.054 (0.00006)

Table 2

Mean squared error for six different smoothing methods, four test signals and five different sample sizes. In brackets the standard errors.

the left-hand side were not detected. However, spurious local extreme values were only occasionally observed. In contrast all other methods produced on average much more complicated models with spline smoothing, the only competitor without a local bandwidth choice, performing worst. Larger number of artefacts were recorded for the rough Blocks and Bumps signals than for the smoother Heavisine and Doppler signals.

The bandwidth choices for the spline, kernel and local polynomial methods are all aimed at minimising the mean squared error. The smooth taut string method aims to minimise the modality instead, but attains nevertheless in half of the 20 test beds a smaller mean squared error than these methods. In

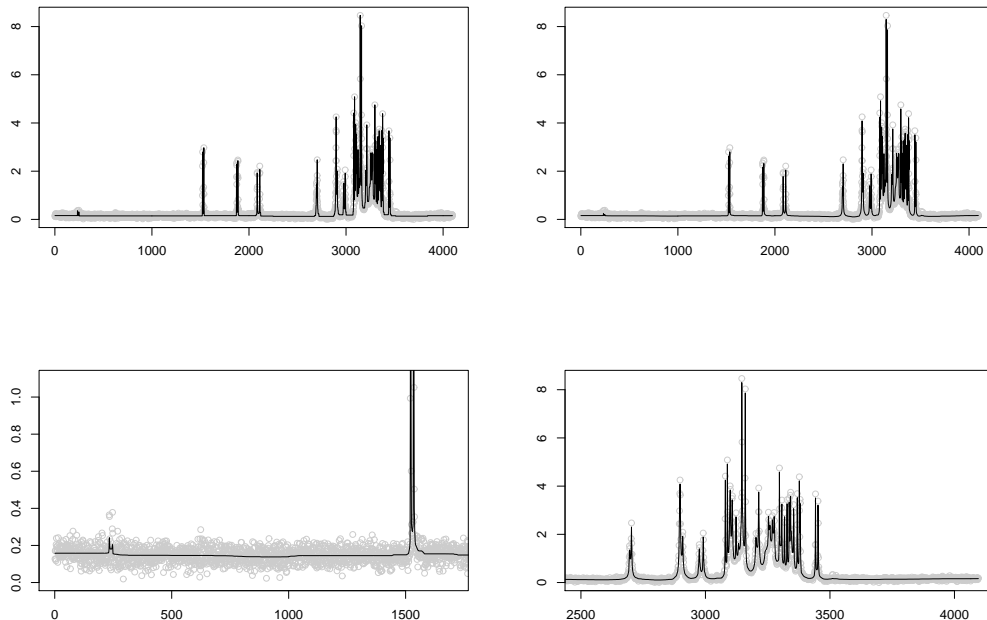


Fig. 8. Data from Spectroscopy. Upper panels: Usual and smooth taut string approximations. Bottom panels: Excerpts of the smooth taut string approximation.

all, but one of these 10 test beds the smooth taut string methods performs also better than the wavelet method. The usual taut string method generates piecewise constant approximations and therefore considerably improves the mean squared error for the Blocks signal. For other data sets the smooth taut string approximations usually decrease the mean squared error.

3.2 Spectroscopy data

In Figure 1 we analysed briefly data from Spectroscopy and an approximation obtained from spline smoothing with the smoothing parameter chosen manually. If the parameter is determined automatically with cross validation the approximation becomes even more wiggly and nearly interpolates the data leading to 2512 local extreme values. The kernel and local polynomial estimators used in the simulation study above need 760 and 876 local extreme values and the wavelet procedure leads to a modality of 284. The taut string and smooth taut string procedures on the other only hand exhibit 82 local extreme values and the chemists could confirm that these extreme values matched exactly the true peaks. Figure 8 shows in the upper panels the approximations obtained from the usual and smooth taut string procedures. Two excerpts of the smooth taut string approximation are shown in the bottom panels and demonstrate that the smooth taut string is able to produce a simple approxi-

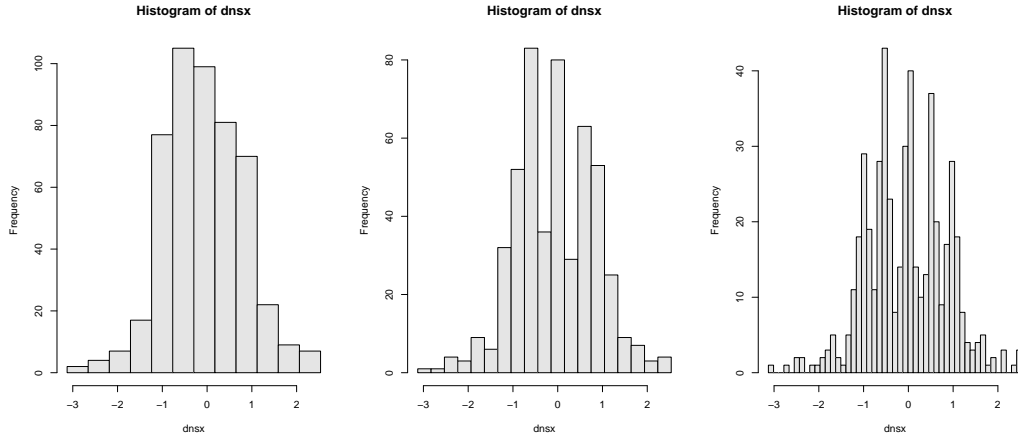


Fig. 9. Three histograms from the same sample, but with different bin widths.

mation in areas without any signal, but at the same time capture all the peaks in regions with more variability in the signal.

4 Further Applications

4.1 Density estimation

Another problem in Statistics where local extreme values often play an important rôle is density estimation. Figure 9 shows three histograms from the same sample, but with different bin widths. The histogram in the left panel seems to suggest that the sample is unimodal whereas the other histograms indicate the presence of three or five modes. Like in the regression case the problem is to decide which modes arise from the data. So given a sample x_1, \dots, x_n our task is again to specify a smooth density function f which is as simple as possible, but such that the corresponding distribution F is an adequate approximation for the data. As in Davies and Kovac (2004) we consider adequacy measures based on generalised Kuiper metrics d_{ku}^κ of order κ defined by κ by

$$d_{ku}^\kappa(F, G) = \max \left\{ \sum_1^\kappa |(F(b_j) - F(a_j)) - (G(b_j) - G(a_j))| \right\} \quad (6)$$

where the maximum is taken over all a_j, b_j with

$$a_1 \leq b_1 \leq a_2 \leq b_2 \cdots \leq a_\kappa \leq b_\kappa.$$

Since the distribution of $d_{ku}^\kappa(F_n, F)$ is independent of F for continuous F , it makes sense to denote the α -quantile by $\text{qu}(n, \alpha, d_{ku}^\kappa)$. Then an adequate approximation to the sample x_1, \dots, x_n is a density f whose distribution F

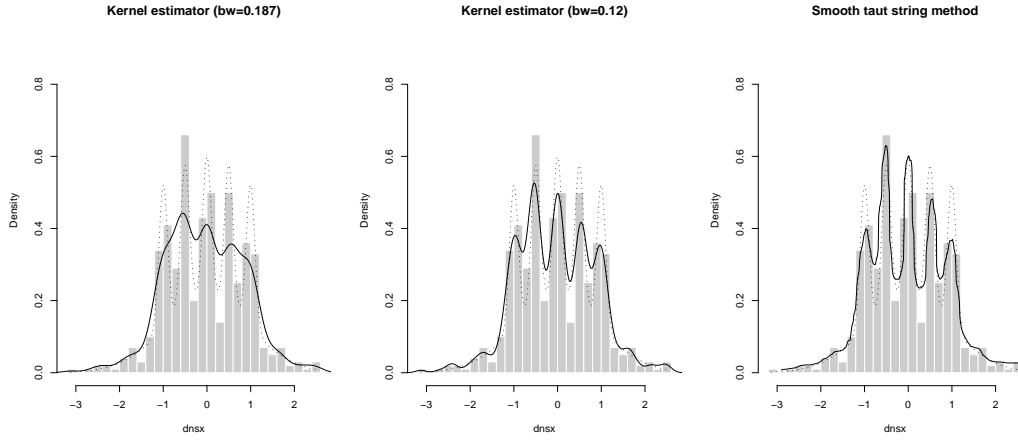


Fig. 10. Histogram of a sample from the claw distribution with two kernel estimates for two different bandwidths in the left and middle panel and smooth taut string approximations in the right panels. The dotted line displays in each plot the true claw density.

satisfies

$$d_{ku}^k(E_n, F) \leq \text{qu}(n, \alpha, d_{ku}^k).$$

where E_n is the empirical distribution.

Again a version of the taut string method can be used to generate a sequence of approximations with increasing complexity. Although not explicitly mentioned in Davies and Kovac (2004) the taut string approximations for density estimation can be regarded as a minimiser of a functional

$$T(f) = \sum_{i=1}^{n-1} (x_{i+1} - x_i) \left((x_{i+1} - x_i) f_i - \frac{1}{n-1} \right)^2 + \sum_{i=1}^{n-1} \lambda_i g_i(f_{i+1} - f_i)$$

where $g_i(x) = |x|$. Again by choosing $g_i(x) = \sqrt{(x_{i+1} - x_i)^2 + x^2}$ we obtain smooth solutions and can run the same algorithm as set out above for the regression case.

The sample that was used in Figure 9 consists of 500 observations from the claw distribution which has been studied by Marron and Wand (1992) and several other authors. The claw density is shown as dotted lines in each of the three panels of Figure 10. In the left panel a kernel estimate is displayed as well. Although difficult to see the kernel estimate has exactly five modes at positions approximately -3.1, -2.3, -0.6, 0 and 0.5. Any smaller bandwidth will enlarge the artefacts at -3.1 and -2.3. In particular for bandwidths that allow the fingers of the claw density to be approximated reasonably well, at the same time a number of artificial extrema at the tails are produced. This problem

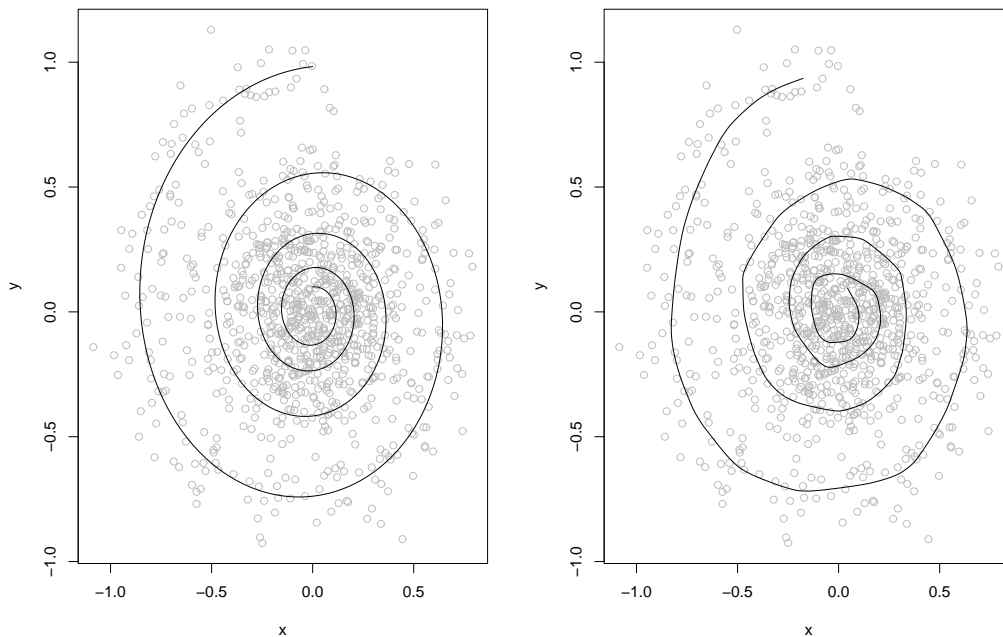


Fig. 11. Spiral with added two-dimensional Gaussian noise. Left panel shows the true spiral, right panel shows approximation derived from two-dimensional smooth taut string functional.

is illustrated in the middle panel of Figure 10. and has been studied in more detail by Davies and Kovac (2004). The smooth taut string approximation on the other hand attains exactly the correct modality and is displayed in the right panel of Figure 10.

4.2 Two-dimensional curves

The left panel of Figure 11 shows a spiral with added two-dimensional Gaussian noise. This setting provides another interesting application of the smooth taut string method. We assume that two-dimensional observations (x_i, y_i) are made at time points t_i and are looking for a function $f = (f_x, f_y)$ such that the two-dimensional residuals $(x_i - f_x(t_i), y_i - f_y(t_i))$ look like noise. A straightforward idea to provide approximations in this situation is to consider the functional

$$\begin{aligned}
 T(f) = & \sum_{i=1}^n (y_i - f_i^y)^2 + \sum_{i=1}^n (x_i - f_i^x)^2 \\
 & + \sum_{i=1}^{n-1} \lambda_i \sqrt{(f_{i+1}^y - f_i^y)^2 + (f_{i+1}^x - f_i^x)^2}
 \end{aligned}$$

The right panel of Figure 11 shows an approximation derived from this functional.

5 Proofs

5.1 Proof of Lemma 1

Since T is convex and differentiable, there exists a unique minimum. Denote the minimiser by \tilde{f} . Then $\frac{\partial T(\tilde{f})}{\partial \tilde{f}_i} = 0$ for all $i = 1, \dots, n$. By adding the partial derivatives with respect to f_j, \dots, f_k we obtain (i). The special case with $j = 1$ and solving for f_{k+1} yields (ii). Let f be some vector that satisfies the equations (4) and $f_i > \tilde{f}_i$ for all $i = 1, \dots, j$, then in particular $\sum_{i=1}^j (f_i - y_i) > \sum_{i=1}^j (\tilde{f}_i - y_i)$. This together with (ii) show (iii) since g' is monotonically increasing.

5.2 Proof of Lemma 3

For $k < n$ define a direction vector ε by $\varepsilon_i = 1\{i \leq k\}, i = 1, \dots, n$ and consider the function $h(t) = T(\tilde{f} + t\varepsilon)$. Then

$$h'(0) = 2 \sum_{i=1}^k (f_i - y_i) - \lambda_k g'_k(\tilde{f}_{k+1} - \tilde{f}_k).$$

To ease notation we assume that $\tau_k = 1$, thus $\tilde{f}_{k+1} \geq \tilde{f}_k$.

We consider the case first where $\sum_{i=1}^k (\tilde{f}_i - y_i) \leq 0$. Assume that $\tilde{f}_{k+1} > \tilde{f}_k$, then $h'(0) < 0$. Then for sufficiently small $\gamma > 0$ and all $t \in (0, \gamma)$ the vectors $\tilde{f} + t\varepsilon$ would still satisfy the monotonicity constraints and $h(t) < h(0)$, thus \tilde{f} would not minimise T . Thus $\tilde{f}_{k+1} = \tilde{f}_k$.

If on the other hand $\sum_{i=1}^k (\tilde{f}_i - y_i) > 0$, but $\tilde{f}_{k+1} - \tilde{f}_k < (g'_k)^{-1} \left(\frac{2}{\lambda_k} \sum_{i=1}^k (\tilde{f}_i - y_i) \right)$, then $h'(0) > 0$. Since $\tilde{f} + t\varepsilon$ satisfies the monotonicity constraints for all $t \leq 0$, we could again find some $q < 0$ such that $h(q) < h(0)$. Similarly we conclude in the case where $\tilde{f}_{k+1} - \tilde{f}_k > (g'_k)^{-1} \left(\frac{2}{\lambda_k} \sum_{i=1}^k (\tilde{f}_i - y_i) \right)$. This proves (i).

For (ii) assume that $f_i \geq \tilde{f}_i$ for all $i = 1, \dots, k$. If $\tau_k = 1$ and $\sum_{i=1}^k (\tilde{f}_i - y_i) > 0$, then because of $\sum_{i=1}^k (f_i - y_i) \geq \sum_{i=1}^k (\tilde{f}_i - y_i)$ and monotonicity of g' also $f_{k+1} > \tilde{f}_{k+1}$. If $\tau_k = 1$ and $\sum_{i=1}^k (\tilde{f}_i - y_i) \leq 0$, then $f_{k+1} \geq f_k \geq \tilde{f}_k = \tilde{f}_{k+1}$. If $\tau_k = -1$ and $\tau_k \sum_{i=1}^k (f_i - y_i) > 0$, then also $\tau_k \sum_{i=1}^k (\tilde{f}_i - y_i) > 0$ and therefore again because of monotonicity of g' we find that $f_{k+1} > \tilde{f}_{k+1}$. Finally if $\tau_k = -1$ and $\tau_k \sum_{i=1}^k (f_i - y_i) \leq 0$ we conclude that $f_{k+1} = f_k > \tilde{f}_k \geq \tilde{f}_{k+1}$.

The computational complexity of the algorithm depends on how often the recursion formula (4) is applied. These applications can be divided into two groups distinguishing between those corresponding to choices of k and f_k with $|f_k - \tilde{f}_k| > \varepsilon/2$ and $|f_k - \tilde{f}_k| \leq \varepsilon/2$. We will see that there are $O(n)$ combinations of the first type, but that for each combination the number of applications of (4) is bounded by some $L_n = O(1/\log(1 + \mu_n))$. The second case occurs for every value of k only up to once. Moreover if it does occur for some $k = \tilde{k}$ and if recursion formula (4) is applied for $M > L_n$ times, then there are no further applications of (4) with $|f_k - \tilde{f}_k| \leq \varepsilon/2$ for $k = \tilde{k}, \tilde{k} + 1, \dots, \tilde{k} + M - L_n$. So the computational complexity of applying (4) in the second case is again of order $O(nL_n)$.

More precisely assume that $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_{k-1}$ have already been calculated for some k and that $f_k > \tilde{f}_k + \varepsilon/2$. Then using a Taylor expansion and Lemma 1 we see that for all $j > k$ such that f_j can be calculated using the recursion formula (4)

$$\begin{aligned} f_j - \tilde{f}_j &\geq f_{j-1} - \tilde{f}_{j-1} + \left(\frac{2}{\lambda_{j-1}} \sum_{i=k}^{j-1} (f_i - \tilde{f}_i) \right) \left((g'_{j-1})^{-1} \right)' \left(\frac{2}{\lambda_{j-1}} \sum_{i=1}^{j-1} (\tilde{f}_i - y) \right) \\ &\geq (f_{j-1} - \tilde{f}_{j-1})(1 + C\mu_n) \end{aligned}$$

for some $C > 0$. Applying this idea several times we see that $f_j - \tilde{f}_j \geq \frac{\varepsilon}{2}(1 + C\mu_n)^{j-k}$ and therefore the recursion formula (4) stops after at most L_n operations where $L_n = O(1/\log(1 + C\mu_n))$.

The number of executions of steps (III) and (IV) of the algorithm for a particular k is bounded by a global constant M that does not depend on k like $M = \log_2((\max(u) - \min(l))/\varepsilon)$ where u and l are the crude initial bounds. Thus the number of applications of (4) corresponding to choices of k and f_k with $|f_k - \tilde{f}_k| > \varepsilon/2$ is bounded by $L_n M n$.

If on the other hand $0 \leq f_k - \tilde{f}_k \leq \varepsilon/2$ for some k and f_k and (4) gets applied $N > L_n$ times, then $0 \leq f_j - \tilde{f}_j \leq \varepsilon/2$ for all $j = k, k + 1, \dots, M := k + N - L_n - 1$. Step (IV) of the algorithm will then move the upper bounds u_k, u_{k+1}, \dots, u_M so closely to $\tilde{f}_k, \tilde{f}_{k+1}, \dots, \tilde{f}_M$ that in subsequent iterations for $k \leq M$ the starting value f_k always satisfies $|f_k - \tilde{f}_k| < -\varepsilon/2$. We conclude similarly in the case where $f_k - \tilde{f}_k \leq \varepsilon/2$ and see that the total number of applications of (4) corresponding to choices of k and f_k with $|f_k - \tilde{f}_k| \leq \varepsilon/2$ is bounded by $L_n n$.

6 Conclusions

We have introduced the smooth taut string method that provides smooth and simple approximations in the context of non-parametric regression without spurious local extreme values. At the same time the new method enjoys a small mean squared error when compared to other smoothing methods in particular when the underlying signal is rough.

Software to this article is made available as part of the `ftnonpar` package for R.

References

- Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations (with discussion). *J. Amer. Statist. Assoc.*, 96:939–967.
- Brockmann, M., Gasser, T., and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88:1302–1309.
- Davies, P. L. (1995). Data features. *Statistica Neerlandica*, 49:185–245.
- Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). *Annals of Statistics*, 29:1–65.
- Davies, P. L. and Kovac, A. (2004). Densities, spectral densities and modality. *Annals of Statistics*, 32:1093–1136.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G., and Picard, D. (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society*, 57:371–394.
- Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society B*, 57:371–394.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.
- Green, P. and Silverman, B. (1994). *Nonparametric regression and Generalized Linear Models: a roughness penalty approach*. Number 58 in Monographs on Statistics and Applied Probability. Chapman and Hall, London.
- Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.*, 31:1600–1635.
- Kovac, A. and Meise, M. (2006). Minimizing total variation under multiresolution constraints. Technical Report, University of Bristol.
- Marron, J. and Wand, M. (1992). Exact mean integrated squared error. *Annals of Statistics*, 20:712–736.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 10:186–190.
- Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth

- selector for local least squares regression. *Journal of the American Statistical Association*, 90:1257–1270.
- Seifert, B., Brockmann, M., Engel, J., and Gasser, T. (1994). Fast algorithms for nonparametric curve estimation. *Journal of Computational and Graphical Statistics*, 3:192–213.
- Silverman, B. W. (1985a). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society*, 47:1–52.
- Silverman, B. W. (1985b). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society*, 47:1–52.
- van de Geer, S. (2001). Least squares estimation with complexity penalties. *Mathematical Methods of Statistics*, 10:355–374.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā*, 26:101–116.