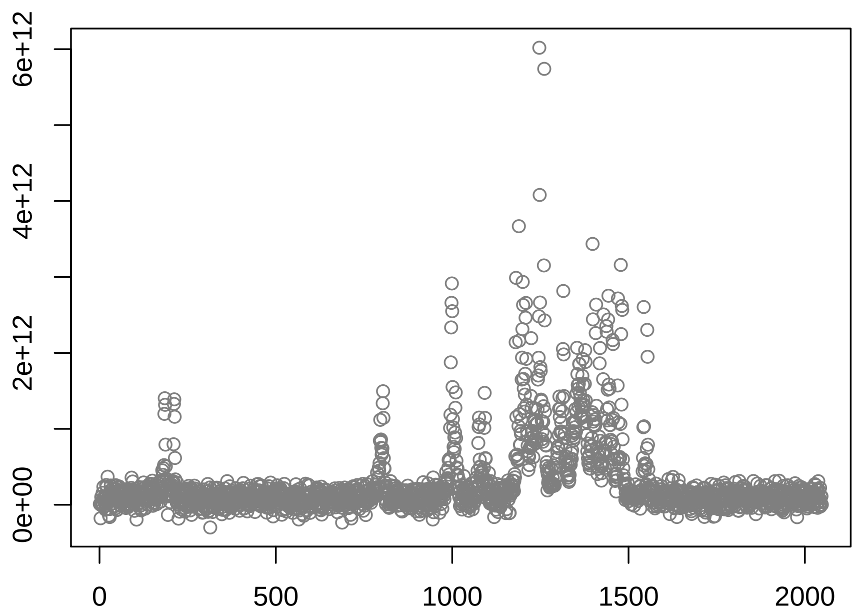
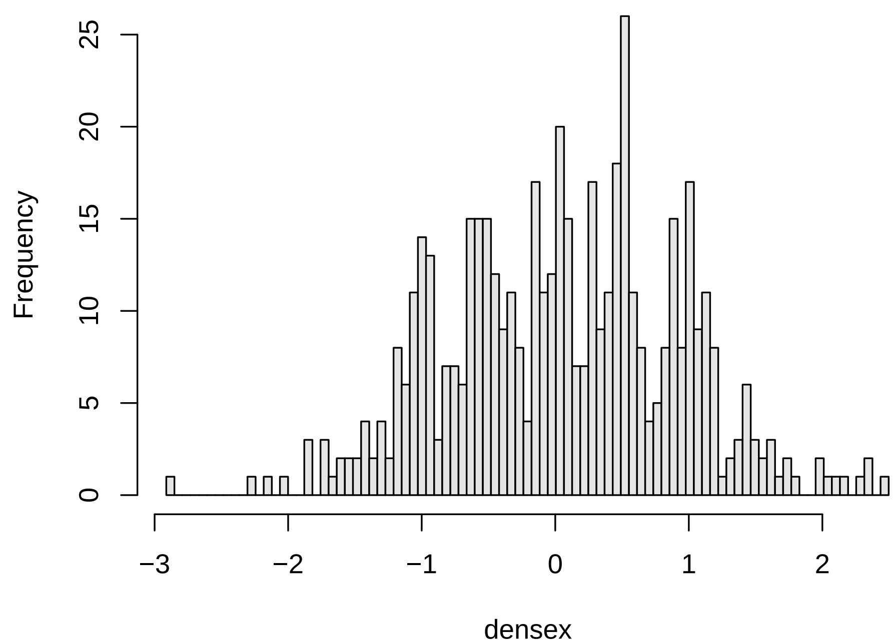
Local extreme values

Arne Kovac

University of Bristol





Overview

1. **Local extreme values and classical methods**
2. **Isotonic regression and the pool-adjacent violator algorithm**
3. **Total variation penalties and the taut string method**
4. **The taut string method and modality**
5. **The multiresolution criterion and global and local squeezing**
6. **The smooth taut string method**
7. **Minimising total variation and a glimpse at two-dimensional problems**
8. **Density estimation**

Local extreme values

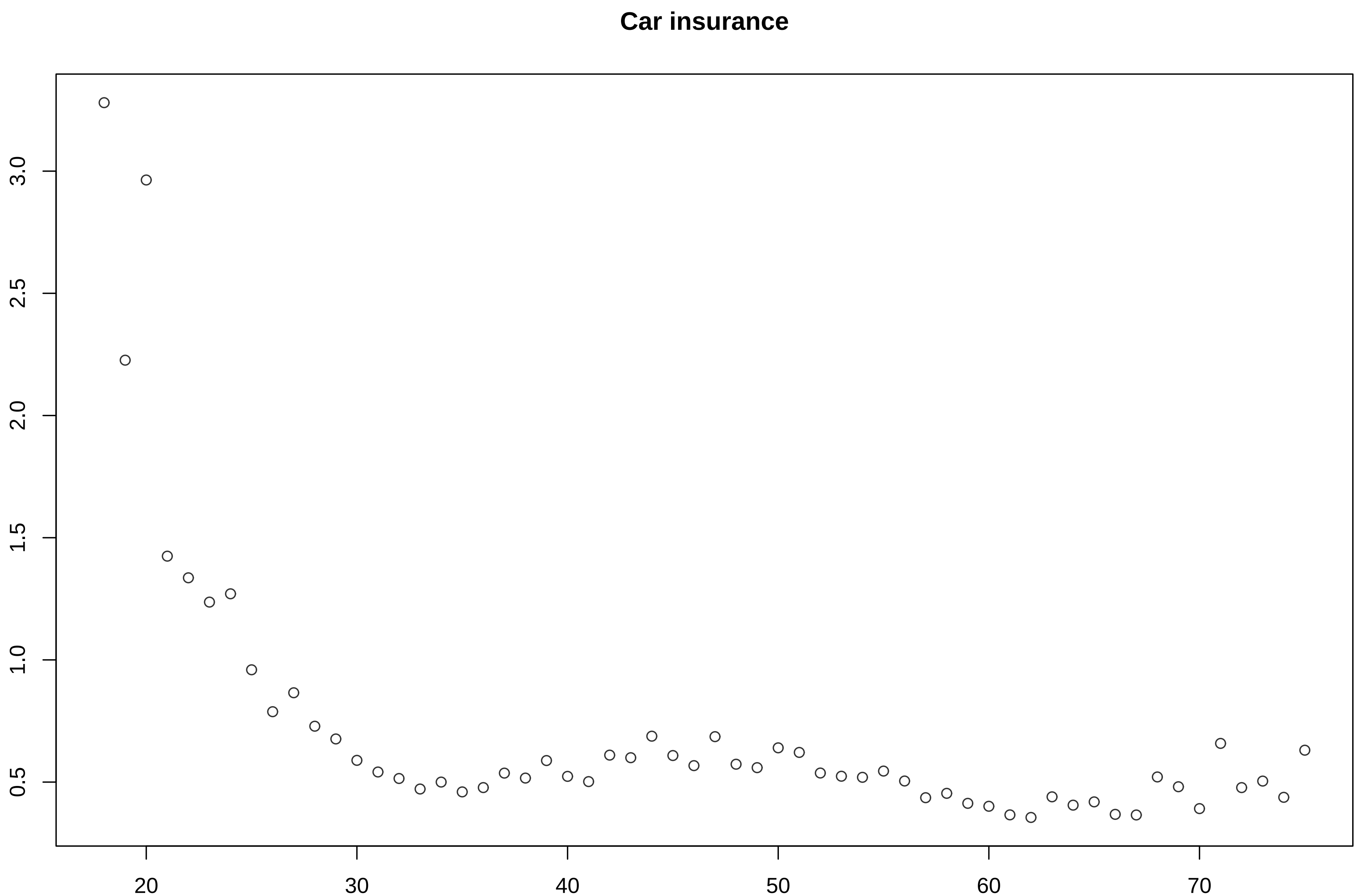
Why interested in local extreme values?

Local extreme values

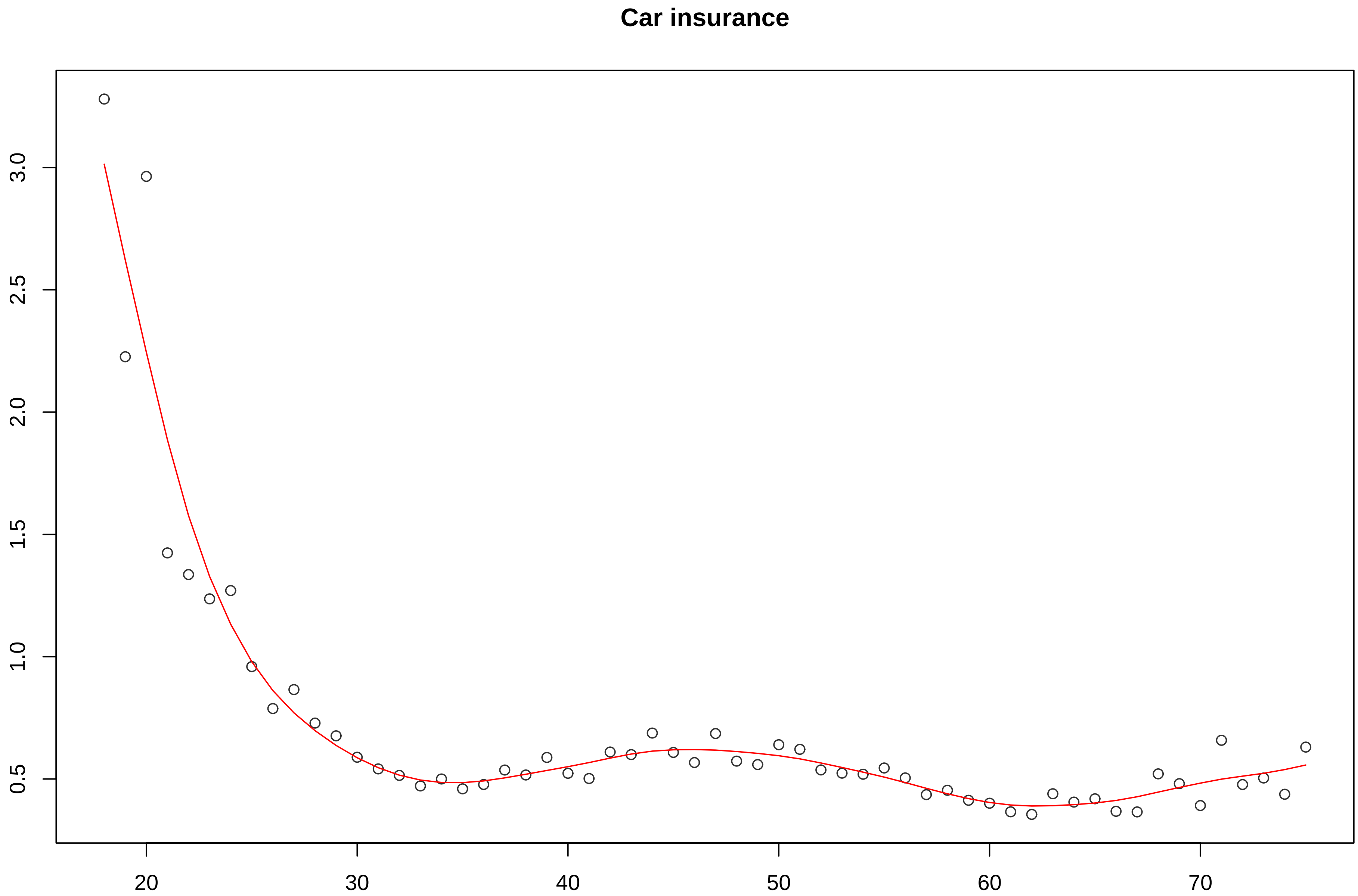
Why interested in local extreme values?

- Local extrema often have an easy and natural interpretation
- In some applications local extrema are subject of interest
- Approximations ($\approx$ estimates) with superfluous extrema do not look nice
- ...

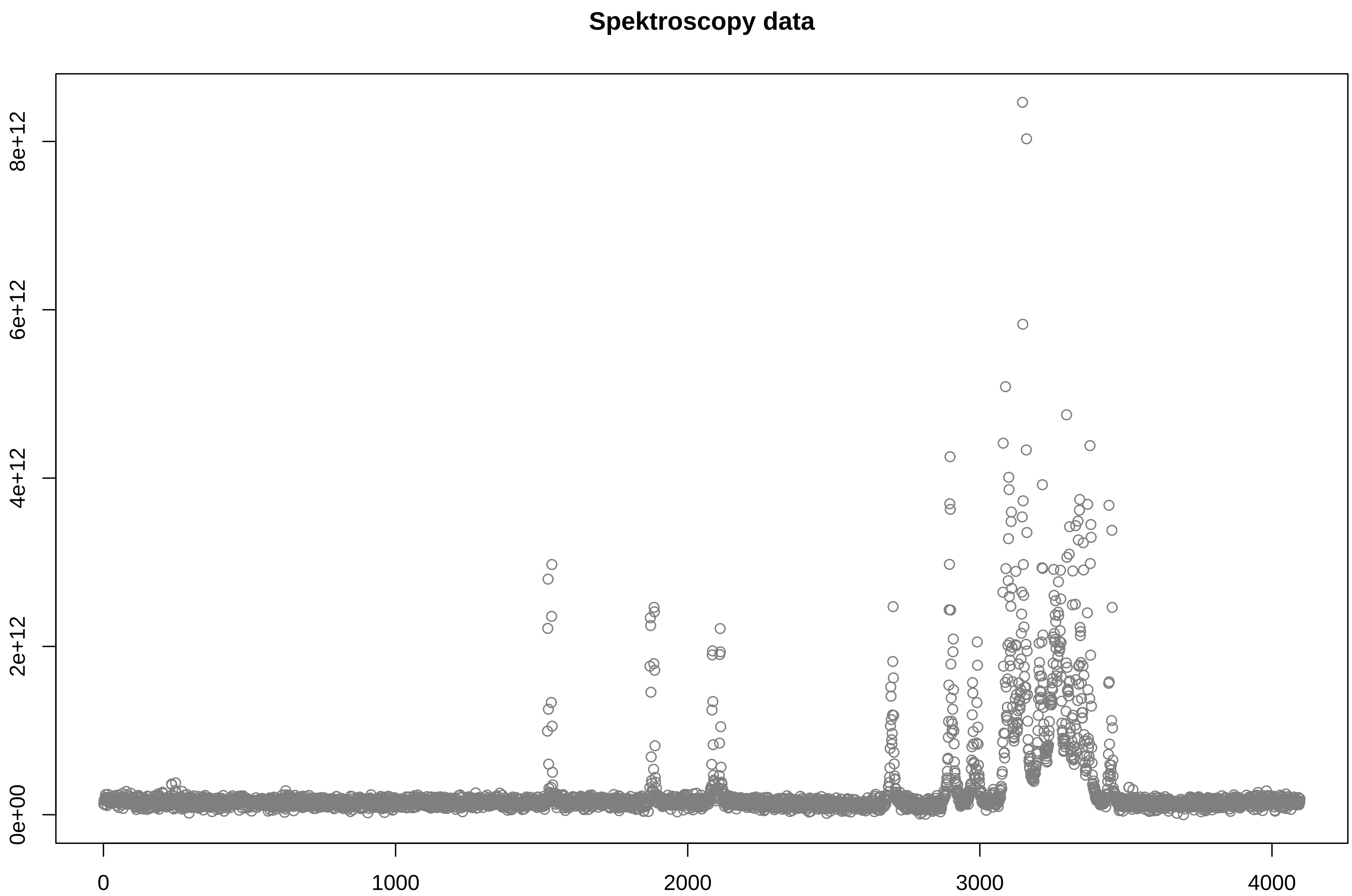
Natural interpretation of extrema



Natural interpretation of extrema



Subject of interest



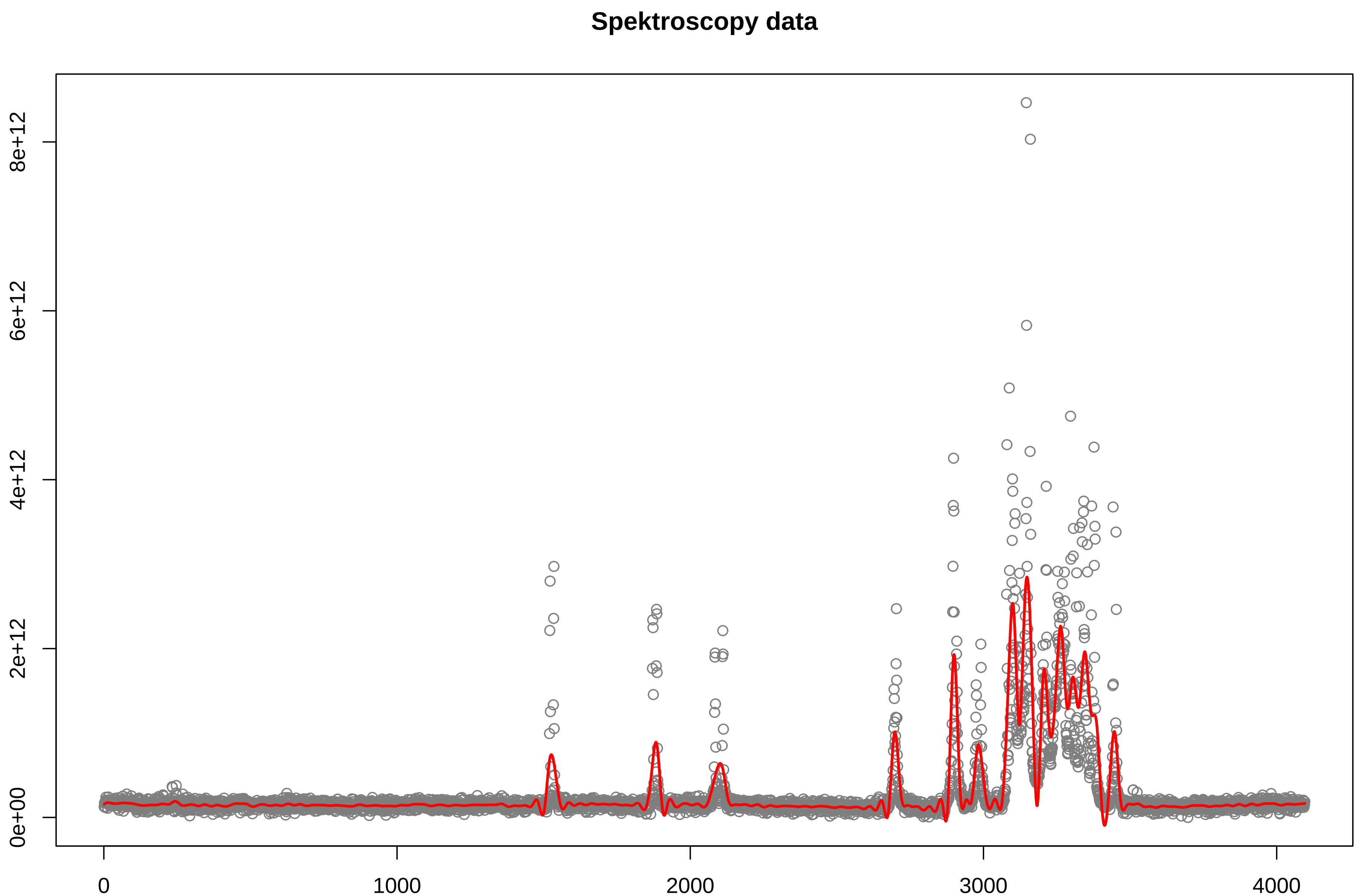
Subject of interest

Data from Spectroscopy:

- Peaks indicate presence of certain structures in substance
- Position: type of structure
- Power: number of structures

Wanted: Automatic procedure which returns positions and power of each peak.

Subject of interest



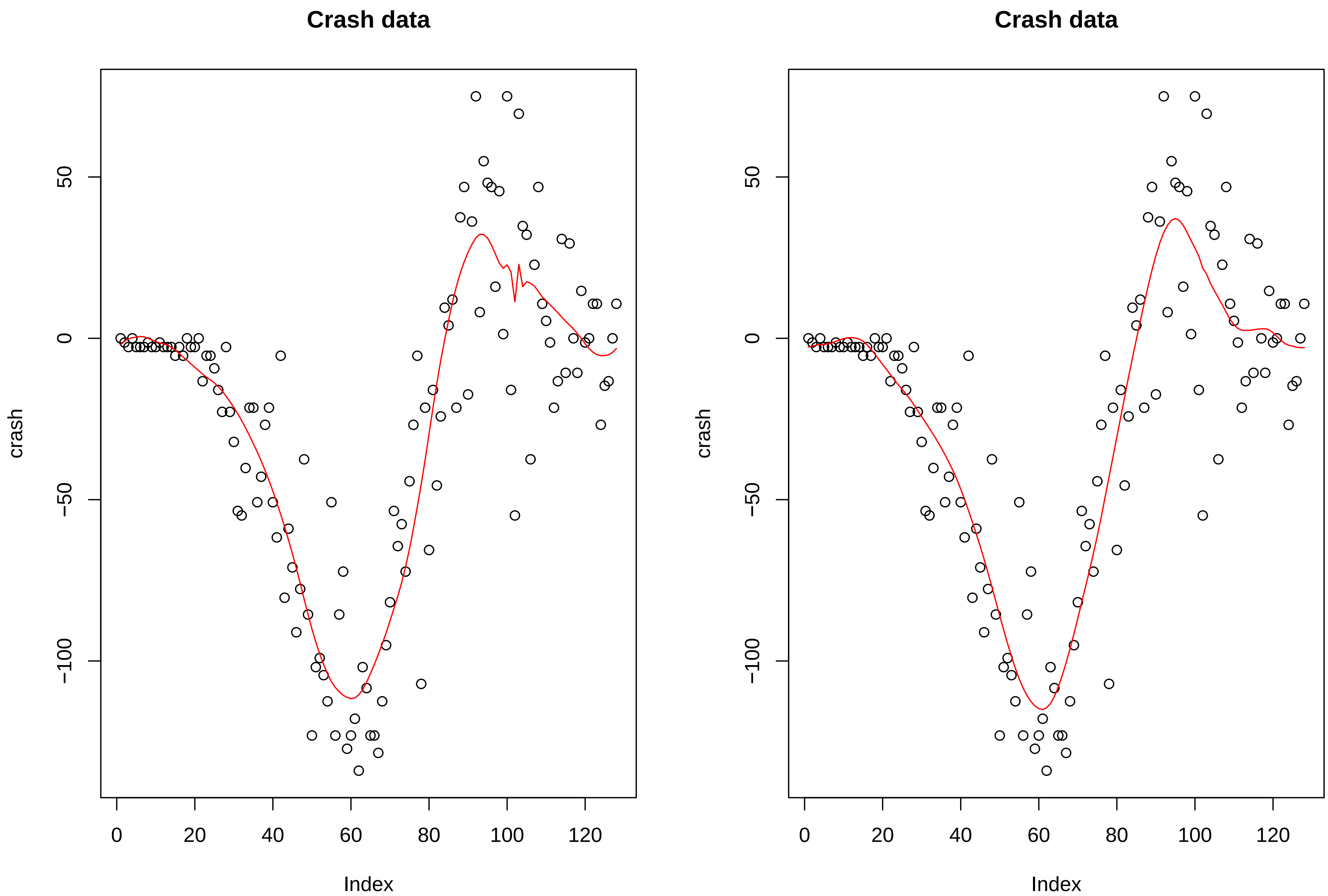
Nice property of data from Spectroscopy

In Statistics often one either

- works with real data and does not know the 'true' signal or
- knows the 'true' signal because the data are simulated.

Spectroscopy data are real data and for some spectroscopy methods noise can be eliminated, it just takes a while. . .

Superfluous extrema



Which one would you prefer?

Superfluous extrema

Kovac's axiom of simplicity:

Even if some paper is not about local extreme values, once it comes to examples, authors will prefer estimates without superfluous extrema.

Local extreme values

Why interested in local extreme values?

- Local extrema often have an easy and natural interpretation
- In some applications local extrema are subject of interest
- Approximations ($\approx$ estimates) with superfluous extrema do not look nice
- Prior knowledge about shape behaviour (monotonicity constraints, convexity constraints etc.)
- Rates of convergence
- . . .

Classical methods and extreme values

We look at three classical smoothing techniques:

- Kernel estimators
- Spline smoothing
- Wavelet thresholding

and see how they perform with respect to local extreme values.

Kernel estimator

Estimate $\hat{f}(t)$ by an weighted average of the given data in a small window centred around t:

$$\hat{f}(t) = \frac{\sum_{i=1}^{n} y_i K(\frac{t-t_i}{\lambda})}{\sum_{i=1}^{n} K(\frac{t-t_i}{\lambda})},$$

Crucial: Choice of bandwidth.

```
> plot(djbumps,col="grey")
> lines(ksmooth(1:2048,djbumps,band=10),col="red")
```

Spline smoothing

Trade-off between goodness-of-fit and smoothness.

Estimate $\hat{f}$ as the twice differentiable function that minimises

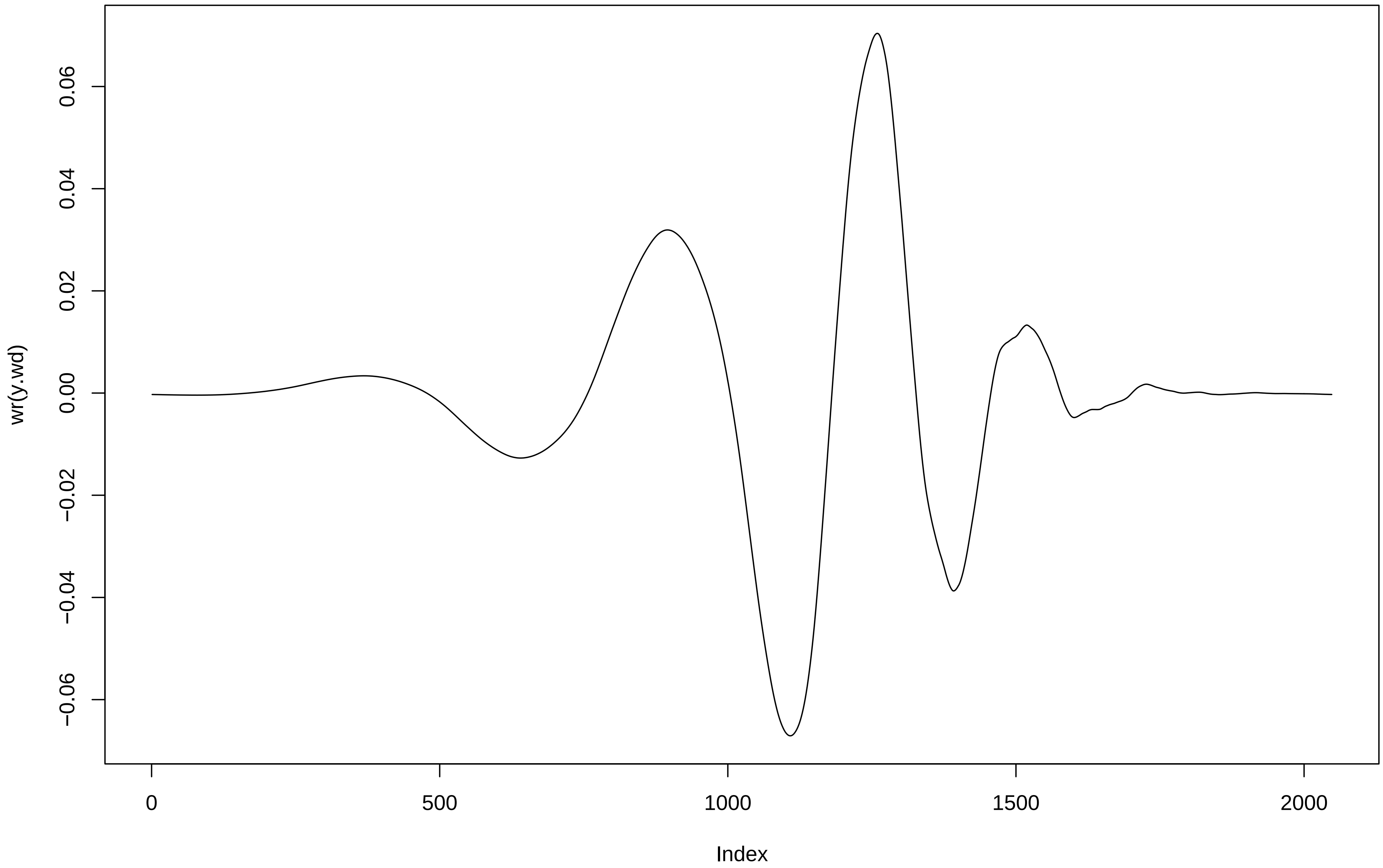
$$S(\hat{f}) = \frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{f}(t_j))^2 + \lambda \int \hat{f}''(x)^2 dx.$$

Solution is cubic spline with knots $(t_i)_{i=1}^{n}$, so that $\hat{f}$ is a cubic polynomial between any two neighbouring time points t_i and t_{i+1}.

Parameter λ controls smoothness of regression function.

```
> plot(djbumps,col="grey")
> lines(smooth.spline(djbumps,spar=0.3),col="red")
```

Wavelet thresholding



$$\hat{f}(t) = \sum_{j,k} w_{j,k}\psi_{j,k}(t), \quad \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k).$$

Wavelet thresholding

- Apply discrete wavelet transform to data

$$w = \mathcal{W}y.$$

- Threshold the wavelet coefficients:

$$w^*_{j,k} = \text{sgn}(w_{j,k})(|w_{j,k}| - \tau)_+,$$

- Estimate f by inverse DWT:

$$\hat{f}_n(t_i)_1^n = \mathcal{W}^T w^*.$$

Wavelet thresholding

```
> djbumps.wd <- wd(djbumps,filter.number=5)
> djbumps.thresh <- threshold(djbumps.wd,dev=
            mymadmad,policy="manual",value=4)
> djbumps.wr <- wr (djbumps.thresh)
> plot(djbumps,col="grey")
> lines(djbumps.wr,col="red")
```

Identifying and eliminating

One possibility (Silverman, 1986; Chaudhuri and Marron, 1999; many other):

- Smoothing methods remain unchanged
- For each local extreme of their output it is decided whether it comes from the underlying signal or not.

Afterwards in can be proceeded in two ways:

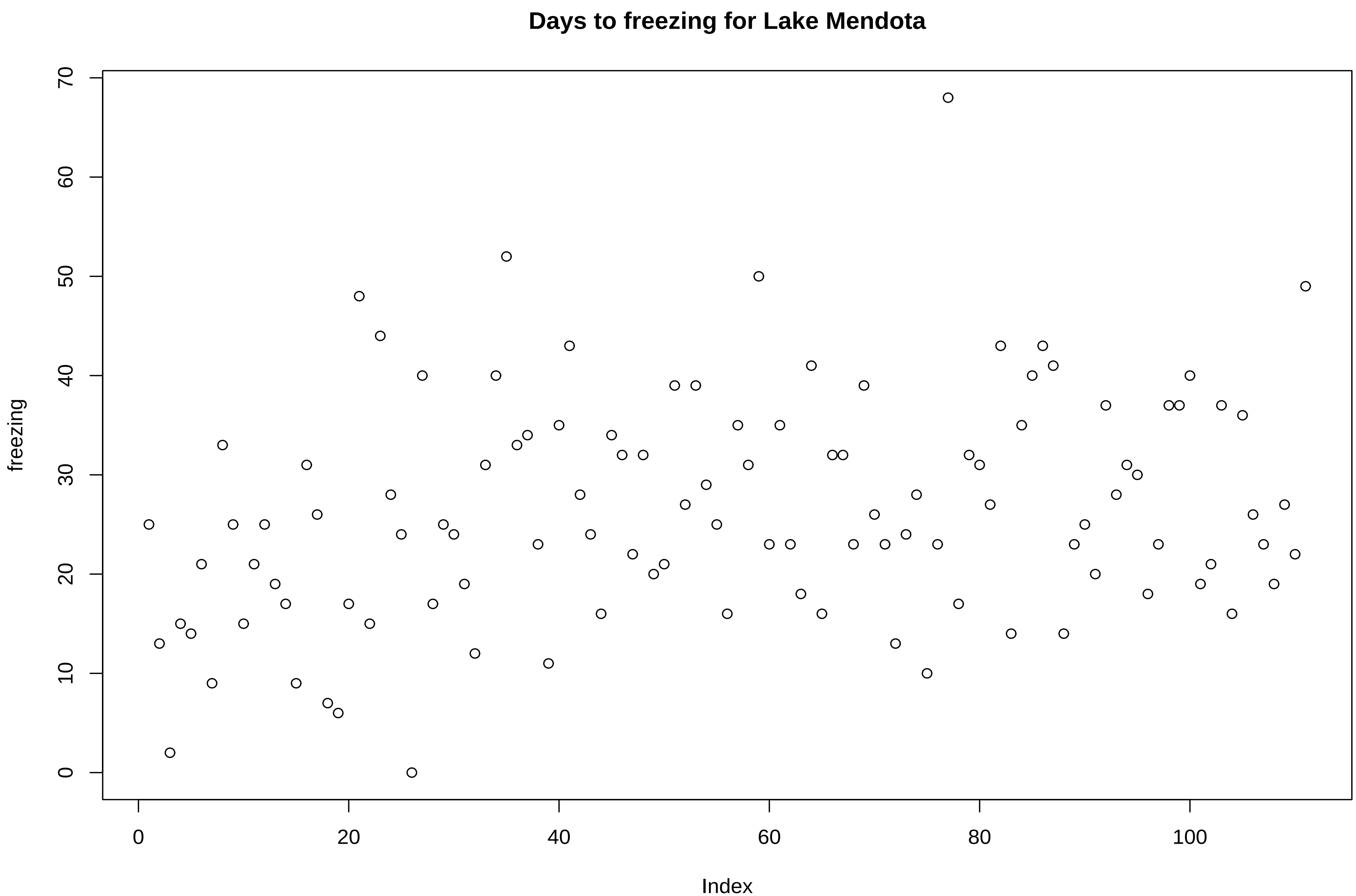
- Choose the smoothing parameter such that only features that arise from the underlying signal are included.
- Eliminate artifacts by projecting subintervals on the space of monotone functions.

Identifying and eliminating

Problems:

- No choice for smoothing parameter might reveal a function with correct modality
- Projecting on space of monotone functions means cutting off extreme values, looks artificial
- Tests based on given estimate not reliable.

Freezing dates



Linear regression

Let $X = \{x_1, \ldots, x_k\}$ where $x_1 < \cdots < x_k$ and $y_j(x_i), j = 1, \ldots, m(x_i)$ observations of a distribution with mean $\mu(x_i)$.

If μ is assumed to be linear in x: linear regression:

$$\sum_{x \in X} \sum_{j=1}^{m(x)} (y_j(x) - f(x))^2 = \min$$

in the class of linear functions which is equivalent to minimise

$$\sum_{x \in X} (\bar{y}(x) - f(x))^2 m(x) = \min, \quad \bar{y}(x) = \sum_{j=1}^{m(x)} y_j(x)/m(x)$$

among all linear functions.

Isotonic regression

A real valued function f on X is *isotonic* if $x < y$ implies $f(x) \leq f(y)$.

An isotonic function g^* on X is an *isotonic regression of* g *with weights* w if it minimises in the class of isotonic functions f

$$\sum_{x \in X} (g(x) - f(x))^2 w(x) = \min$$

Cumulative sum diagram

Cumulative sums

$$G_j = \sum_{i=1}^{j} g(x_i)w(x_i), W_j = \sum_{i=1}^{j} w(x_i), j = 1, 2, \ldots, k.$$

Points $P_j = (W_j, G_j)$ constitute *cumulative sum diagram* (CSD). The *slope* of the segment joining P_{j-1} to P_j is just $g(x_j)$, the slope of the chord joining P_{i-1} to $P_j, (i \leq j)$ is weighted average

$$\mathrm{Av}\{x_i, \ldots, x_j\} = \sum_{r=i}^{j} g(x_r)w(x_r) / \sum_{r=i}^{j} w(x_r).$$

Greatest convex minorant

Isotonic regression of $g =$ slope of the greatest convex minorant (GCM) of the CSD.

This is the graph of the supremum of all convex functions whose graphs lie below the CSD.

Graphically, the GCM is the path along which a taut string lies if it joins P_0 and P_k and is constrained to lie below the CSD.

Some properties

CSD and GCM coincide at P_k, ie $G_k^* = G_k$.

If for some index i the GCM at P_{i-1}^* lies strictly below the CSD at P_{i-1}, then the slopes of the GCM entering P_{i-1}^* from the left and leaving to the right are the same:

$$G_{i-1}^* < G_{i-1} \Rightarrow g_i^* - g_{i-1}^* = 0, i = 1, 2, \ldots, k.$$

If $P_r = P_r^*, P_s = P_s^*, P_t = P_t^*$ for $r < s < t$, then for all $r < j < s$ the slope of $P_j P_s$ is smaller than the slope of $P_j^* P_s^*$. If $s < j < t$, then the slope of $P_s P_j$ is larger than the slope of $P_j^* P_s^*$.

The GCM and isotonic regression

Theorem: The slope g^* of the GCM furnishes the isotonic regression of g. Indeed, if f is isotonic on X then

$$\sum_{x \in X} (g(x) - f(x))^2 w(x) \geq \sum_{x \in X} (g(x) - g^*(x))^2 w(x) + \sum_{x \in X} (g^*(x) - f(x))^2 w(x).$$

The isotonic regression is unique.

Proof makes use of *partial summation formula (Abel's lemma):* Suppose $\{u_k\}$ and $\{v_k\}$ are two sequences. Then

$$\sum_{k=m}^{n} u_k(v_{k+1} - v_k) = u_{n+1}v_{n+1} - u_m v_m - \sum_{k=m}^{n} u_{k+1}(v_{k+1} - v_k).$$

Pool-Adjacent Violators Algorithm

Idea: If for some i, $g(x_{i-1}) > g(x_i)$, then graph of the part of the GCM between points P^*_{i-2} and P^*_i is a straight line segment. Thus CSD could be altered by connecting P_{i-2} with P_i by a straight line segment without changing the GCM.

Pool-Adjacent Violators Algorithm

Idea: If for some i, $g(x_{i-1}) > g(x_i)$, then graph of the part of the GCM between points P^*_{i-2} and P^*_i is a straight line segment. Thus CSD could be altered by connecting P_{i-2} with P_i by a straight line segment without changing the GCM.

Isotonic regression g^* partitions X into sets on which it is constant, i.e. into *level sets* for g^*, called *solution blocks*. On each of these solution blocks g^* takes the weighted average of the values of g over the block, using weights w.

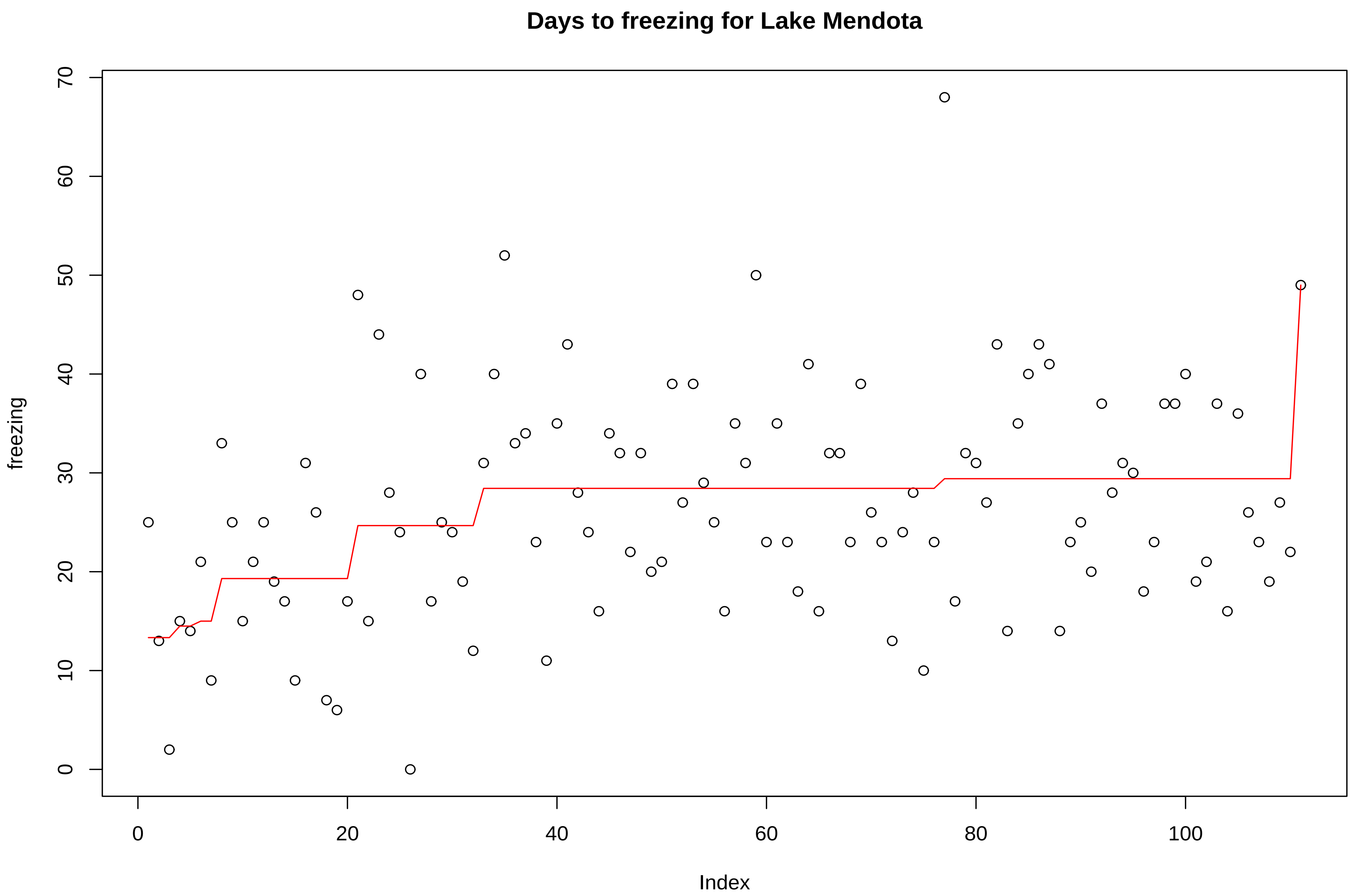
Pool-Adjacent Violators Algorithm

Idea: If for some i, $g(x_{i-1}) > g(x_i)$, then graph of the part of the GCM between points P^*_{i-2} and P^*_i is a straight line segment. Thus CSD could be altered by connecting P_{i-2} with P_i by a straight line segment without changing the GCM.

Isotonic regression g^* partitions X into sets on which it is constant, i.e. into *level sets* for g^*, called *solution blocks*. On each of these solution blocks g^* takes the weighted average of the values of g over the block, using weights w.

If $g(x_1) \leq g(x_2) \leq \cdots \leq g(x_k)$, then the initial partition is also final partition and $g^*(x_i) = g(x_i)$ for all i. If not select any of the pairs of violators such that $g(x_i) > g(x_{i+1})$ and pool these two values of g. Iterate until final partition reached.

Freezing dates revisited



Some notation

Observations $y_1, \ldots, y_n$ at time points $t_1 \leq \cdots \leq t_n$

Model:

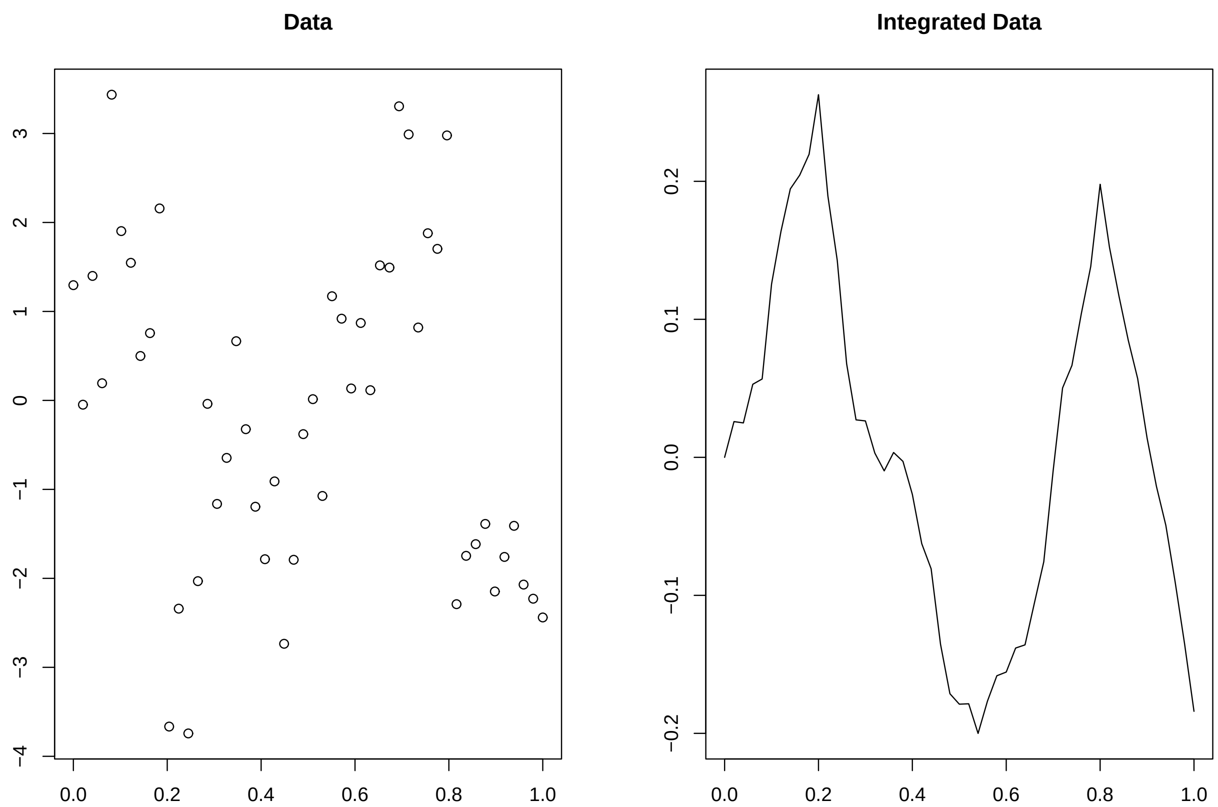
$$y_i = f(t_i) + \varepsilon_i.$$

Noise $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. $N(0, \sigma^2)$.

The taut string method

Integrated process with linear interpolation between design points

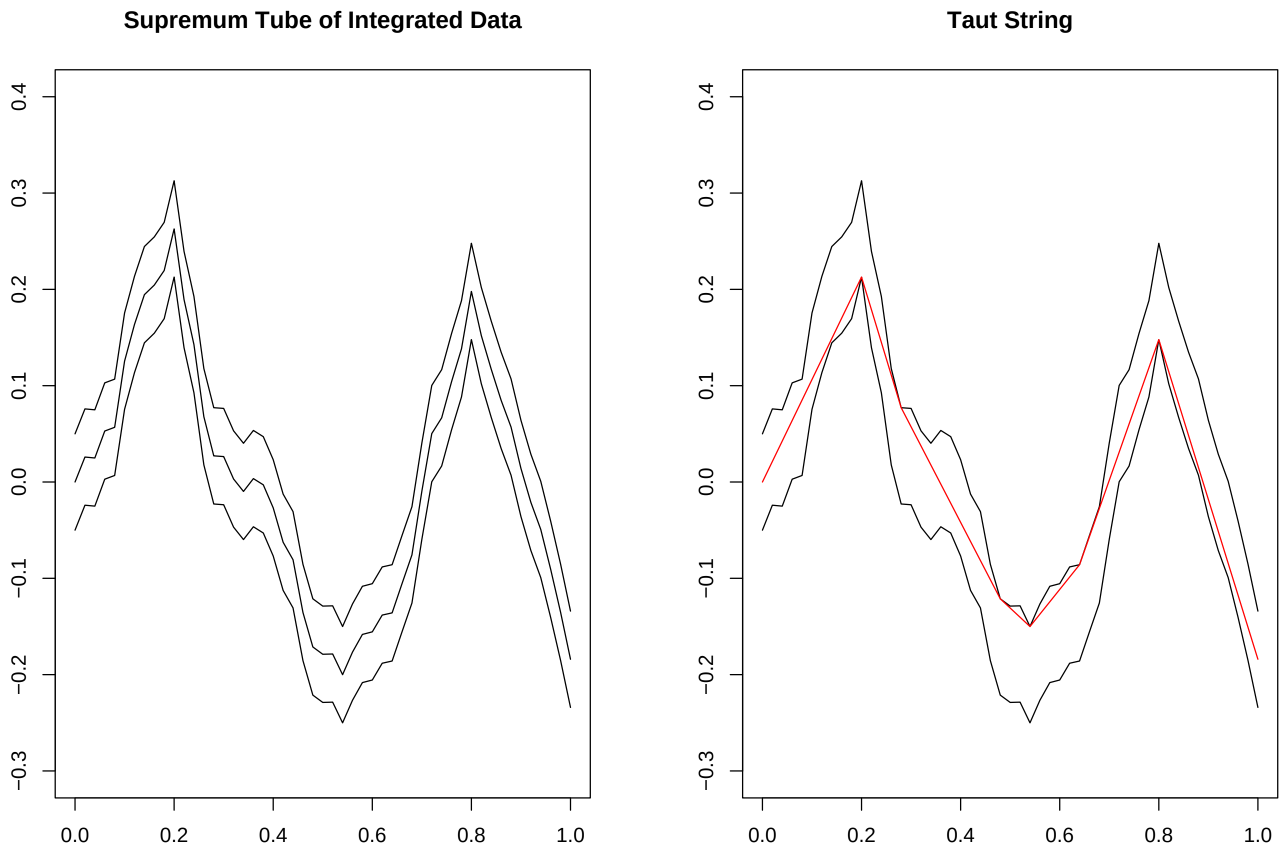
$$Y_0 := 0, \quad Y_j = \sum_{i=1}^{j} y_i \quad (j = 1, \ldots, n)$$



The taut string method

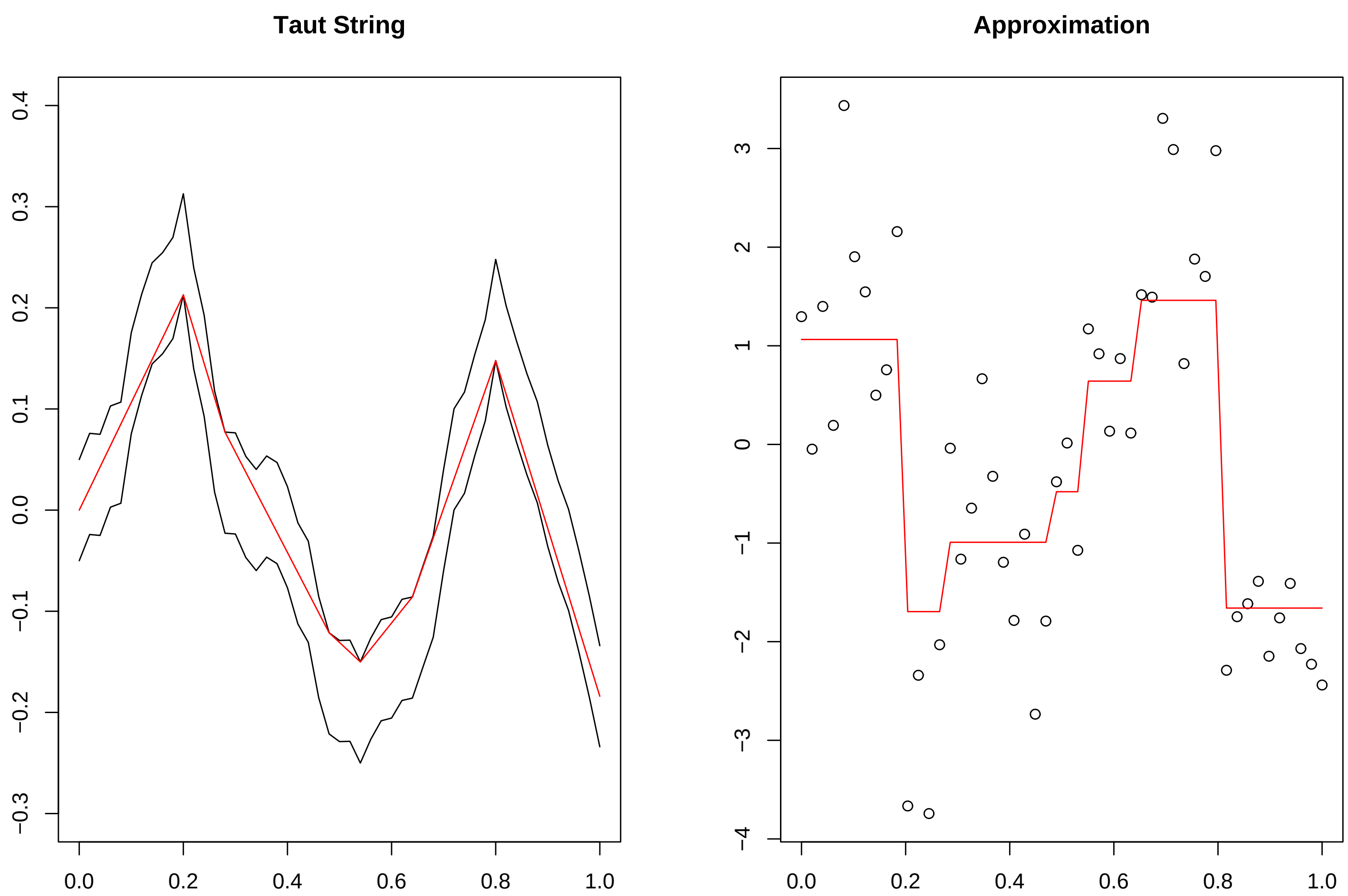
Tube $[Y - \lambda, Y + \lambda]$

A string inside the tube is tightened $\longrightarrow F^{\lambda}$.



The taut string method

Differentiating the taut string F^λ yields approximation f^λ.



Some history

- Isotonic least squares regression (Barlow et al, 1972)
- Test of unimodality (Hartigan and Hartigan, 1985)
- Density estimation (Davies, 1995; Davies and Kovac, 2004)
- Spectral densities (Davies and Kovac, 2004)
- Non-parametric regression (Mammen and van de Geer, 1997; Davies and Kovac, 2001; Kovac, 2006)

Properties of taut string

- Piecewise constant
- Calculation is possible in O(n) steps

Properties of taut string

- Piecewise constant
- Calculation is possible in O(n) steps

Some more properties we investigate in detail in next lecture:

- Modality increases monotonically with tube decreasing tube width
- Minimal modality among all functions such that the integral lies inside the tube.
- Asymptotic consistency of modality

ftnonpar package

Implemented in `R` as function `pmreg` in package `ftnonpar`:

```
> library(ftnonpar)
> data(djdata)
> tmp <- pmreg(djdoppler,band=0.001)
> plot(djdoppler,col="grey")
> lines(tmp$y, col="red")
> tmp <- pmreg(djdoppler,band=0.03,verb=T)
```

The `ftnonpar` *package needs to be installed first using* `install.packages(ftnonpar")`

More properties

- If taut string touches upper bound in i and k with $i < k$

$$F_i^\lambda = Y_i + \lambda, \quad F_k^\lambda = Y_k + \lambda$$

  and does not touch lower bound in between

$$F_j^\lambda > Y_j - \lambda, \quad (j = i, \ldots, k)$$

  then F^λ is GCM on $[i, k]$ and $f_{i+1}^\lambda, \ldots, f_k^\lambda$ is isotonic regression for $y_{i+1}, \ldots, y_k$.

- Similarly, taut string yields antionic regression on intervals where taut string only touches lower bound.

More properties

- If taut string touches upper bound in i and lower bound in k and does not touch either bound in between, then f^λ takes local maximum on $[t_{i+1}, t_k]$.
- Vice versa for local minimum.

Calculation of taut string

Calculation is possible in O(n) steps:

- Solution is calculated from left to right
- Suppose taut string known up to k, then successively add new observations and
- calculate greatest convex minorant of upper bound and least concave majorant of lower bound
- Solution is extended once GCM initially smaller than LCM.

Taut strings and total variation

Total variation of real-valued function f on interval $[0,1]$:

$$\sup_P \sum_i |f(x_{i+1}) - f(x_i)|$$

the supremum running over all partitions $P = (x_1, \ldots, x_n)$ of the interval $[a, b]$. (Wikipedia)

Taut strings and total variation

Total variation of real-valued function f on interval $[0, 1]$:

$$\sup_P \sum_i |f(x_{i+1}) - f(x_i)|$$

the supremum running over all partitions $P = (x_1, \ldots, x_n)$ of the interval $[a, b]$. (Wikipedia)

Taut string method: Fast algorithm for minimising

$$T(f) = \sum_{i=1}^{n} (y_i - \hat{f}_i)^2 + \sum_{i=1}^{n-1} \lambda_i |\hat{f}_{i+1} - \hat{f}_i|$$

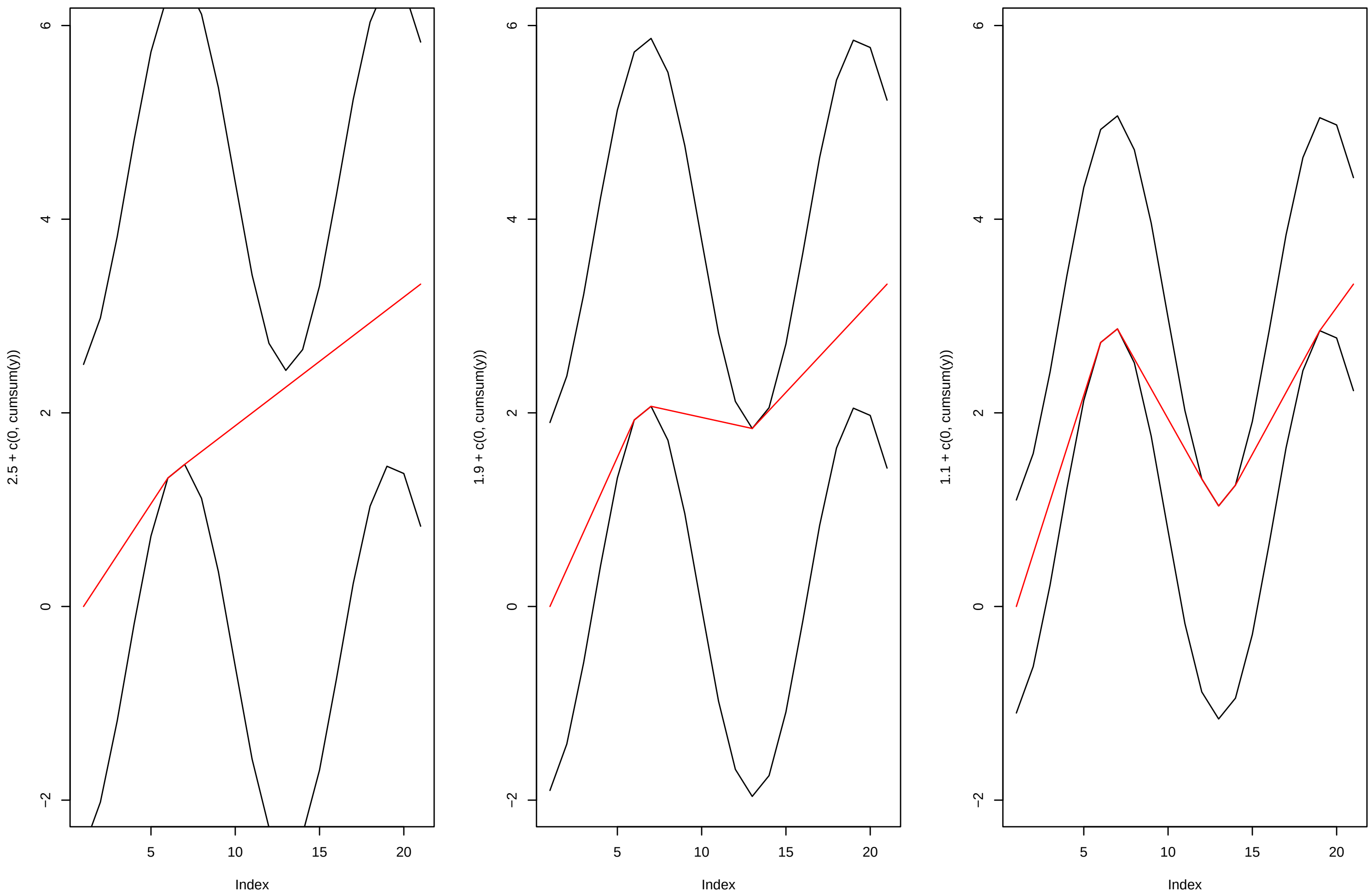
Taut strings and total variation

Since functional T is convex, a vector f minimises T if and only if

$$DT(f,\delta) := \lim_{\varepsilon \downarrow 0} \frac{T(f+\varepsilon\delta) - T(f)}{\varepsilon} \geq 0 \quad \text{for any } \delta \in \mathbb{R}^n.$$

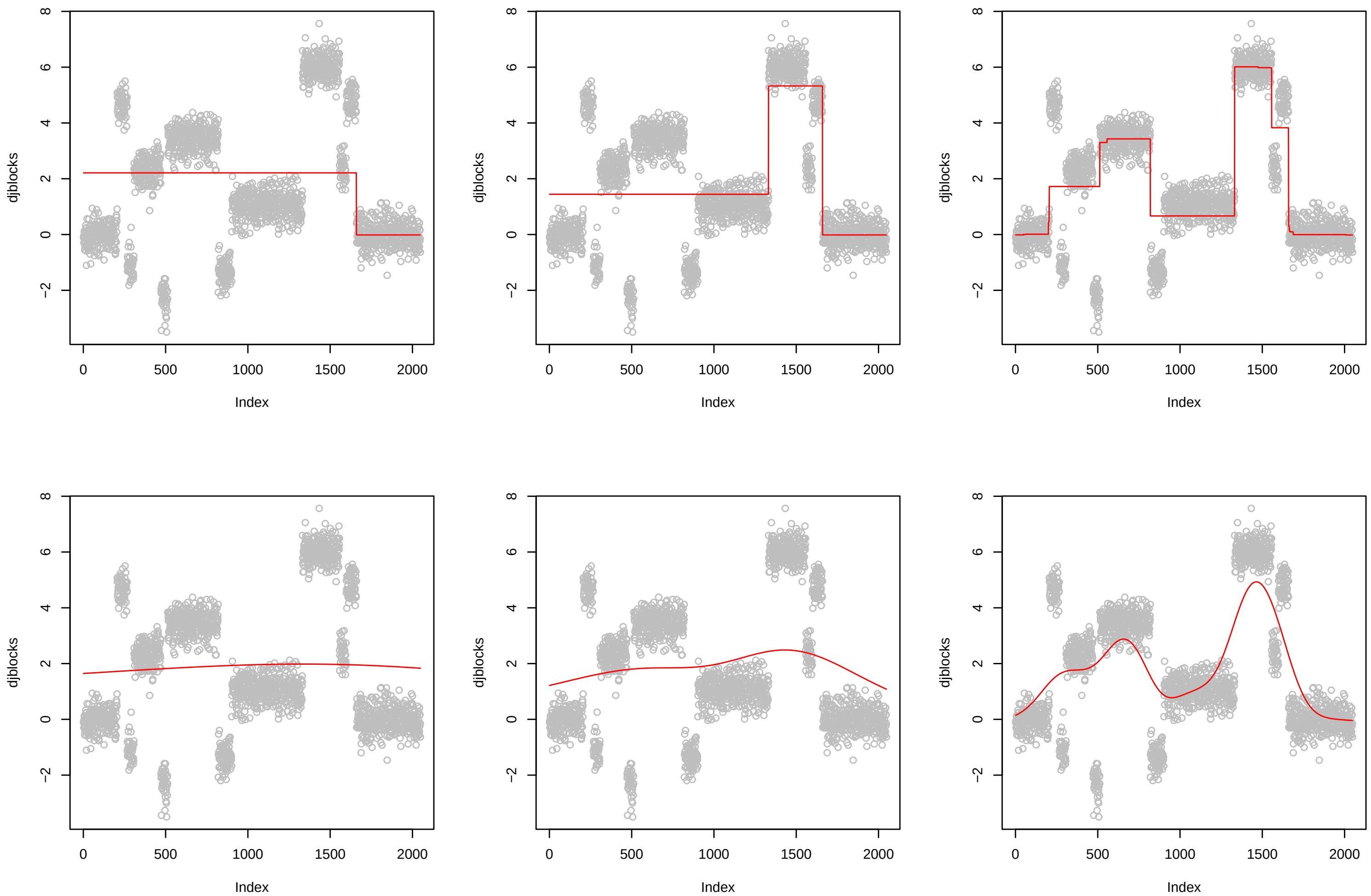
Taut strings and modality

Modality increases monotonically with decreasing tube width



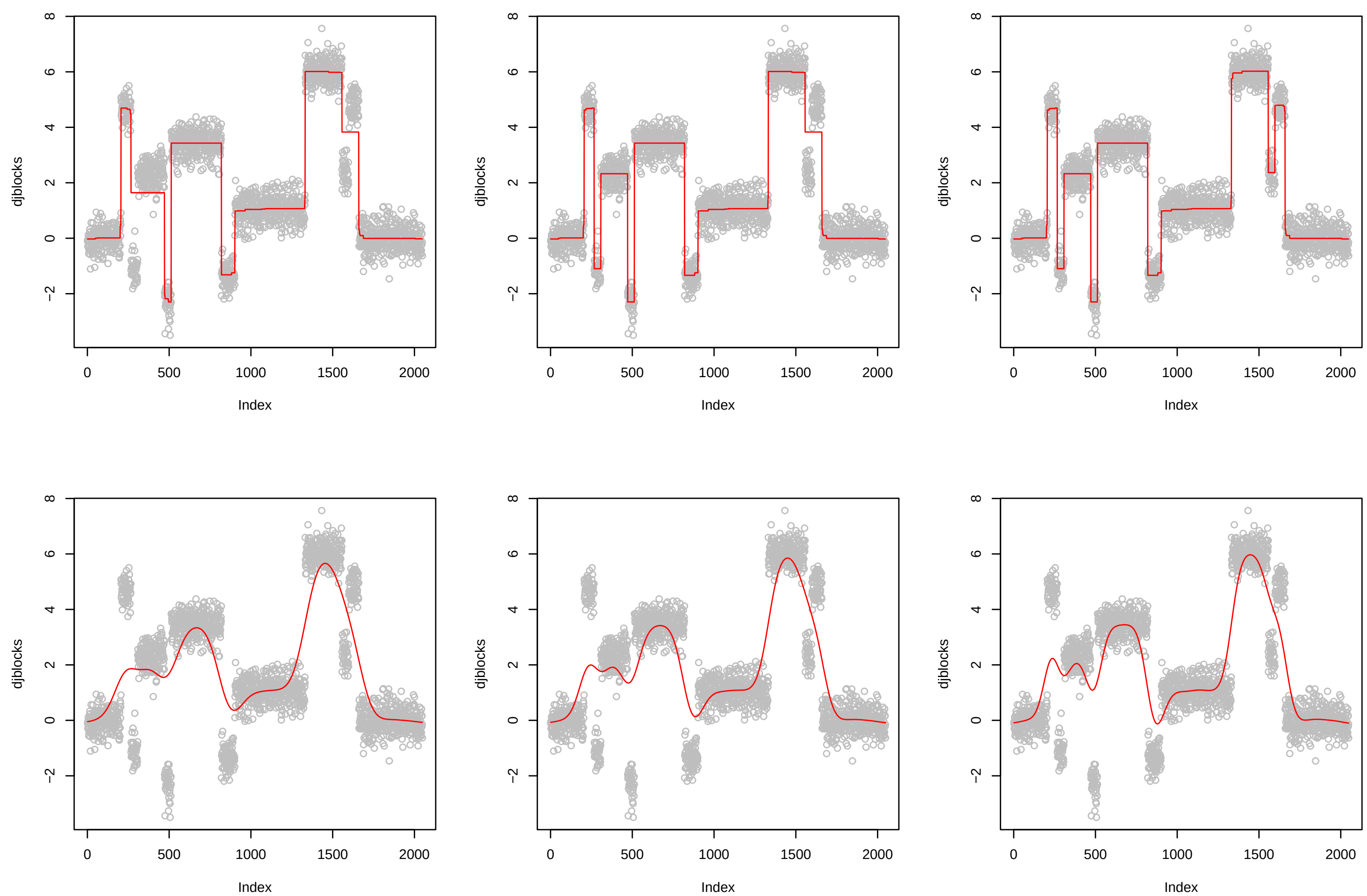
Taut strings and modality

Order in which peaks are added depends on power of peaks in contrast to kernel estimators:



Taut strings and modality

Order in which peaks are added depends on power of peaks in contrast to kernel estimators:



Taut strings and modality

Denote by $\mathcal{F}$ the set of all functions f such that the integral $F(x) = \int_0^x f(t)dt$ lies inside the tube with width λ and $\int_0^1 f(t) = Y(1) = \frac{1}{n}\sum_{i=1}^n y_i$. Then the derivative f^λ of the taut string minimises the modality among all $f \in \mathcal{F}$.

Proof: Denote the intervals where f^λ takes local extreme values (including extremes at the boundaries) by

$$I_i = [t_i^l, t_i^r], \quad i = 0, \ldots, k+1$$

where $0 = t_0^l < t_0^r < t_1^l < \cdots < t_{k+1}^l < t_{k+1}^r = 1.$

Taut strings and modality

We assume that f^λ is increasing on$[0, t_1^r]$. In this case f^λ takes a local maximum on I_i whenever i is odd and a local minimum whenever i is even. Furthermore for every function $f \in \mathcal{F}$

$$\max_{t \in I_i} f(t) \ge f^\lambda(I_i), \quad \text{if } i \text{ is odd,}$$

and

$$\min_{t \in I_i} f(t) \le f^\lambda(I_i), \quad \text{if } i \text{ is even.}$$

For example the first inequality is proved by noting that for every $f \in \mathcal{F}$ and odd i

$$\max_{t \in I_i} f(t) \ge \frac{F(t_i^r) - F(t_i^l)}{t_i^r - t_i^l} \ge \frac{Y(t_i^r) - \lambda - (Y(t_i^l + \varepsilon)}{t_i^r - t_i^l} = f^\lambda(I_i).$$

Taut strings and modality

Thus there are points $s_i \in I_i$ such that for every even i

$$f(s_i) \le f(I_i) < f(I_{i+1}) \le f(S_{i+1})$$

and for every odd i

$$f(s_i) \ge f(I_i) > f(I_{i+1}) \ge f(S_{i+1}).$$

Therefore every function $f \in \mathcal{F}$ has at least k local extreme values.

Taut strings and modality

Assume that $y_i = f(t_i) + \varepsilon_i$ with ε_i i.i.d. $\mathcal{N}(0, \sigma^2)$.

For bandwidths of order $C/\sqrt{n}$ the modality of f^λ is asymptotically consistent:

$$\lim_{C\to\infty} \liminf_{n\to\infty} \mathbb{P}(\text{Modality}(\tilde{f}_n^{C/\sqrt{n}}) = \text{Modality}(f)) = 1.$$

Taut strings and modality

Proof for showing that Modality($\tilde{f}_n^{C/\sqrt{n}}$*)*$\le$ *Modality(*f*)):*

Let $E_k = \sum_{i=1}^{k} \varepsilon_i$, then $\sqrt{n}E_t$ converges weakly to σW where W denotes the standard Wiener process. In particular

$$\lim_{n\to\infty} \mathbb{P}(\max|\sqrt{n}E_t| \le x) = \mathbb{P}(\max|W(t)| \le \frac{x}{\sigma}).$$

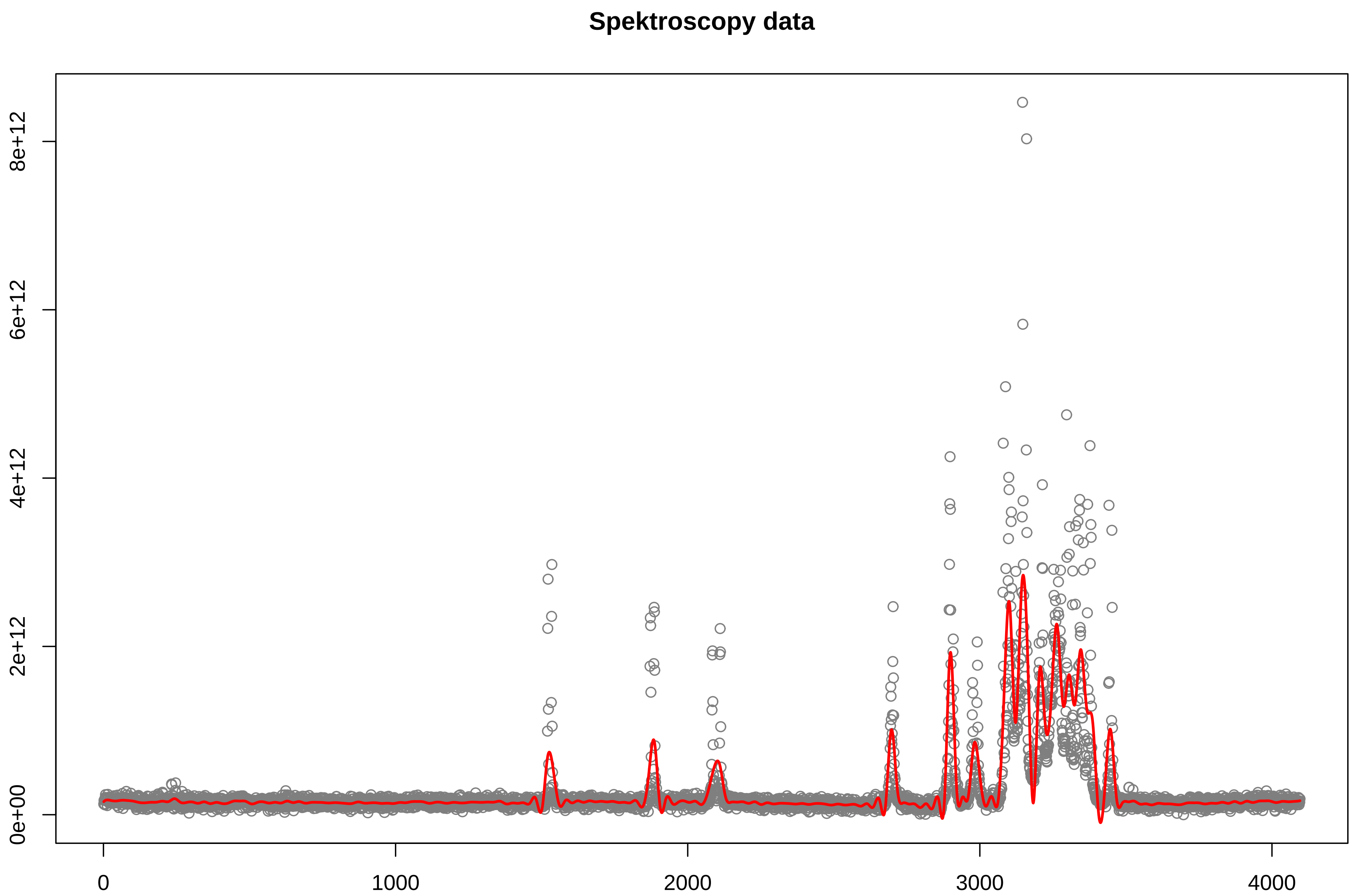
Therefore

$$\lim_{n\to\infty} \mathbb{P}(\max|Y_t - f(t)| \le \frac{C}{\sqrt{n}}) = \mathbb{P}(\max|W(t)| \le \frac{C}{\sigma}).$$

As n tends to infinity the probability that the function f lies in the tube with radius $C/\sqrt{n}$ tends to $\mathbb{P}(\max|W(t)| \le \frac{C}{\sigma})$. As the taut string minimises the modality we see that

$$\lim_{n\to\infty} \mathbb{P}(\max|Y_t - f^{C/\sqrt{n}}(t)| \le \frac{C}{\sqrt{n}}) = \mathbb{P}(\max|W(t)| \le \frac{C}{\sigma}).$$

Approximation



Adequate approximations

The approximation to the Spectroscopy data on the last slide was gathered from some classical method. It is

- not simple (Many artificial local extrema, some of them bigger than peaks approximating true features) and
- not adequate (Local maxima severely underestimated, some true features only weakly approximated).

Approximation and simplicity

Approximation

Adequate function: Function such that residuals 'look like' noise.

Any adequate function represents a good model for the data in the sense that data will look like a typical sample (Davies, 1995).

Simplicity

Find simplest adequate function.

Multiresolution Criterion

Check residuals on different scales and locations:

$$\left| \sum_{i \in I} (y_i - f_i) \right| < w_I \cdot \sigma$$

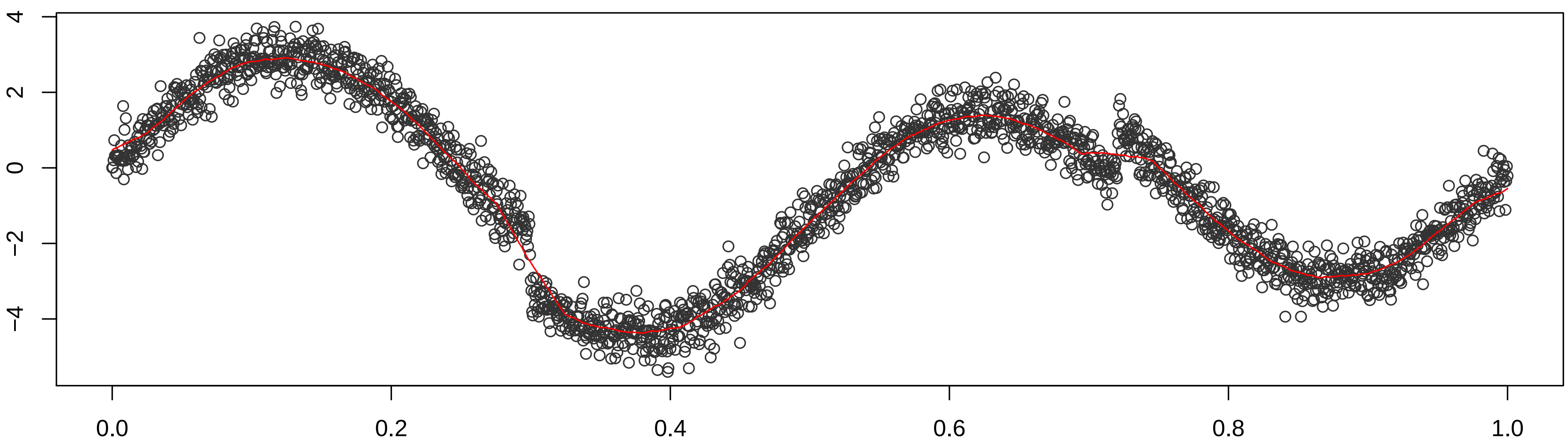
with $w_I = \sqrt{|I| \cdot 2\log(n)}$ for all intervals I of some family $\mathcal{I}$ of subintervals of $\{1, \ldots, n\}$. (Davies and Kovac, 2001)

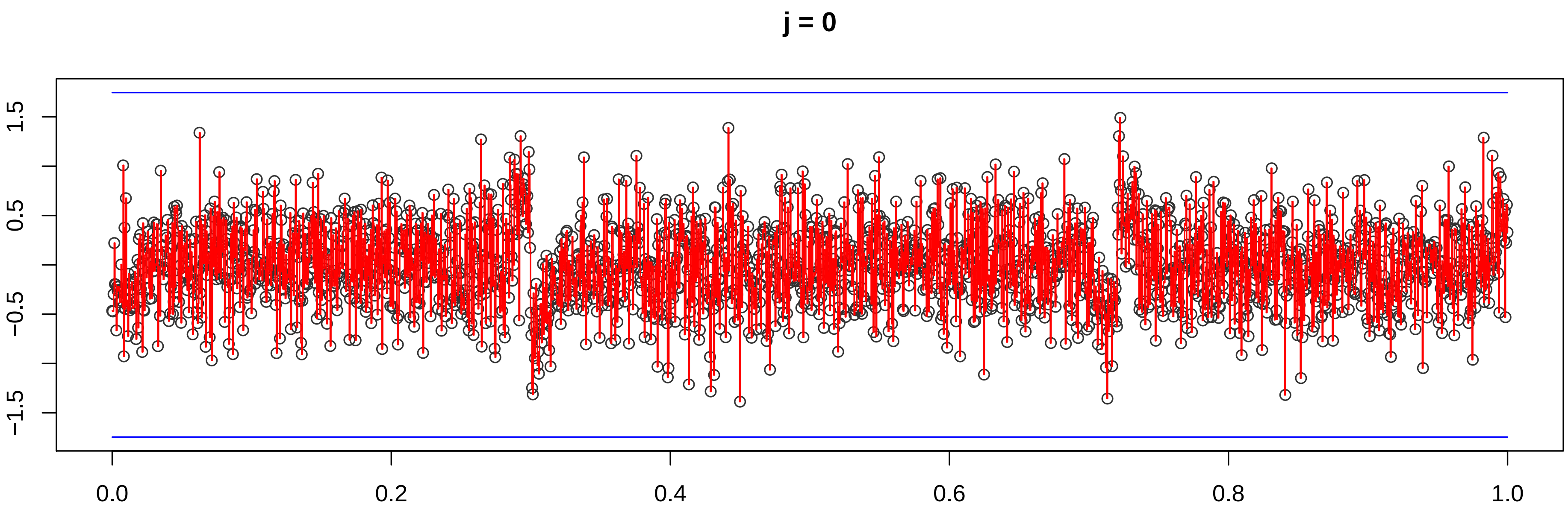
Theorem about maximum of white noise:

Let (X_n) be i.i.d. $\mathcal{N}(0,1)$. Then

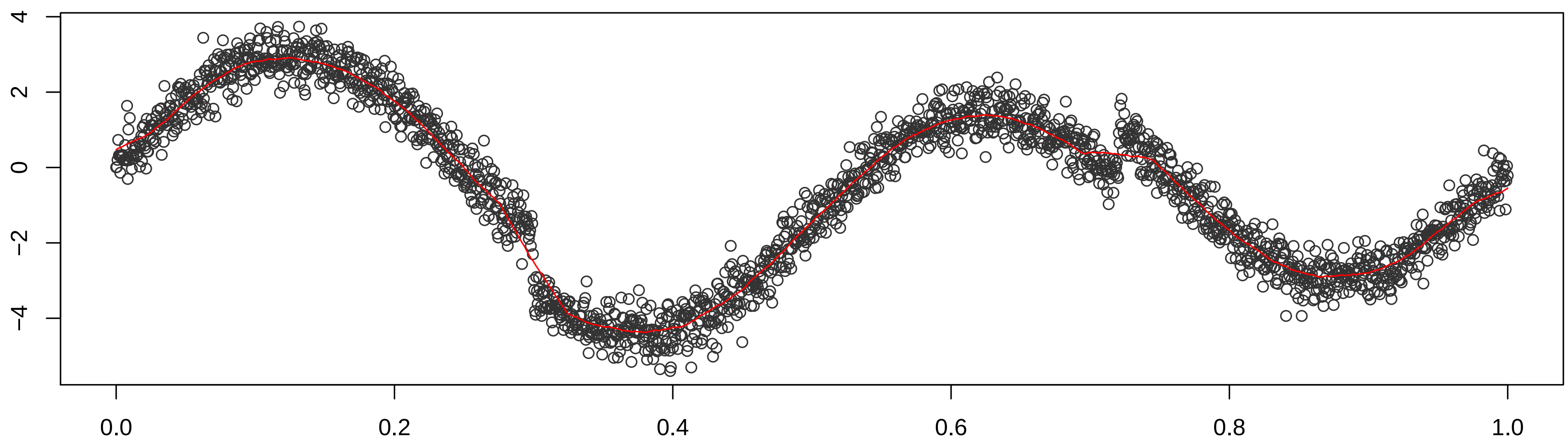
$$\mathbb{P}\left(\left\{\max_i |X_i| \leq \sqrt{2\log(n)}\right\}\right) \to 1, \quad n \to \infty.$$

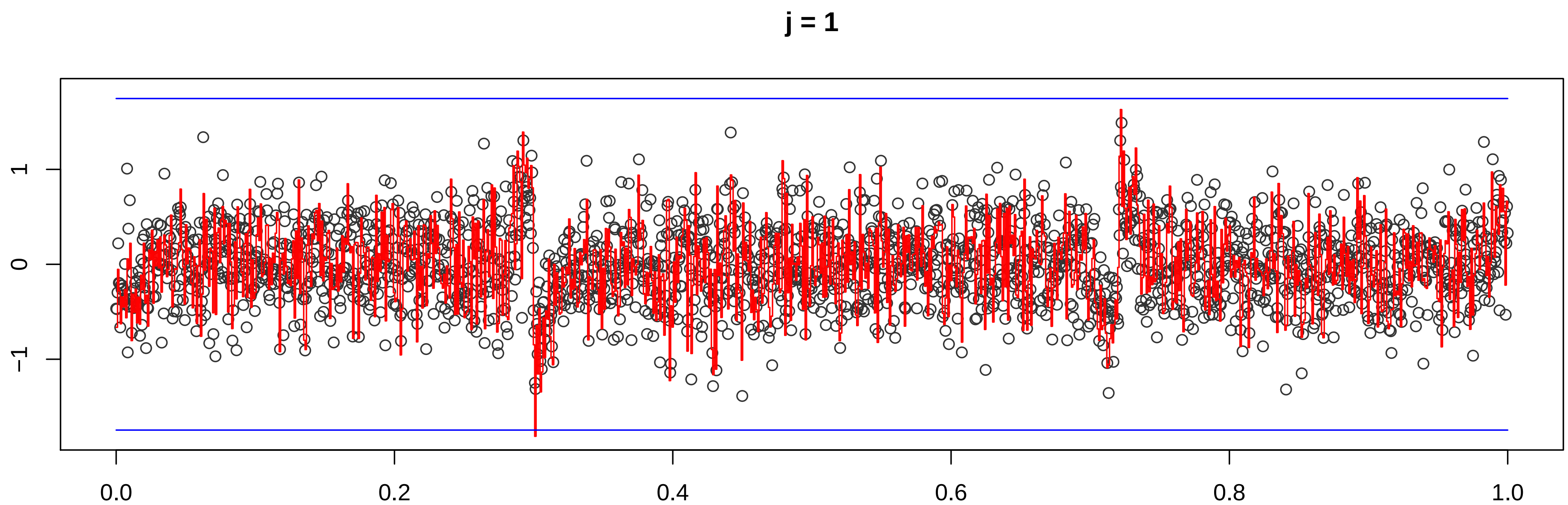
The Multiresolution Criterion



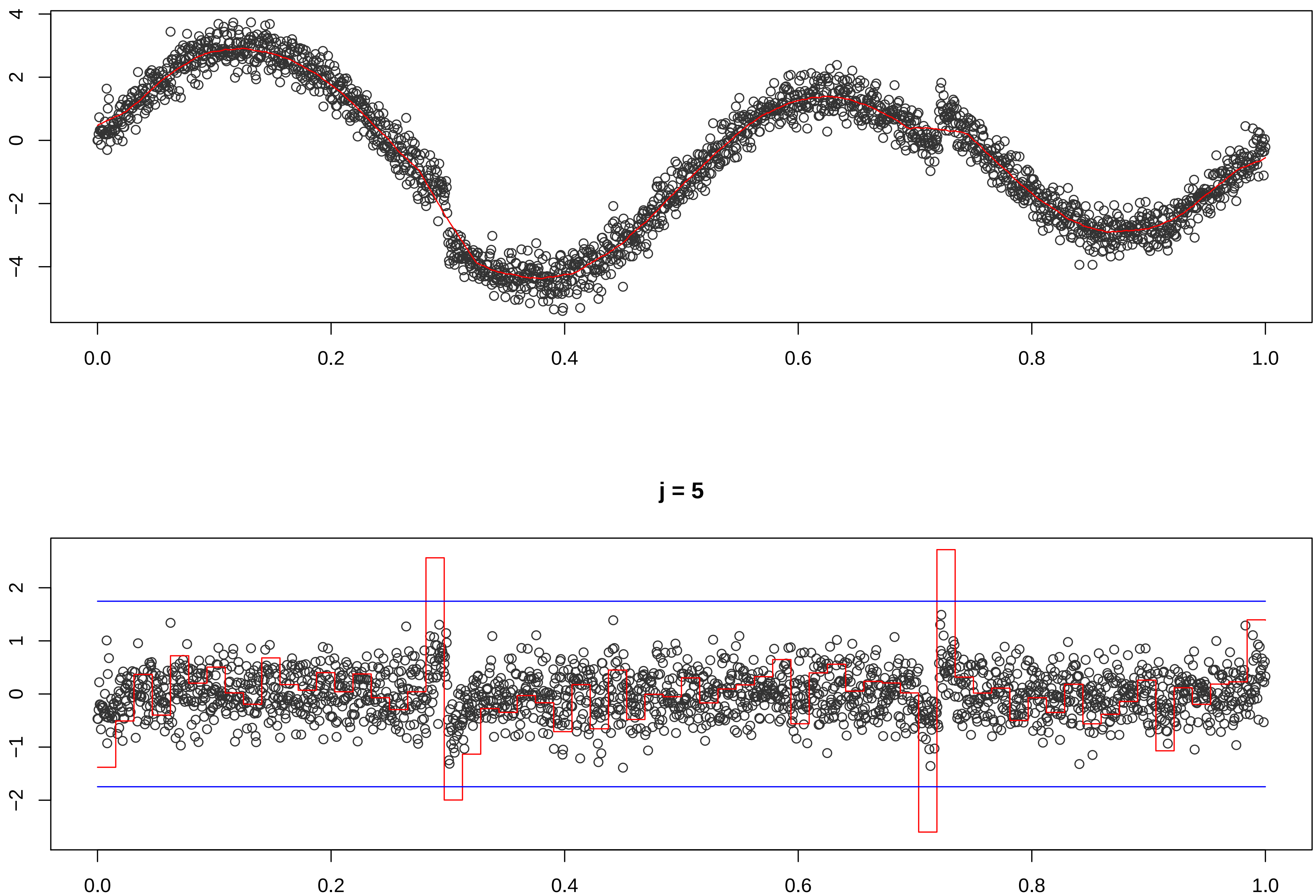


The Multiresolution Criterion





The Multiresolution Criterion



Approximation and simplicity

Approximation

Adequate function: Function such that residuals look like noise.

- Multiresolution criterion

Simplicity

Find simplest adequate function, eg minimize modality (number of local extreme values).

The taut string method

Minimization of number of local extreme values often difficult.

- Produce sequence of candidate functions $f_1, f_2, \ldots$ with increasing number of local extreme values.
- Stop once an adequate approximation f_k is produced.

One method to produce candidate functions:

- Taut string method

Global squeezing

Reducing width of tube

$\longrightarrow$ Increasing number of local extreme values.

Global squeezing

Reducing width of tube

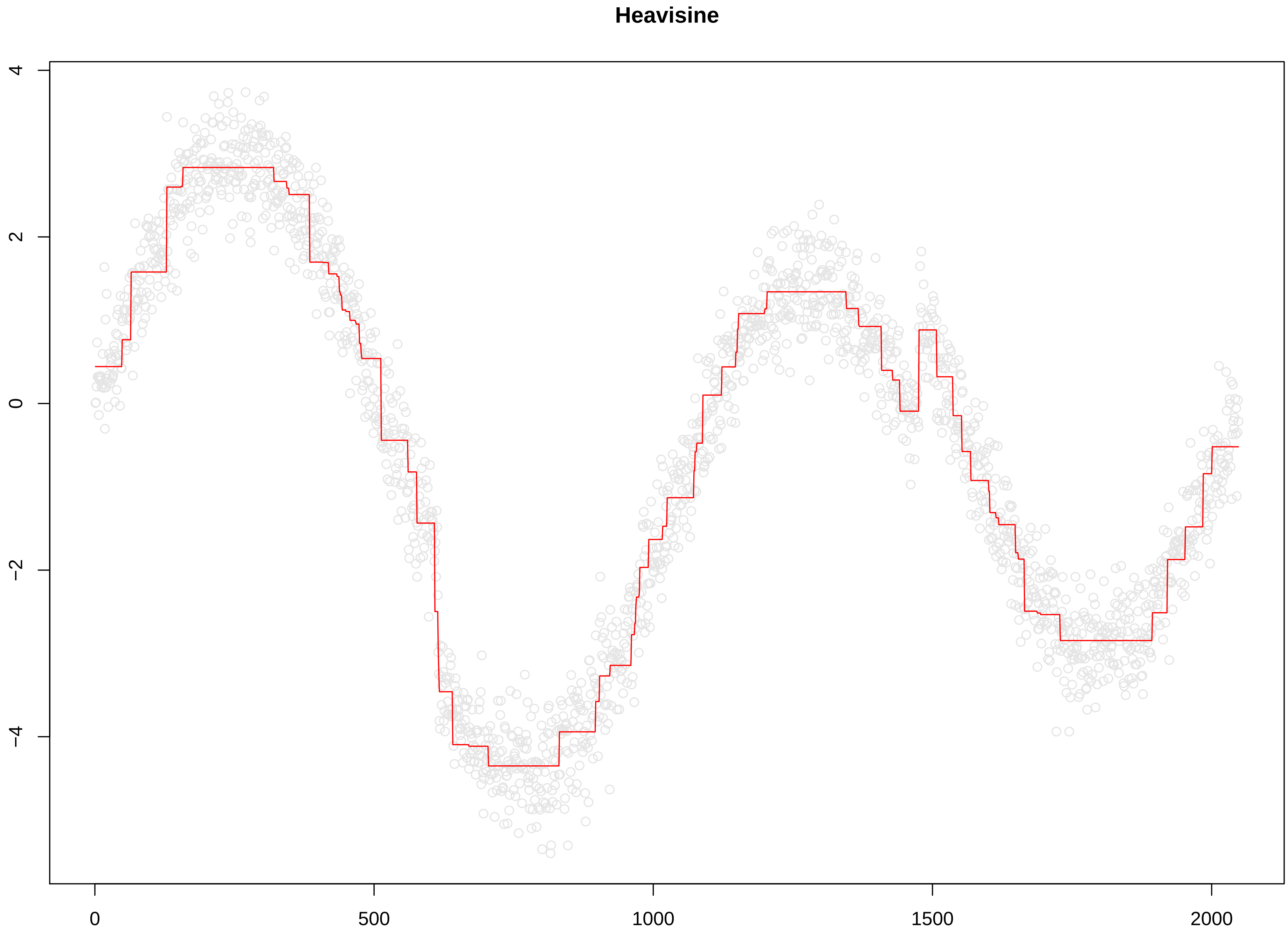
$\longrightarrow$ Increasing number of local extreme values.

Global squeezing:

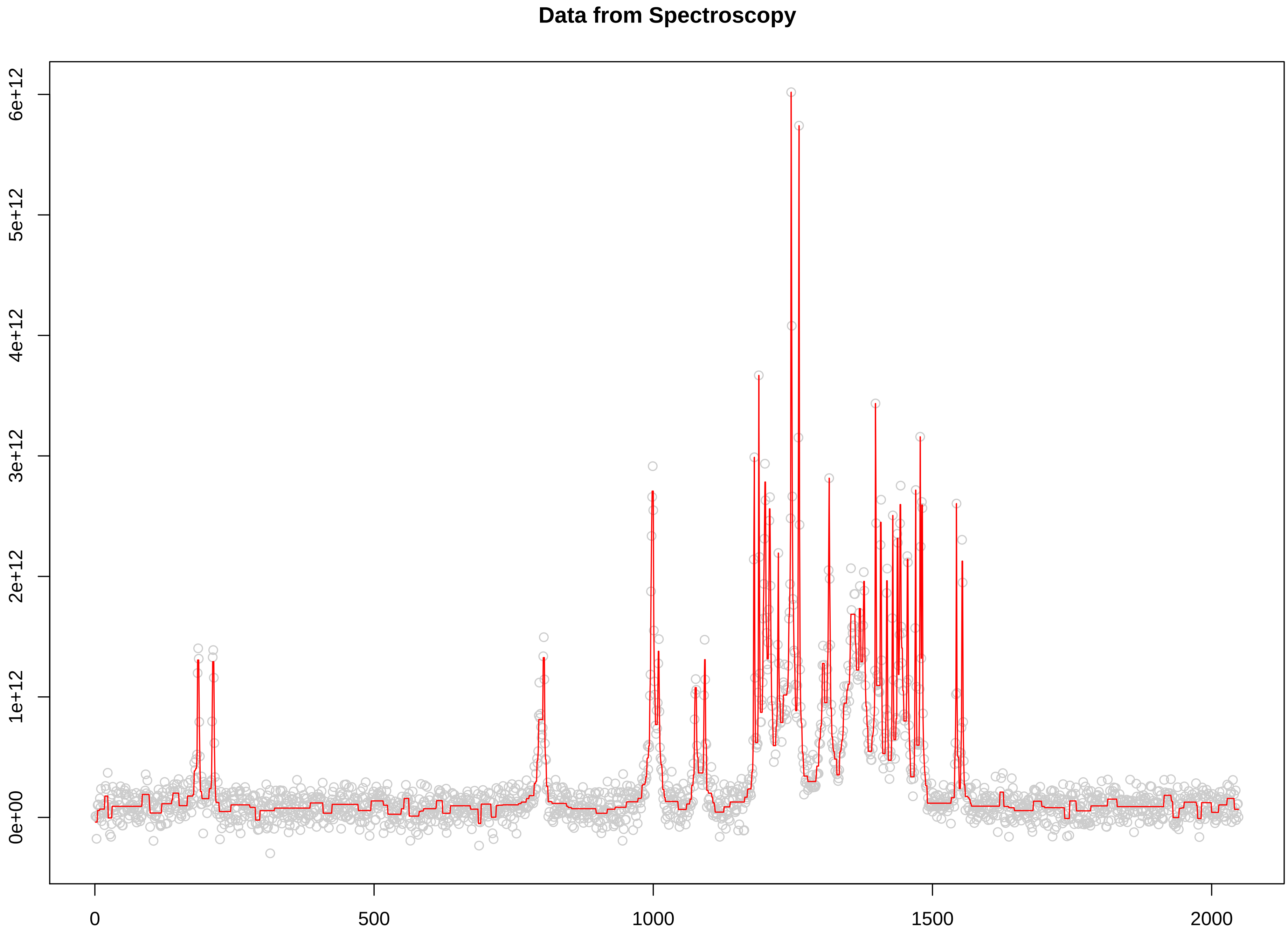
- Start with large tube width.
- Calculate taut string.
- Gradully reduce tube width.
- Stop if all multiresolution coefficients small enough.

Heavisine Data with Global Squeezing

The Heavisine Data



Data from Spectroscopy



Local Squeezing

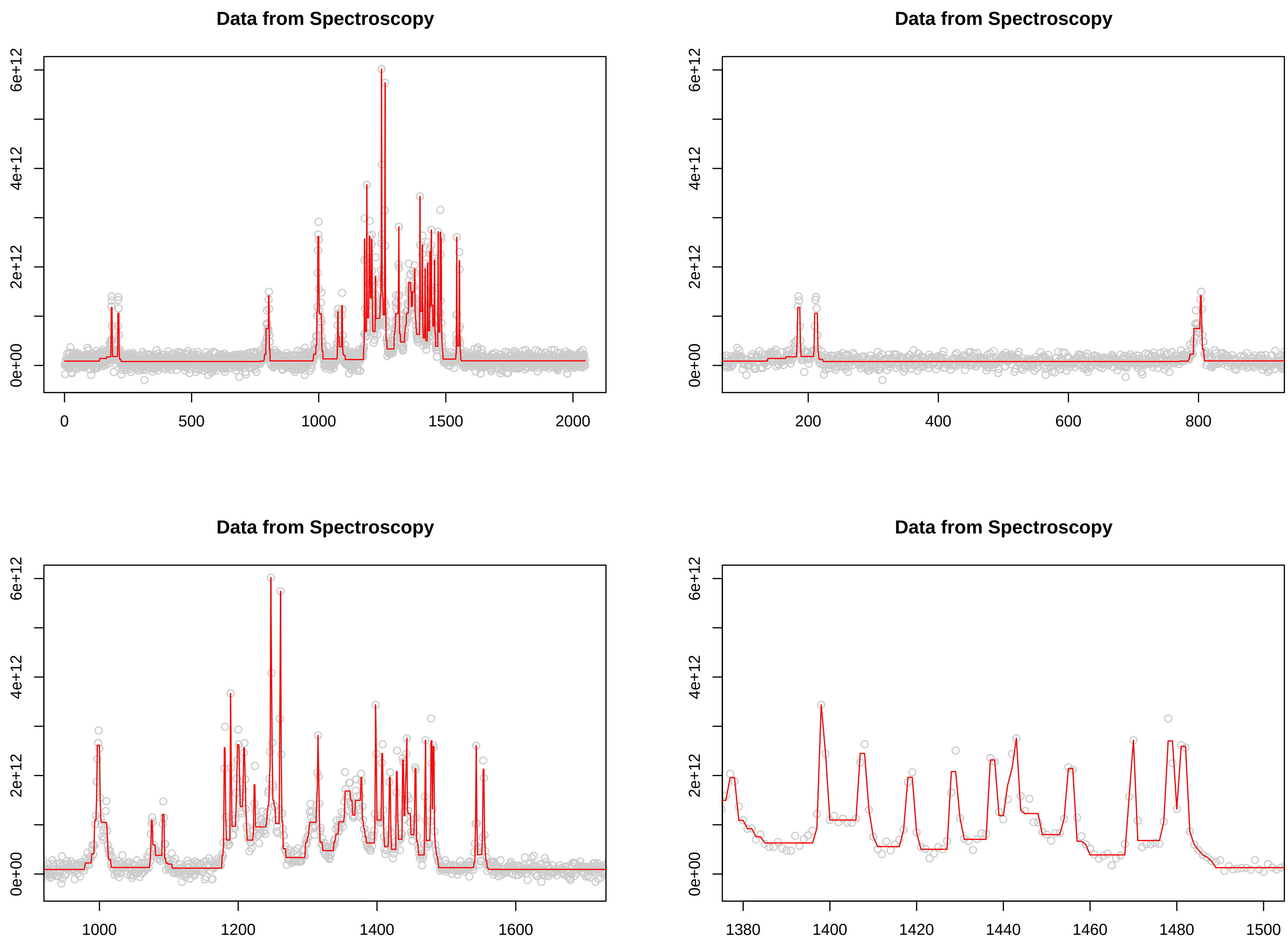
Reduce width of tube locally.

Local squeezing:

- Start with large, global bandwidth.
- Calculate taut string.
- Narrow tube on intervals where constraint is not satisfied.
- Stop when all constraints satisfied.

Spectroscopy Data with Local Squeezing

Data from Spectroscopy



The smoothness problem

Reformulation of the problem: Given noisy data $y_1, \ldots, y_n$ at points $x_1, \ldots, x_n$.

Find function f

- which fits the data (Multiresolution Criterion),
- simple (minimum number of local extreme values),
- is smooth.

Quickly!

Smooth taut string functional

- Taut string functional: Penalize differences in y-direction

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

Smooth taut string functional

- Taut string functional: Penalize differences in y-direction

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

- Now: Penalize Euclidean distances between points and minimize

$$T(f) = \sum_{i=1}^{n} (y_i - f_i)^2 + \sum_{i=1}^{n-1} \lambda_i \sqrt{\varepsilon (x_{i+1} - x_i)^2 + (f_{i+1} - f_i)^2}.$$

Algorithm

$T(f)$ is differentiable, so minimization possible with standard techniques like steepest descent method:

(1) Start with $f^0 = y$ and $k = 1$.

(2) T is decreasing in direction of $-\nabla f$.

(3) Determine $\lambda > 0$ and $f^k = f^{k-1} - \lambda \nabla f$ such that

$$T(f^k) < T(f^{k-1})$$

(4) If $\max |\nabla f| > 10^{-10}$ increase $k = k + 1$ and go to step (2).

Problem: Speed of convergence decays rapidly.

New attempt

Find minimiser $\tilde{f}$ of functional

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

- Suppose: $\tilde{f}_1$ known.
- T convex and differentiable, then

$$0 = \frac{\partial T(\tilde{f})}{\partial f_1} = 2(\tilde{f}_1 - y_1) - \lambda_1 g'(\tilde{f}_2 - \tilde{f}_1)$$

- Solve for $\tilde{f}_2 \longrightarrow \tilde{f}_1$ and $\tilde{f}_2$ known.

New attempt

Find minimiser $\tilde{f}$ of functional

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

- Suppose: $\tilde{f}_1$ and $\tilde{f}_2$ known.
- T convex and differentiable, then

$$0 = \frac{\partial T(\tilde{f})}{\partial f_2} = 2(\tilde{f}_1 - y_1) + \lambda_1 g'(\tilde{f}_2 - \tilde{f}_1) - \lambda_2 g'(\tilde{f}_3 - \tilde{f}_2)$$

- Solve for $\tilde{f}_3 \rightarrow \tilde{f}_1, \tilde{f}_2$ and $\tilde{f}_3$ known.

New attempt

Idea:

- Once $\tilde{f}_1$ is known, possible to calculate $\tilde{f}$ easily.
- Solution $\tilde{f}$ satisfies

$$\sum_{i=1}^{n} (y_i - \tilde{f}_i) = 0.$$

- Moreover: If $f_1 > \tilde{f}_1$, then $f_j > \tilde{f}_j$ for all j. Similarly if $f_1 < \tilde{f}_1$,
- Using nested intervals for f_1 yields $\tilde{f}_1$ and thus $\tilde{f}$.

New attempt

Idea:

- Once $\tilde{f}_1$ is known, possible to calculate $\tilde{f}$ easily.
- Solution $\tilde{f}$ satisfies

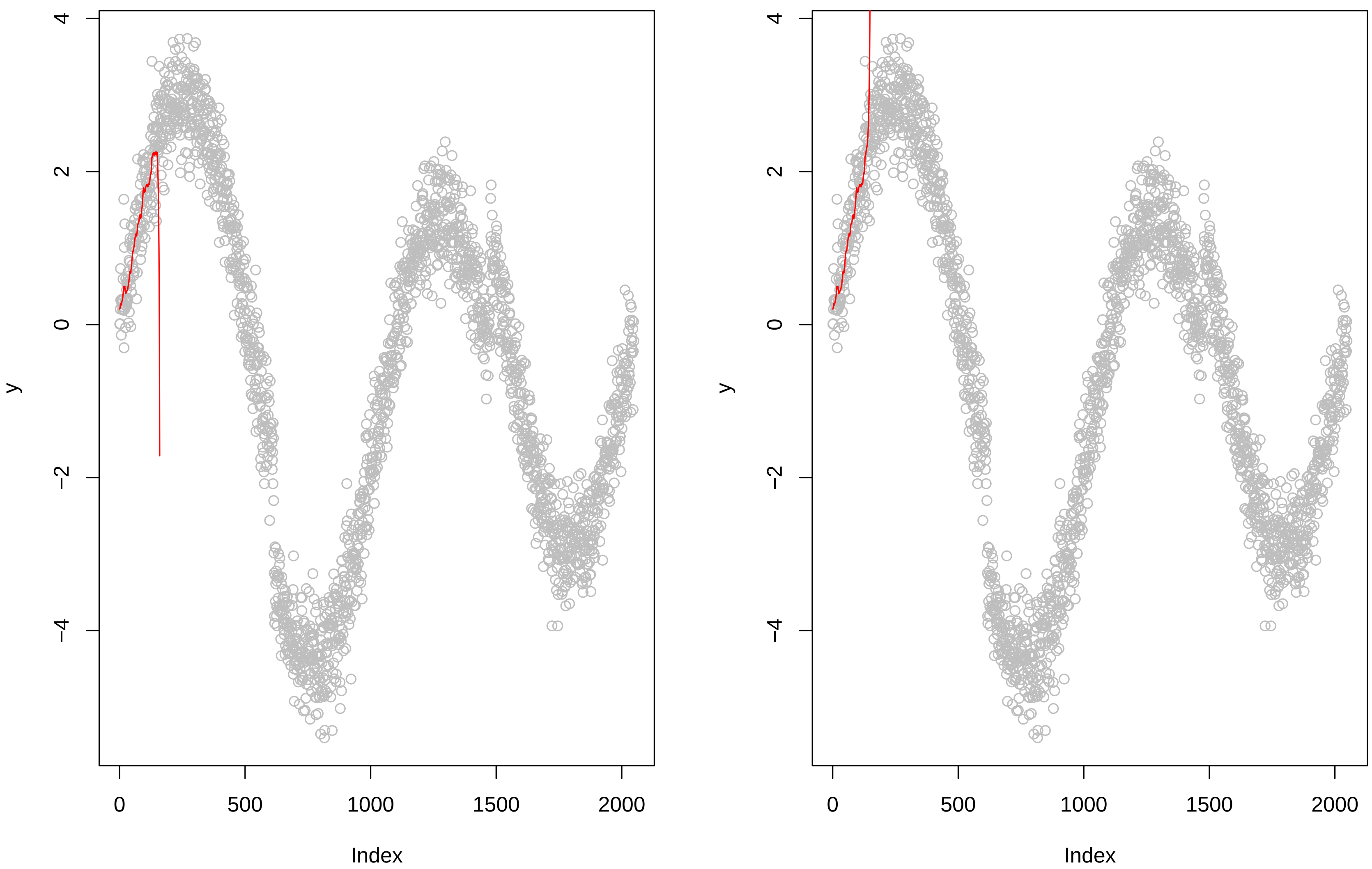
$$\sum_{i=1}^{n}(y_i - \tilde{f}_i) = 0.$$

- Moreover: If $f_1 > \tilde{f}_1$, then $f_j > \tilde{f}_j$ for all j. Similarly if $f_1 < \tilde{f}_1$,
- Using nested intervals for f_1 yields $\tilde{f}_1$ and thus $\tilde{f}$.

Theoretically. . .

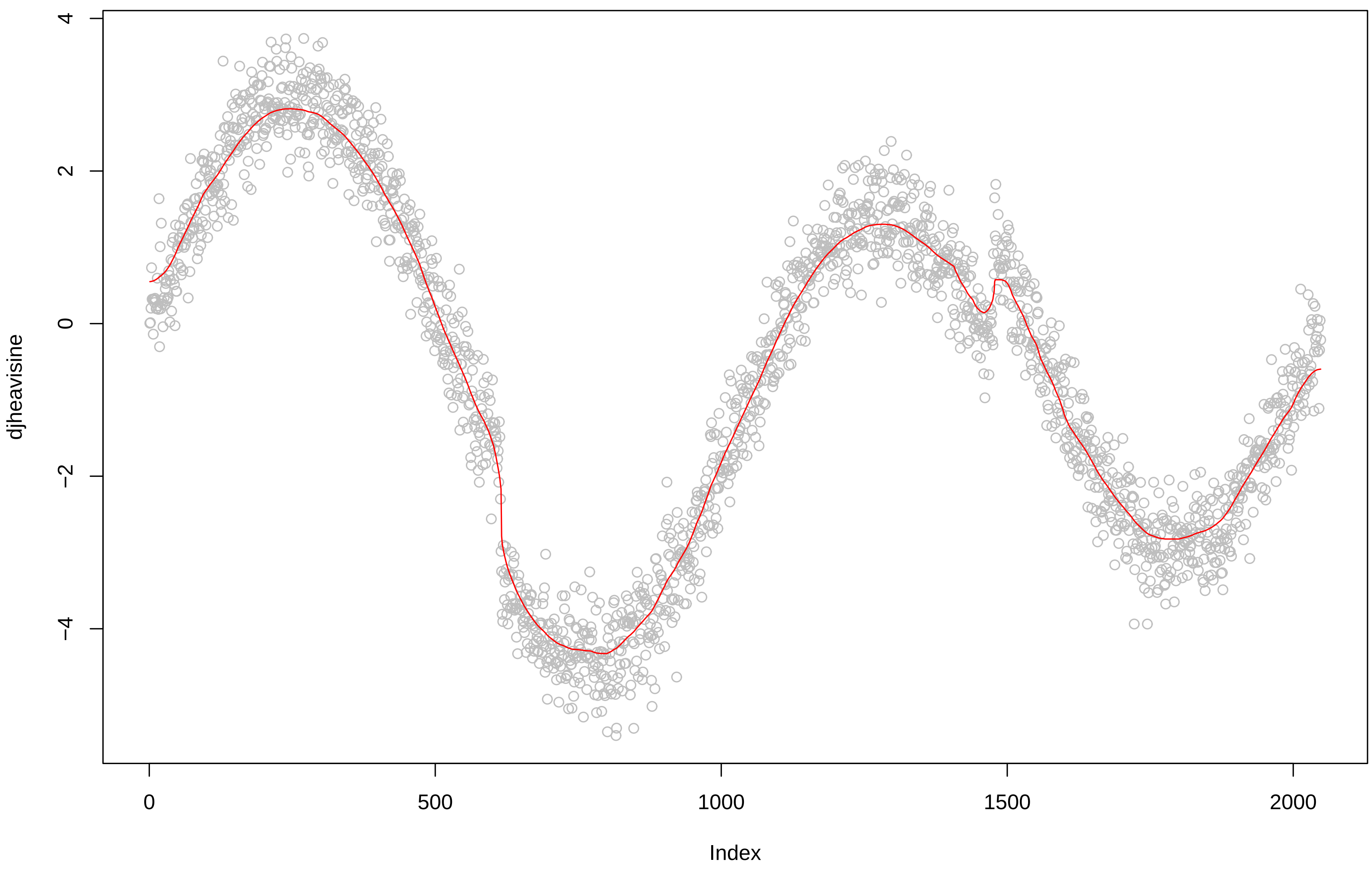
Sensitivity regarding starting point

$f_1 = 0.2017878903881263$ und $f_1 = 0.2017878903881264$



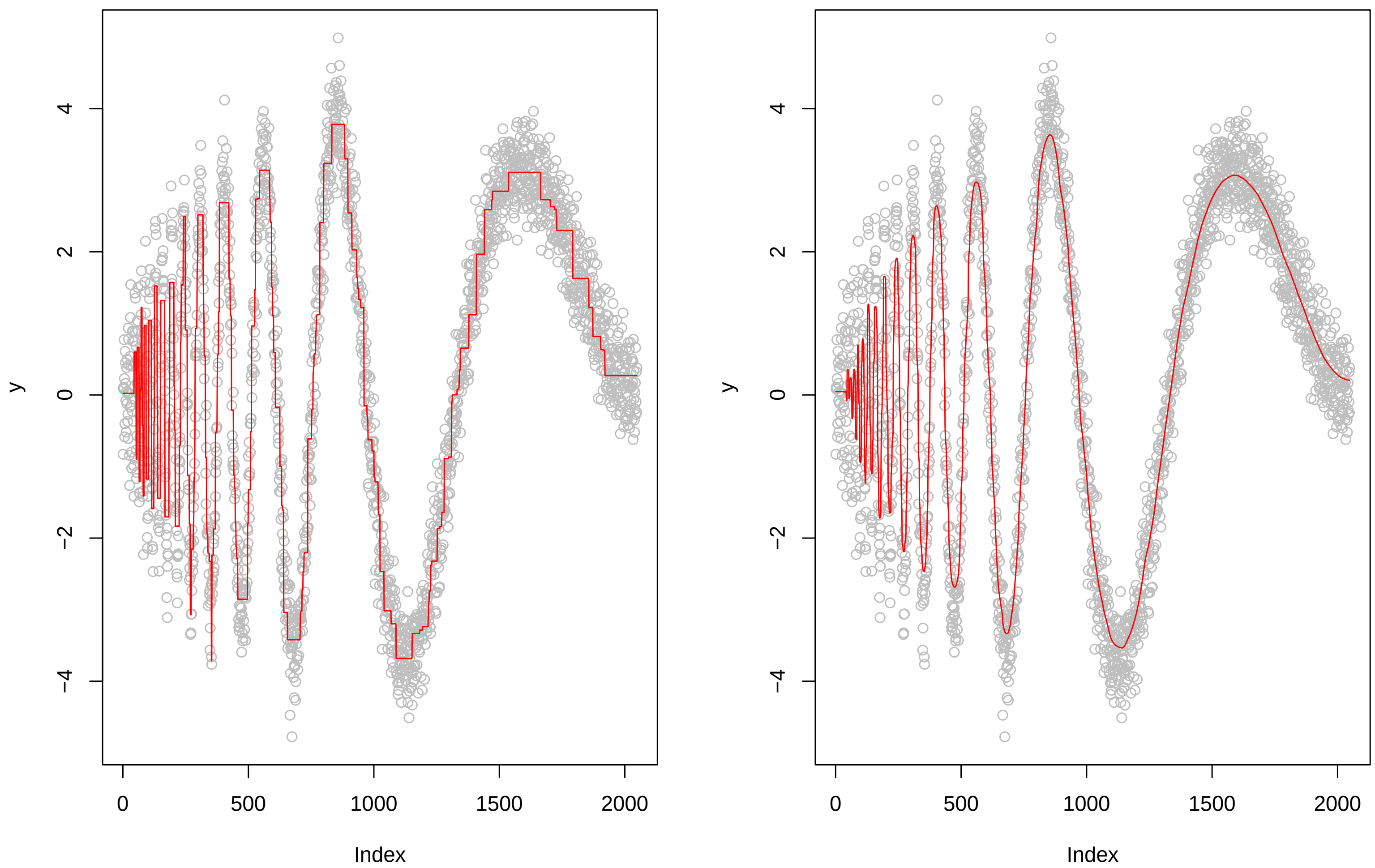
Correction

Use nested intervals for each data point

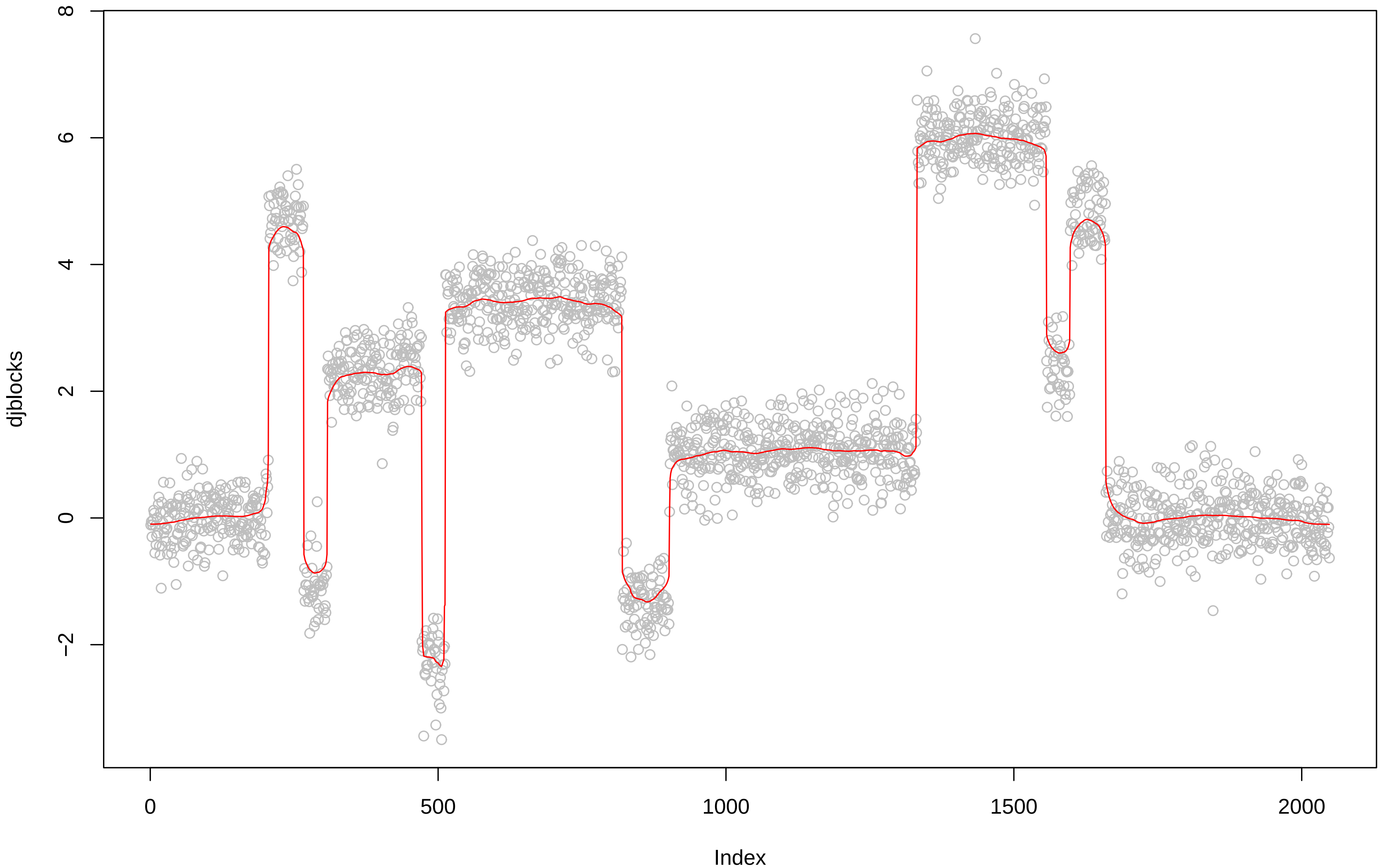


Determination of local penalties

Use local squeezing to determine local penalties automatically.



Monotonicity constraints



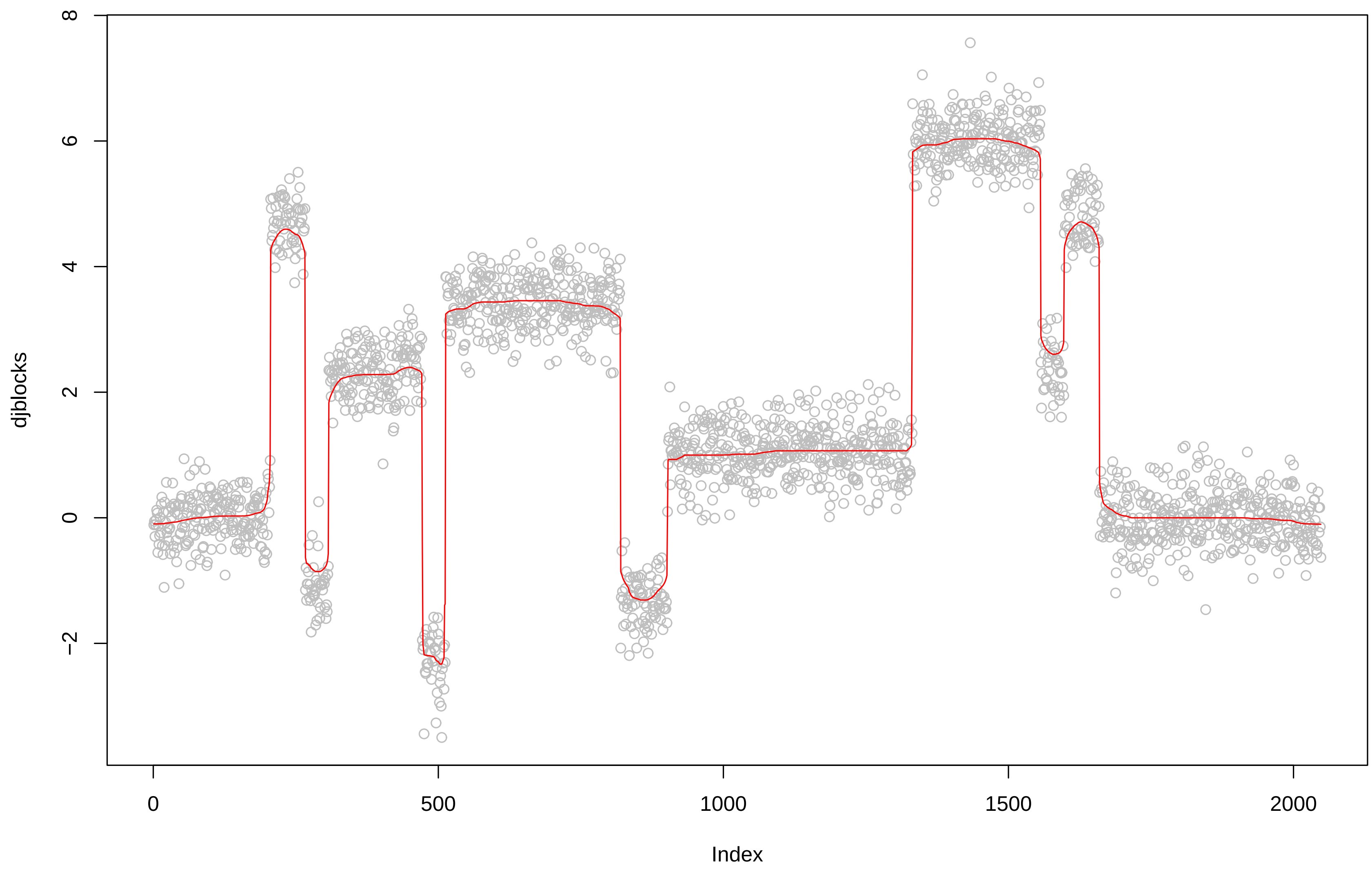
Monotonicity constraints

Derive monotonicity behaviour from usual taut string.

For all $\tau_1, \tau_2, \ldots, \tau_{n-1} \in \{-1, 1\}$ minimise $T(f)$ among all f such that $(f_{i+1} - f_i)\tau_i \geq 0$.

Modified procedure easily adaptable to this situation. If monotonicity behaviour is determined with taut string method or TV minimization, then existence is guaranteed.

Monotonicity constraints



Approximation and simplicity

The original problem revisited:

Approximation

Adequate function: Function such that residuals look like noise.

- Multiresolution criterion

Simplicity

Find simplest adequate function, eg minimize modality (number of local extreme values).

Minimizing total variation

Low modality $\sim$ small total variation

Minimize total variation among all adequate functions.

$$\min \sum_{i=1}^{n-1} |f_{i+1} - f_i| \text{ s.t. } \frac{1}{\sqrt{|I|}} \left| \sum_{i \in I} y_i - f_i \right| < \sqrt{2\log(n)} \cdot \sigma$$

for all $I \in \mathcal{I}$.

Linear programming

- Problem of linear programming

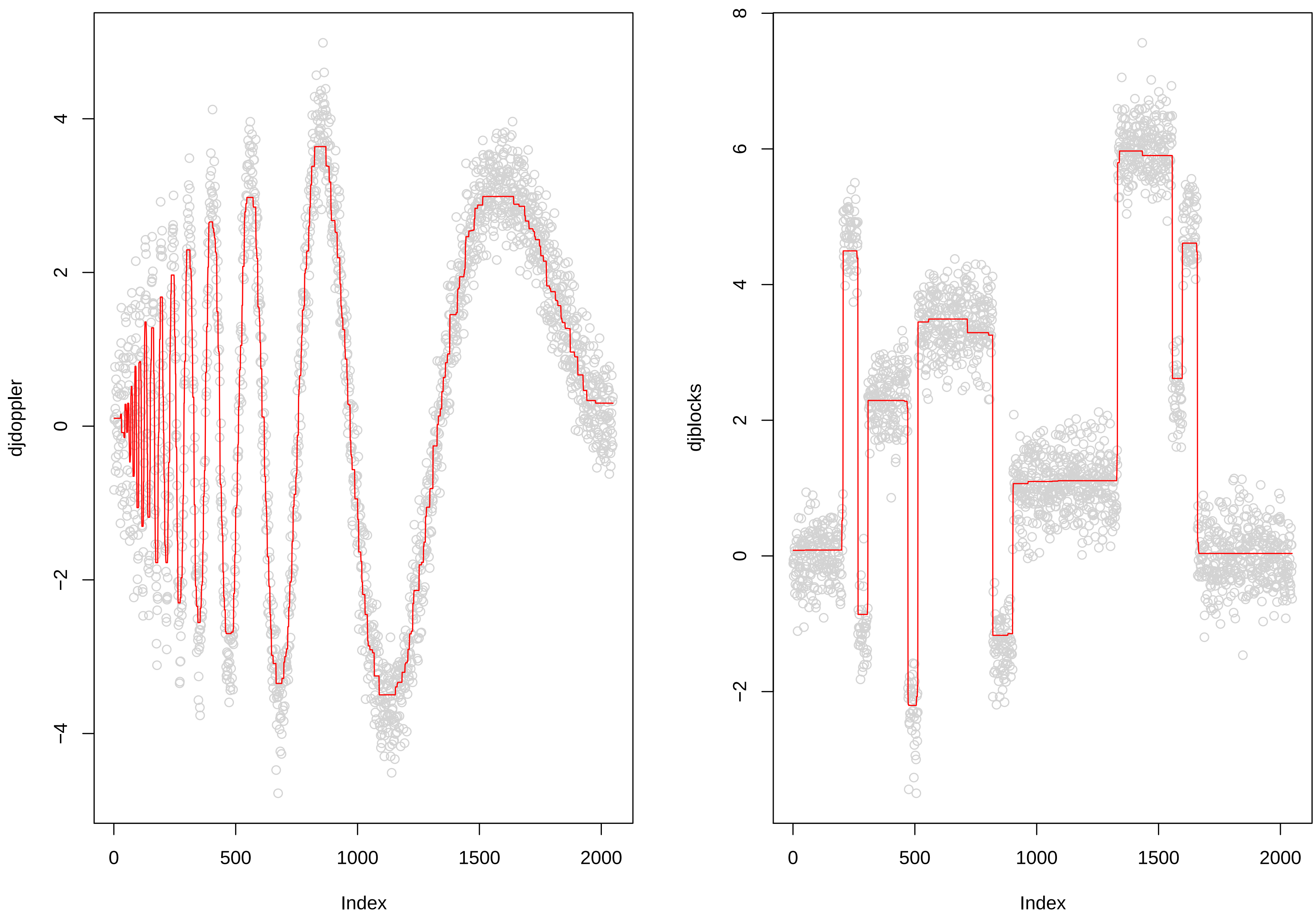
$$\min c^t x \quad \text{s.t. } Ax = b,\ x \geq 0$$

- Standard algorithms applicable like Simplex, Interior Point Methods etc.

- Huge dimensionality

- Using the structure of A yields simplex iterations of order $O(n)$

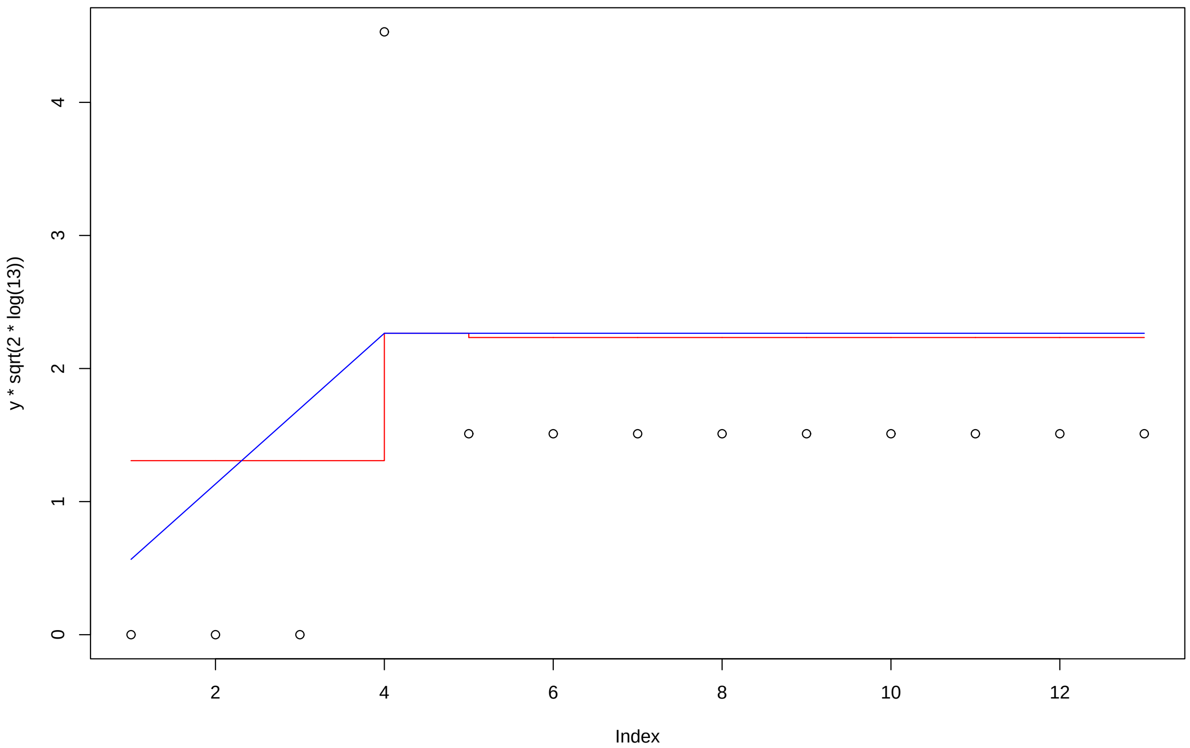
Minimizing total variation



Minimizing total variation

Minimization of TV = minimization of modality?

In general: No! — Usually: Yes!



But at least: can calculate lower bound for modality of adequate functions.

Minimizing total variation

No. constant pieces = no. of active MR constraints

In practice the assumption of the following lemma is often satisfied:

Let $\tilde{f} \in \mathcal{F}$ such that for each interval I on which $\tilde{f}$ takes a local extreme value there is a subinterval $J \subset I, J \in \mathcal{I}$ with

$$\frac{1}{\sqrt{|J|}} \sum_J (\tilde{f}_i - y_i) = \sqrt{2 \log(n)} \sigma$$

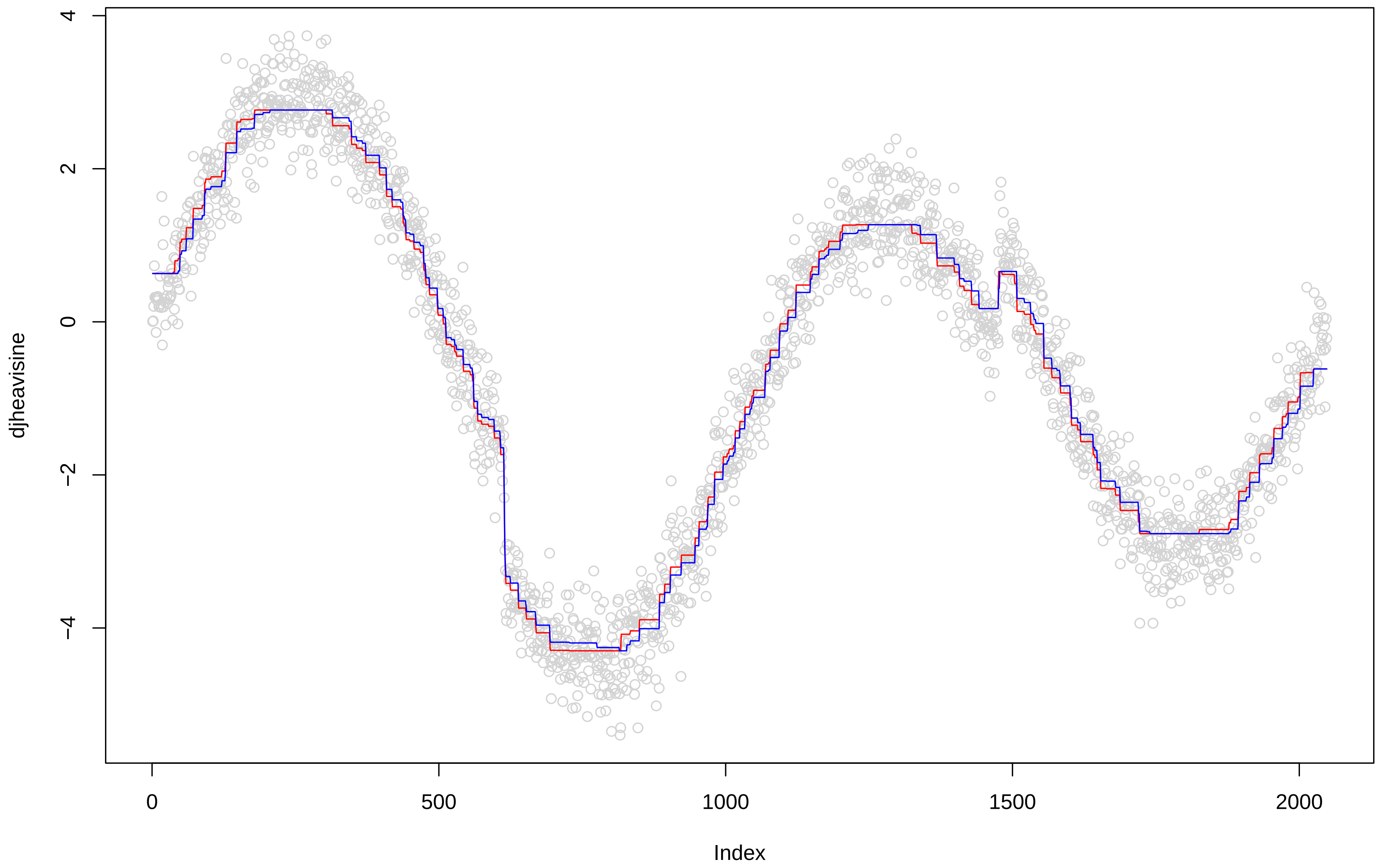
for a local minimum and

$$\frac{1}{\sqrt{|J|}} \sum_J (\tilde{f}_i - y_i) = -\sqrt{2 \log(n)} \sigma$$

for a local maximum.

Then $\tilde{f}$ attains the minimum modality among all $f \in \mathcal{F}$.

No uniqueness of solution



As is often the case for L_1-problems solution is not unique:

```
plot(djheavisine,col="grey")
lines(mintvmon(djheavisine,method=0)$y,col="red")
lines(mintvmon(djheavisine[2048:1],method=0)$y[2048:1],col="blue")
```

Smoother approximations

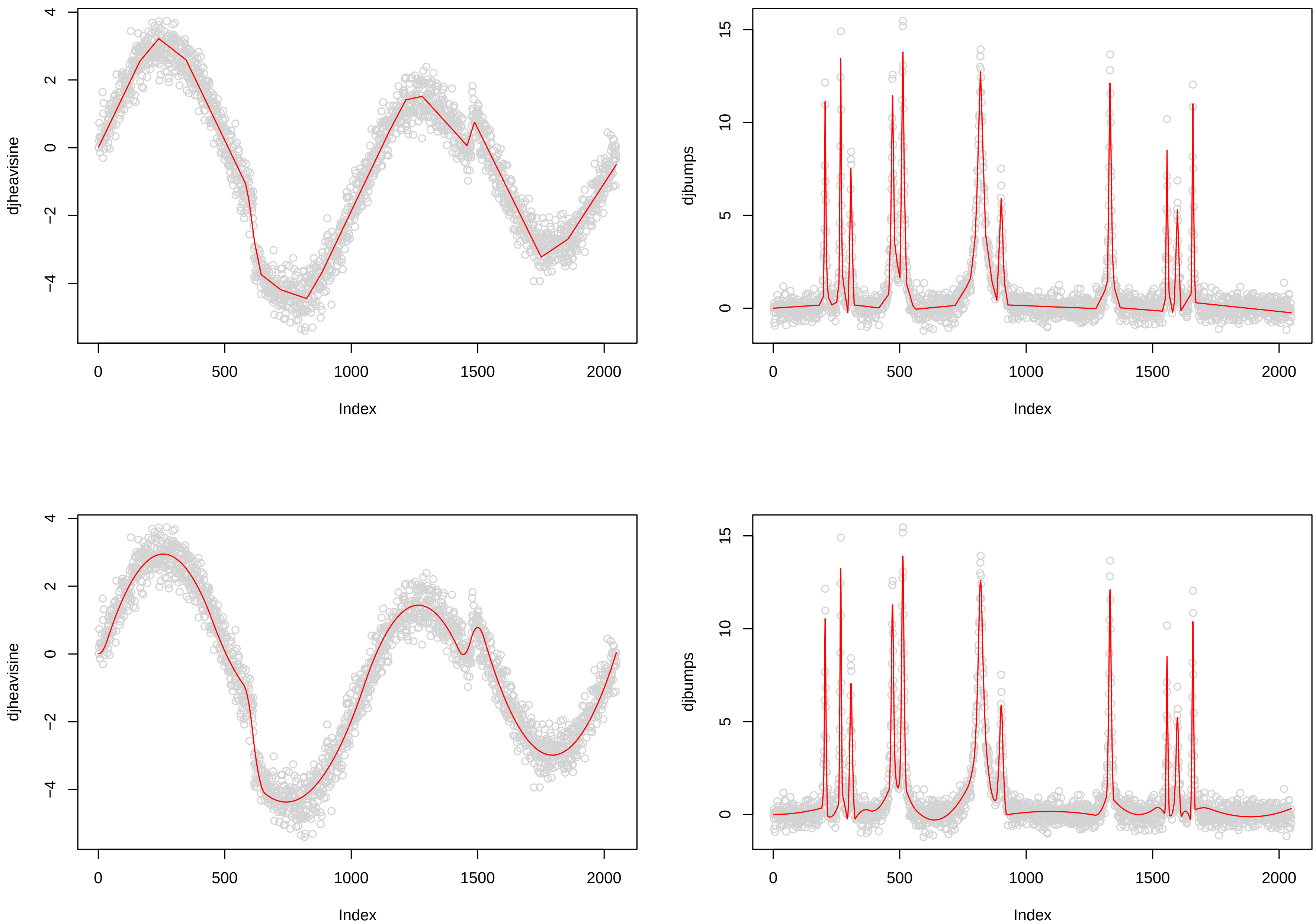
Minimizing $\mathrm{TV}(f')$ or $\mathrm{TV}(f'')$ yields smoother approximations.

$$\min \sum_{i=1}^{n-2} |f_{i+2} - 2f_{i+1} + f_i|$$

$$\text{or } \min \sum_{i=1}^{n-3} |f_{i+3} - 3f_{i+2} + 3f_{i+1} - f_i|$$

$$\text{s.t. } \frac{1}{\sqrt{|I|}} \left| \sum_{i \in I} y_i - f_i \right| < \sqrt{2\log(n)} \cdot \sigma \text{ for all } I \in \mathcal{I}.$$

Smoother approximations



Smoothness and Modality

Minimize $\mathrm{TV}(f)$ s.t. MR constraints:

$\rightarrow f_0$

Smoothness and Modality

Minimize $\mathrm{TV}(f)$ s.t. MR constraints:

$\rightarrow f_0$

Minimize $\mathrm{TV}(f')$ s.t. MR constraints and monotonicity constraints obtained from f_0:

$\rightarrow f_1$

Smoothness and Modality

Minimize $\mathrm{TV}(f)$ s.t. MR constraints:

$\rightarrow f_0$

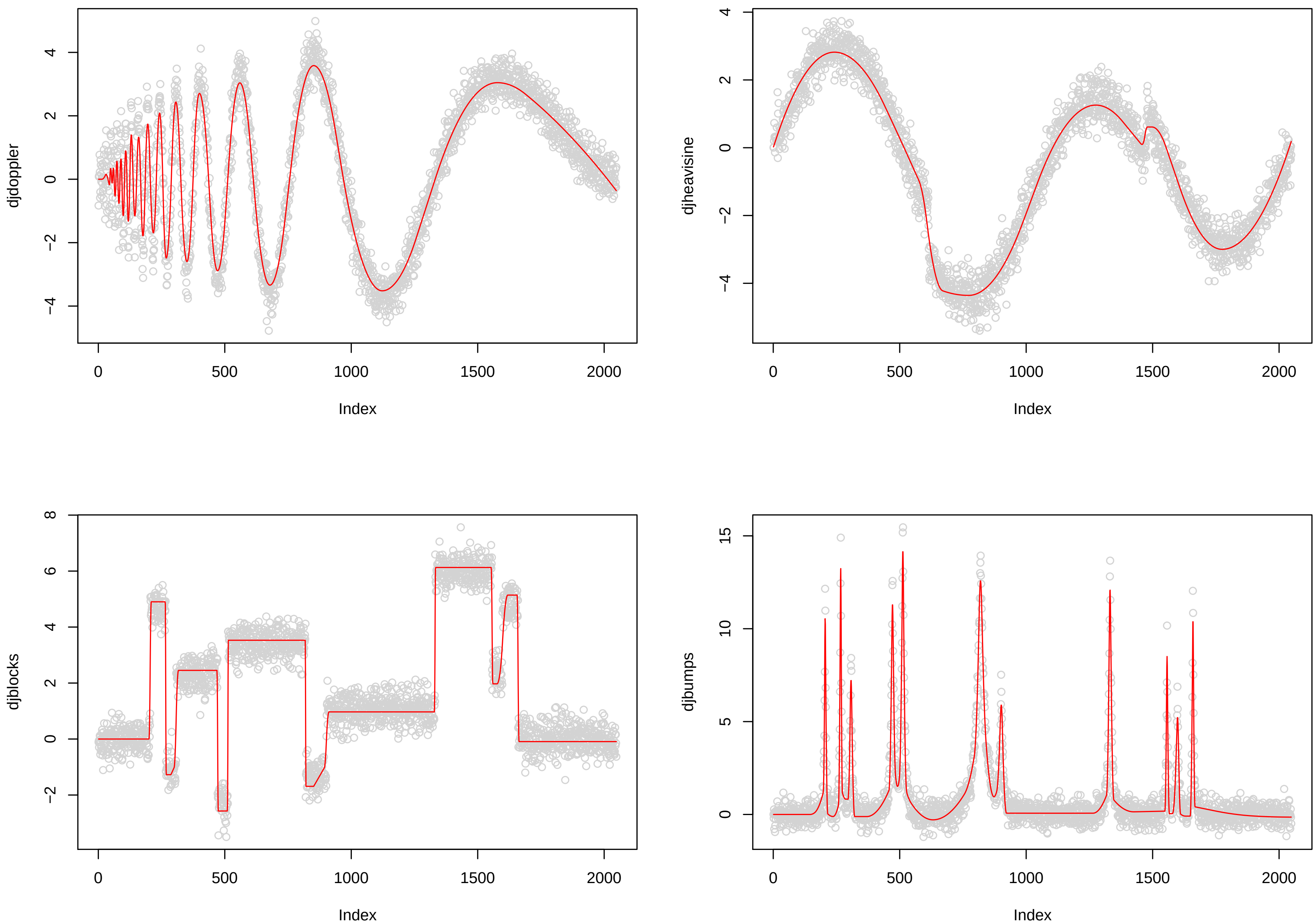
Minimize $\mathrm{TV}(f')$ s.t. MR constraints and monotonicity constraints obtained from f_0:

$\rightarrow f_1$

Minimize $\mathrm{TV}(f'')$ s.t. MR constraints, monotonicity constraints obtained from f_0, and convexity constraints gathered from f_1:

$\rightarrow f_2$

Monotonicity and convexity constraints



Pros and cons

Pros:

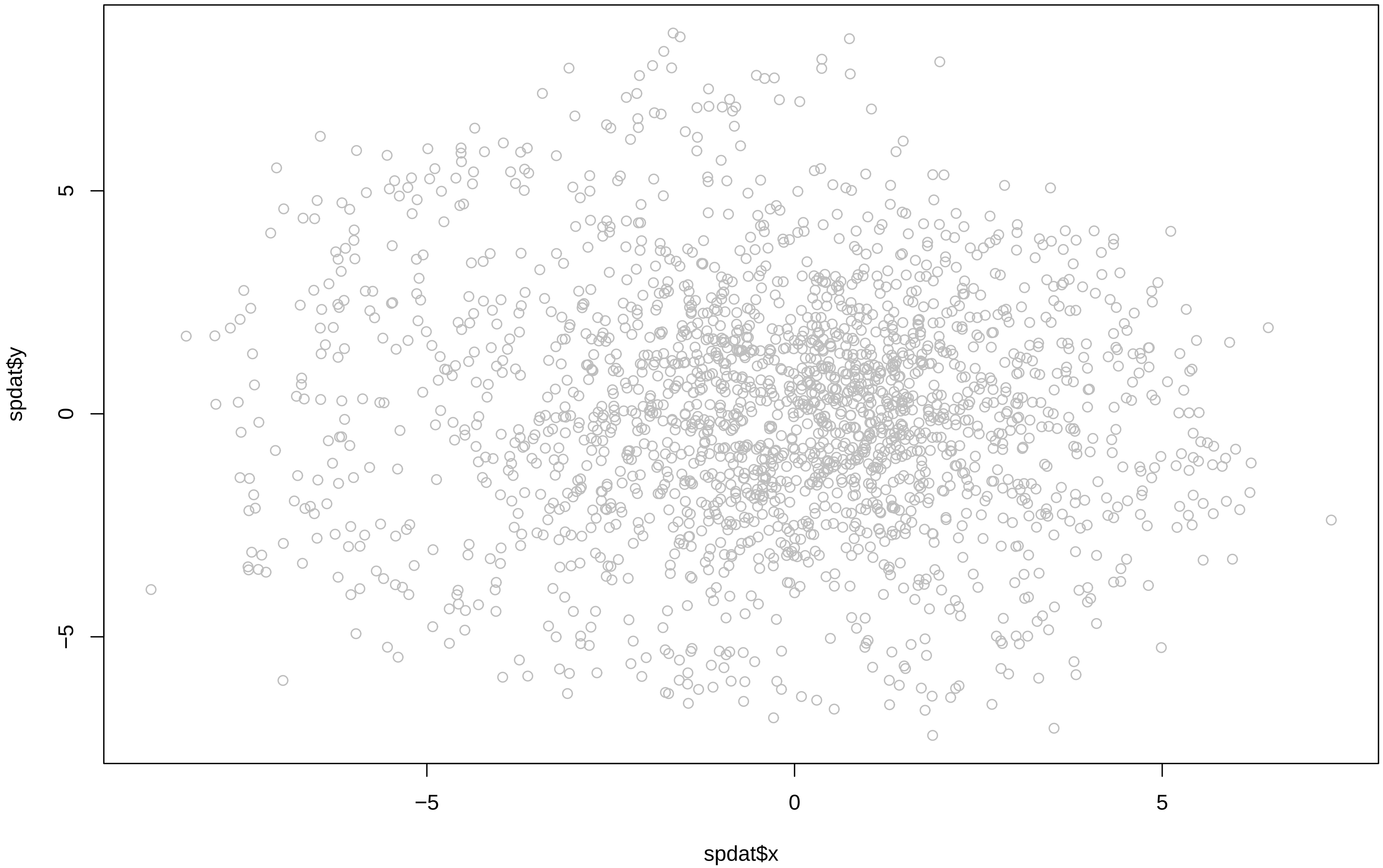
- Nonparametric regression in one line
- Mathematically simple
- Problem of linear programming

  $\longrightarrow$ general solution methods exist

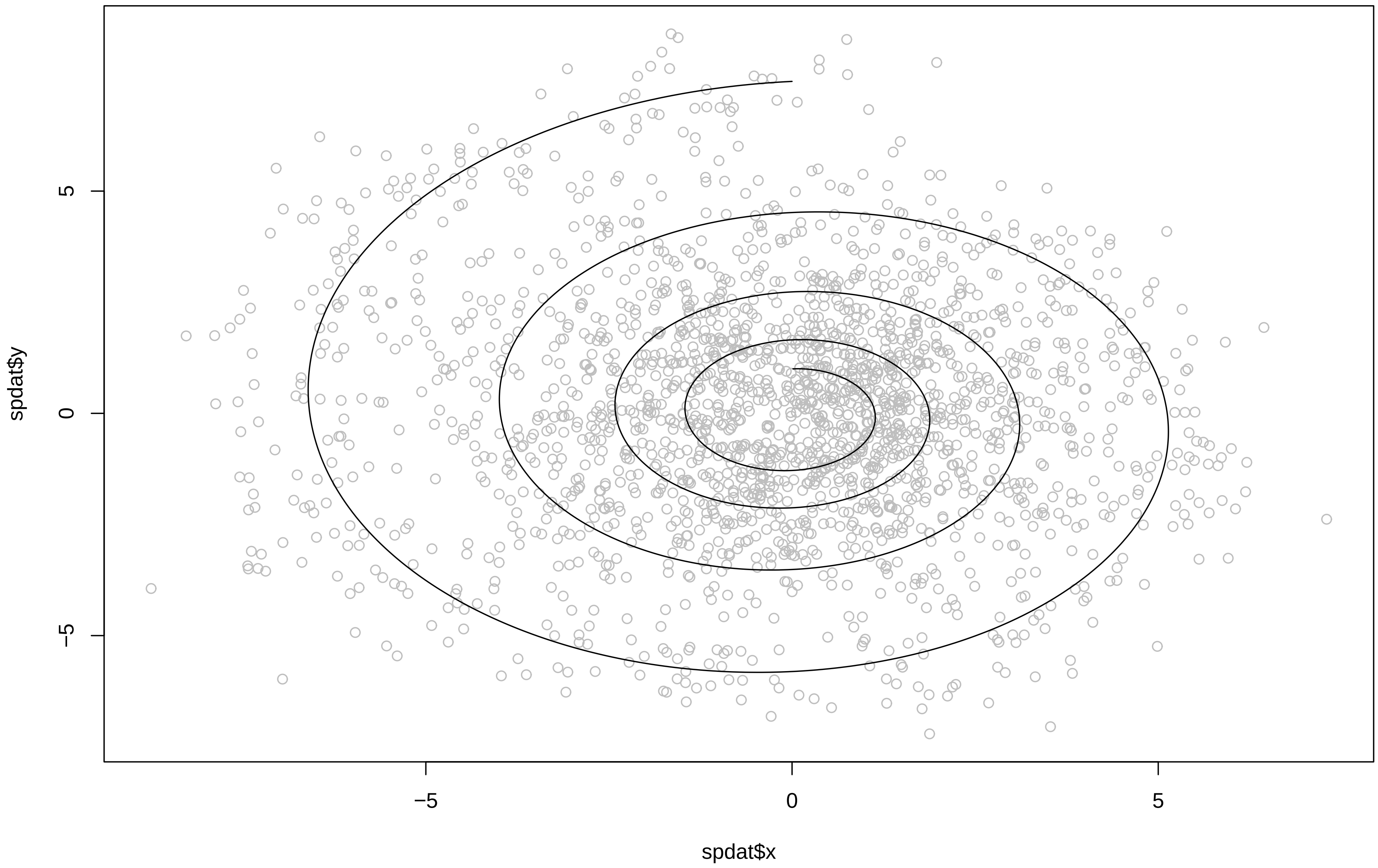
Cons:

- Minimization of f'' can be slow, in particular if data are not smooth.
- No unique solution. Special solution of algorithm has positive bias on isotonic intervals, negative bias on antitonic intervals.

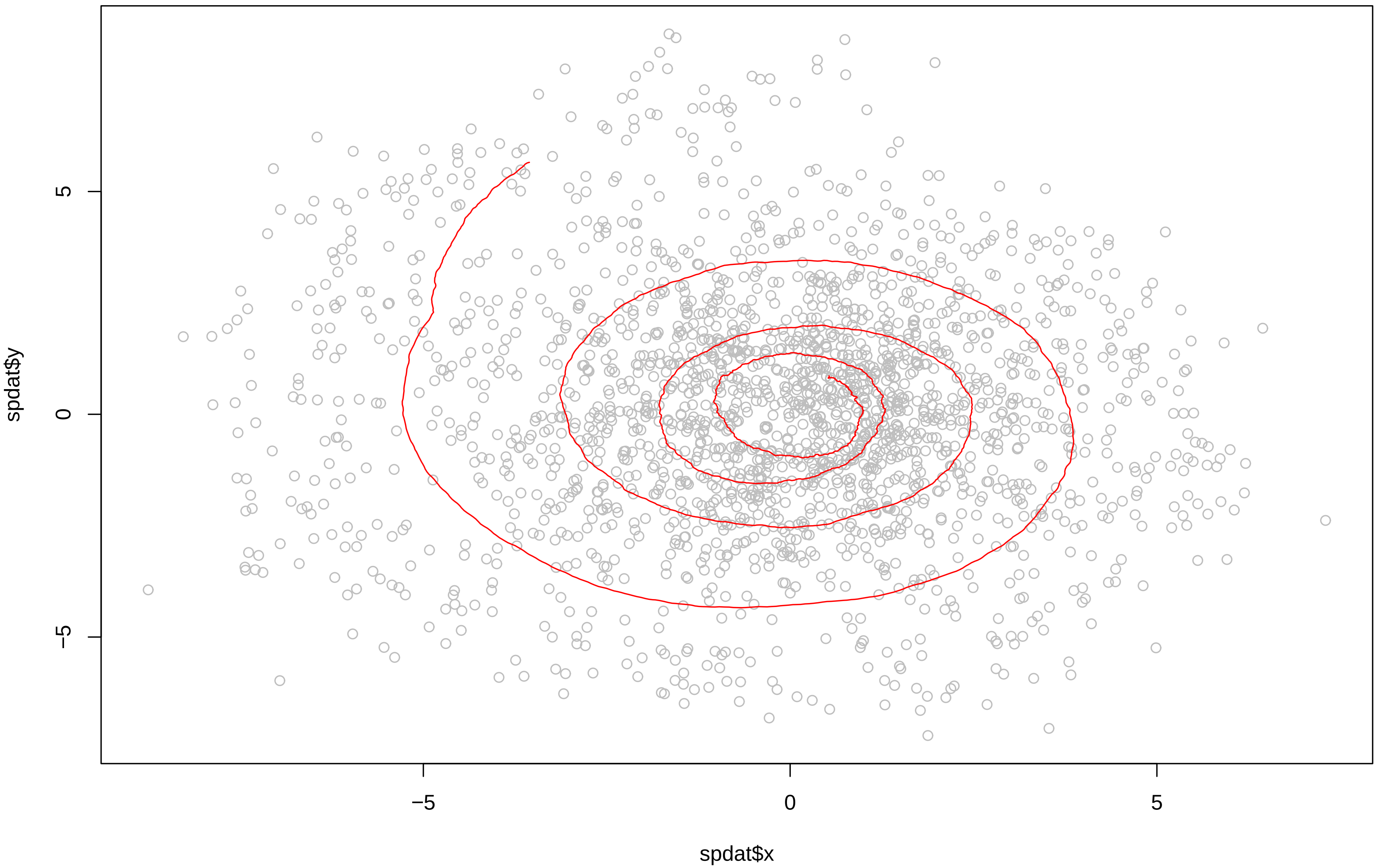
White noise?



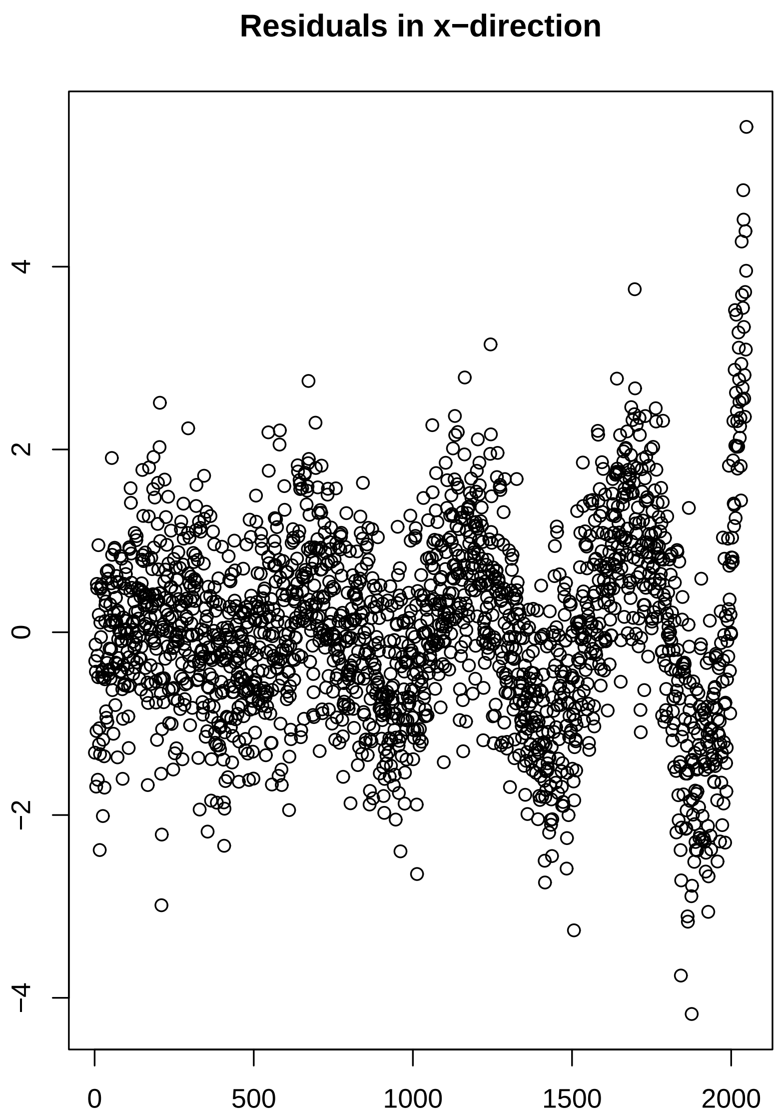
Noisy spiral

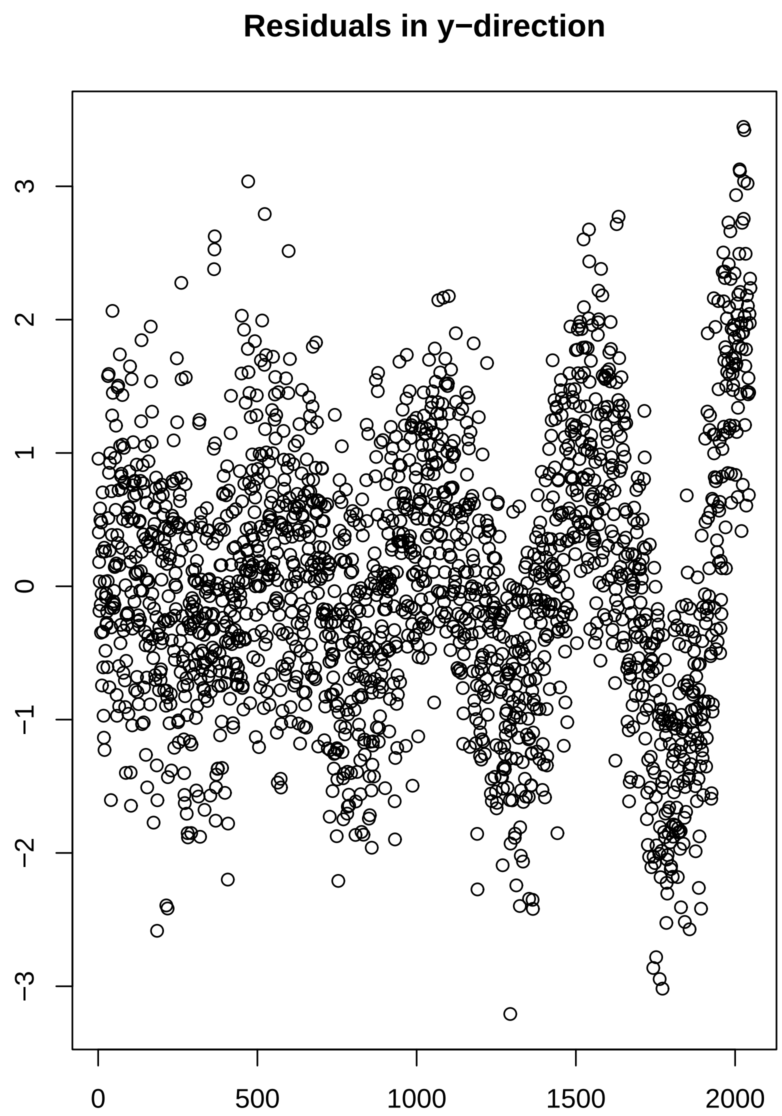


Some classical estimator



Some classical estimator





Multiresolution Criterion

Check residuals in x- and y-directions on different scales and locations:

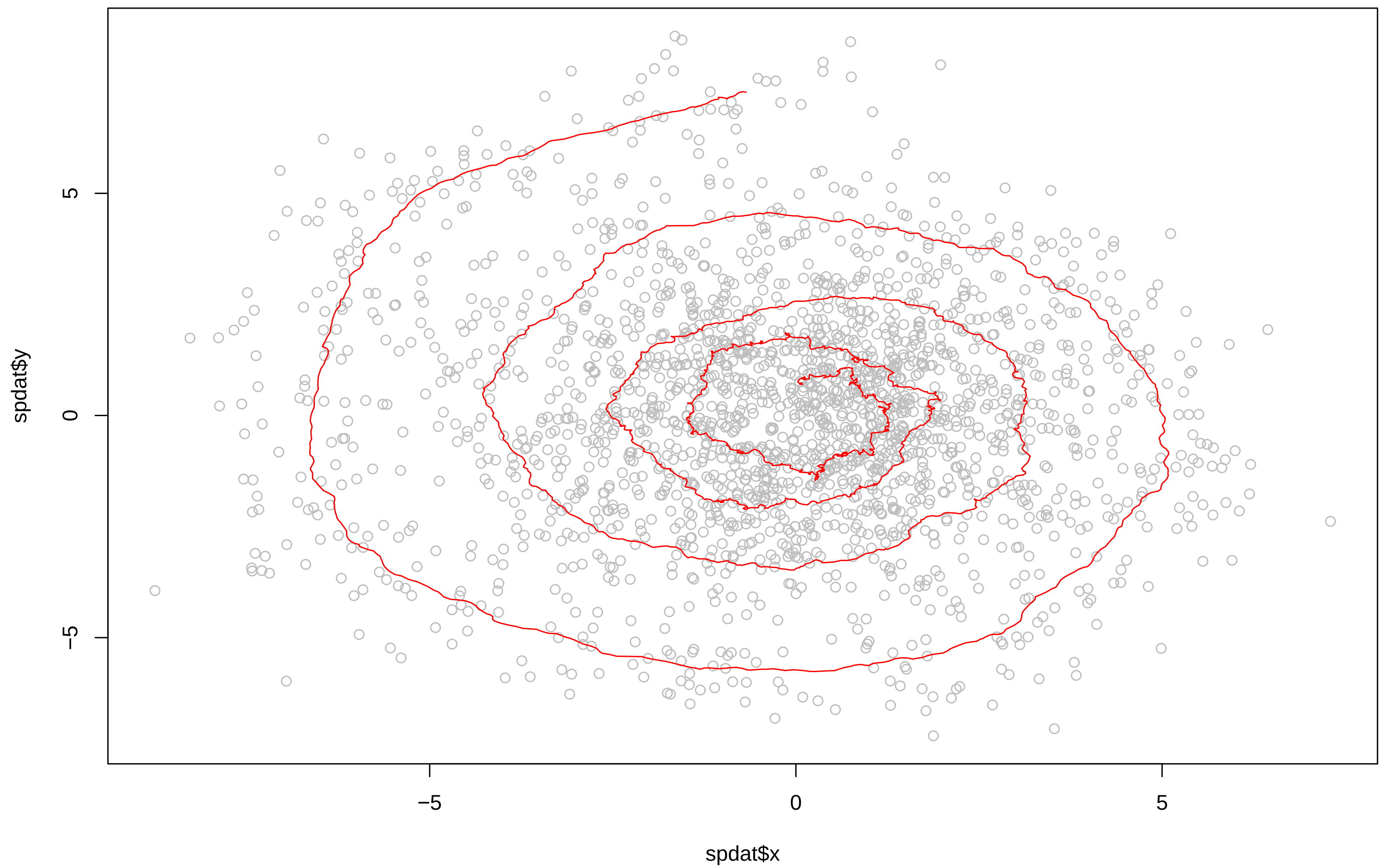
$$\left|\sum_{i\in I}(y_i - f_i^y)\right| < w_I \cdot \sigma$$

$$\left|\sum_{i\in I}(x_i - f_i^x)\right| < w_I \cdot \sigma$$

with $w_I = \sqrt{|I| \cdot 2\log(n)}$ for all intervals I of some family $\mathcal{I}$ of subintervals of $\{1, \ldots, n\}$. (Davies and Kovac, 2001)

Some classical estimator

Largest bandwidth such that multiresolution constraints satisfied.



Problem

Noisy bivariate data $(x_1, y_1), \ldots, (x_n, y_n)$ at time points $t_1, \ldots, t_n$.

Find curve $f = (f^X, f^Y)$ which

- approximates the data and
- is simple (no artificial local extreme values).

The 2D-taut string method

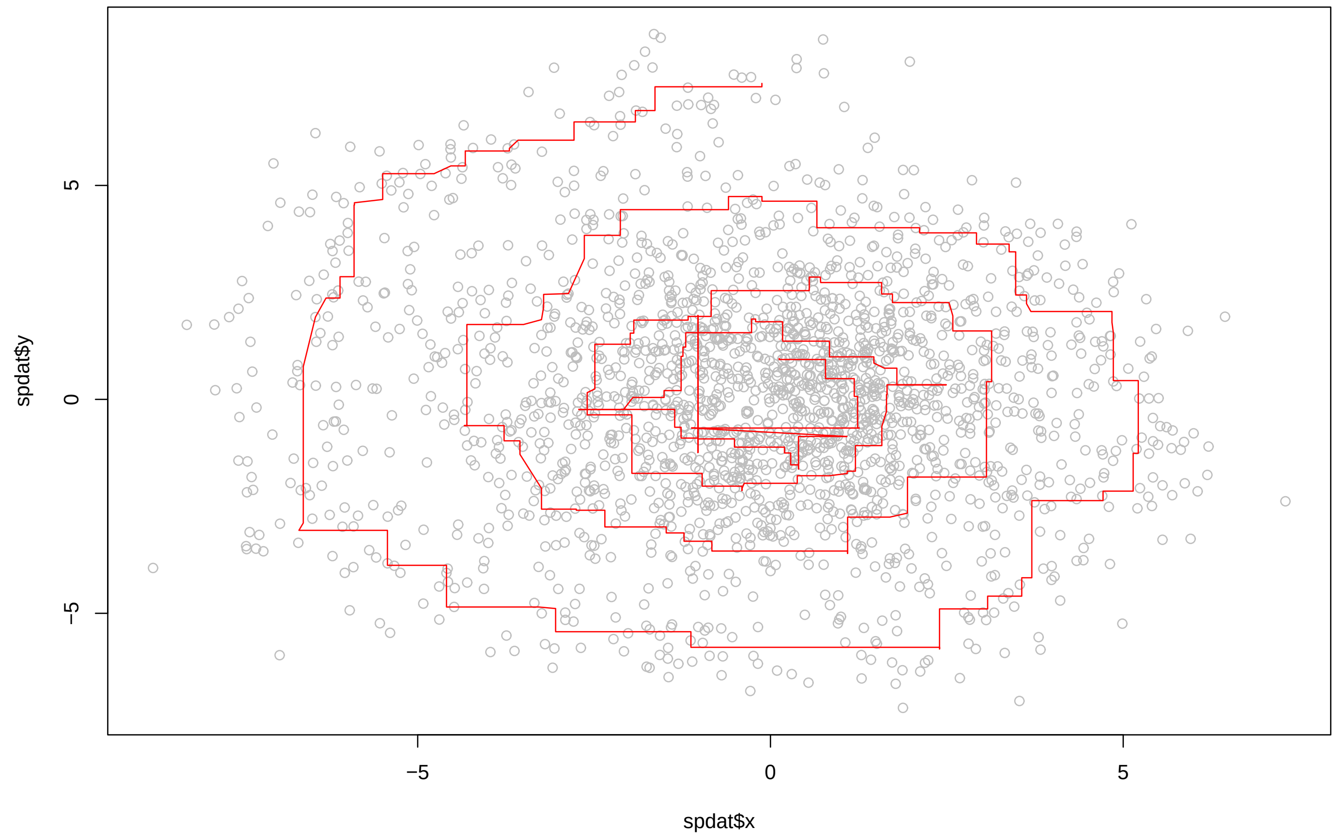
$$\sum_{i=1}^{n}(x_i - f_i^X)^2 + \sum_{i=1}^{n}(y_i - f_i^Y)^2 + \sum_{i=1}^{n-1}\lambda_i|f_{i+1}^X - f_i^X| + \sum_{i=1}^{n-1}\mu_i|f_{i+1}^Y - f_i^Y|$$

Two applications of taut string to x and y-data.

The 2D-taut string method

$$\sum_{i=1}^{n}(x_i - f_i^X)^2 + \sum_{i=1}^{n}(y_i - f_i^Y)^2 + \sum_{i=1}^{n-1}\lambda_i|f_{i+1}^X - f_i^X| + \sum_{i=1}^{n-1}\mu_i|f_{i+1}^Y - f_i^Y|$$

Two applications of taut string to x and y-data.



The 2D-taut string method

Euclidean distances:

$$T(f) = \sum_{i=1}^{n}(x_i - f_i^X)^2 + \sum_{i=1}^{n}(y_i - f_i^Y)^2$$

$$+ \sum_{i=1}^{n-1} \lambda_i \sqrt{(f_{i+1}^X - f_i^X)^2 + (f_{i+1}^Y - f_i^Y)^2}$$

Interpretation as a string

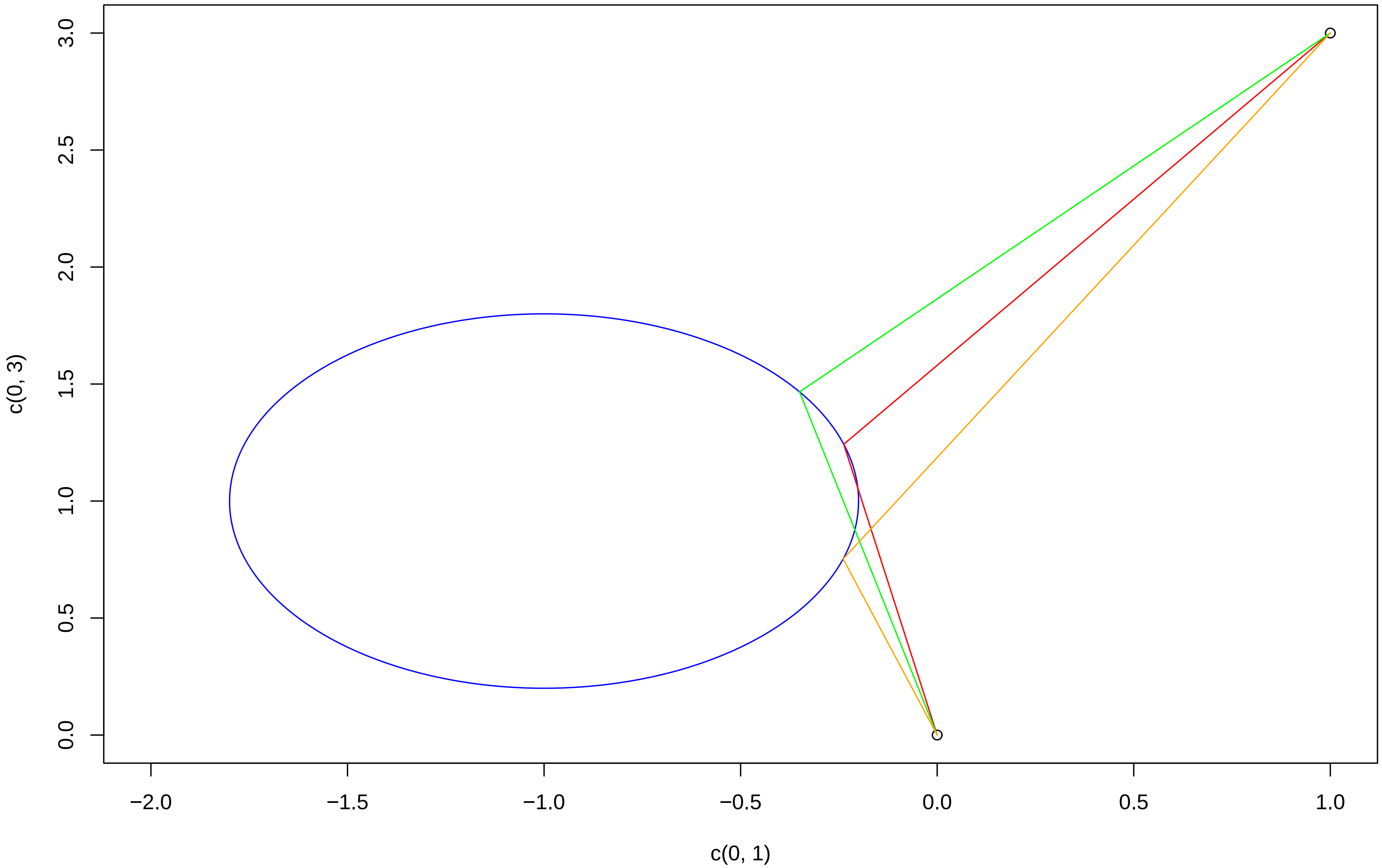
- X and Y: Cumulative sums of the data

$$X_0 = 0, \quad X_i = X_{i-1} + x_i, \quad i = 1, \ldots, n.$$

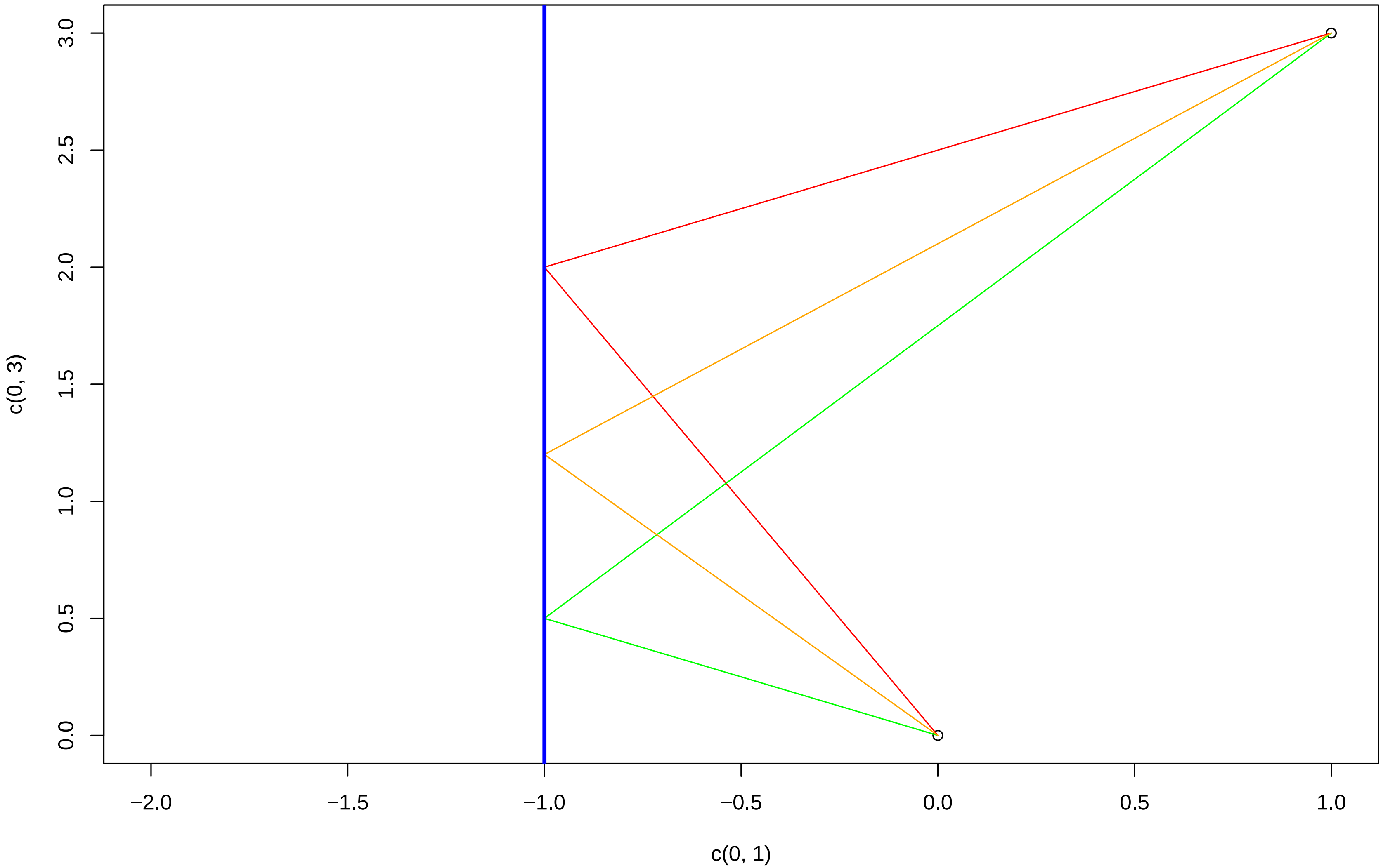
- F^X and F^Y: Cumulative sums of minimiser $f = (f^X, f^Y)$
- (F_i^X, F_i^Y) lies in circle centred at (X_i, Y_i) with radius λ_i.
- F is linear between each two points that lie on the border of a circle.

Is F a taut string?

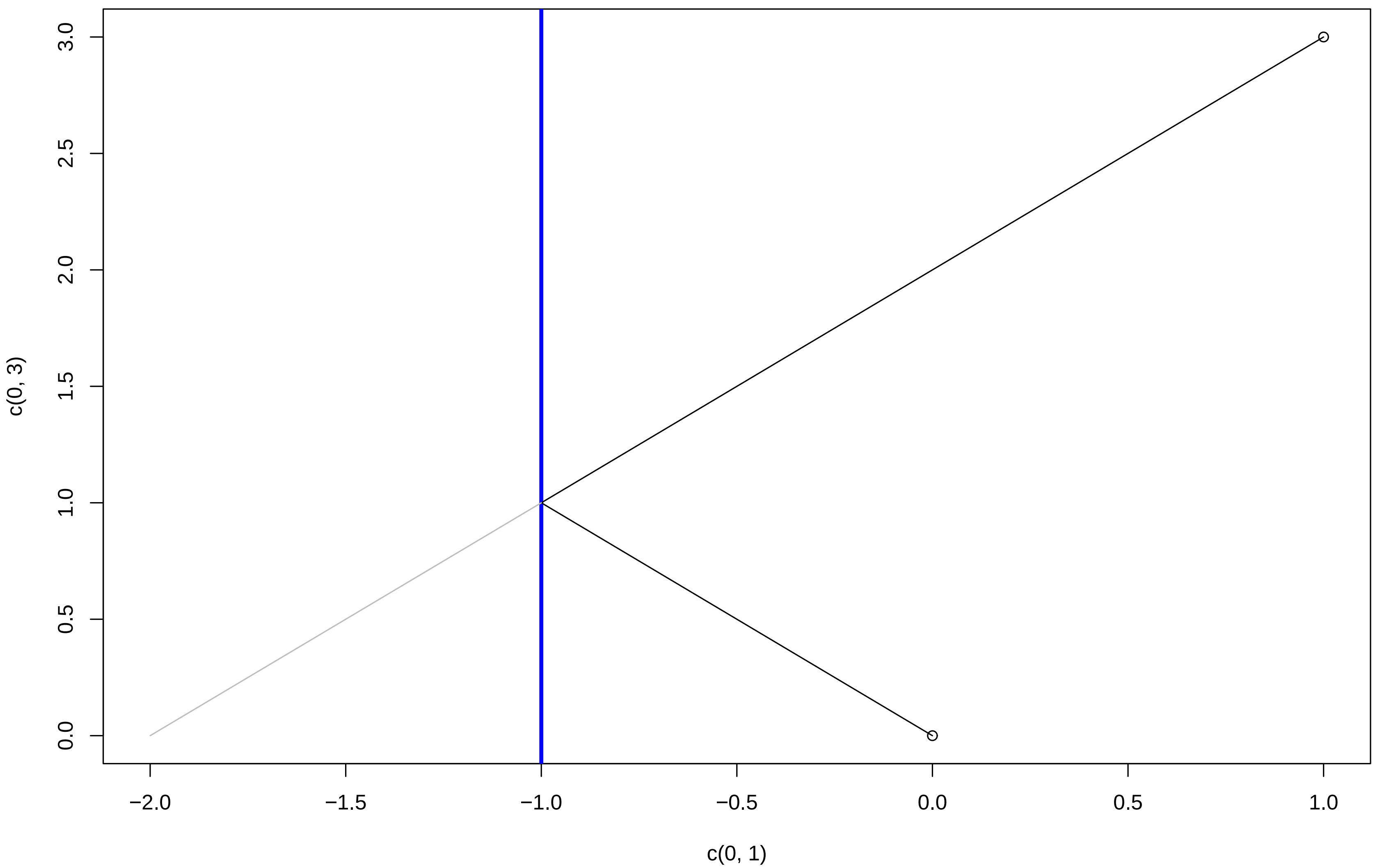
A toy example



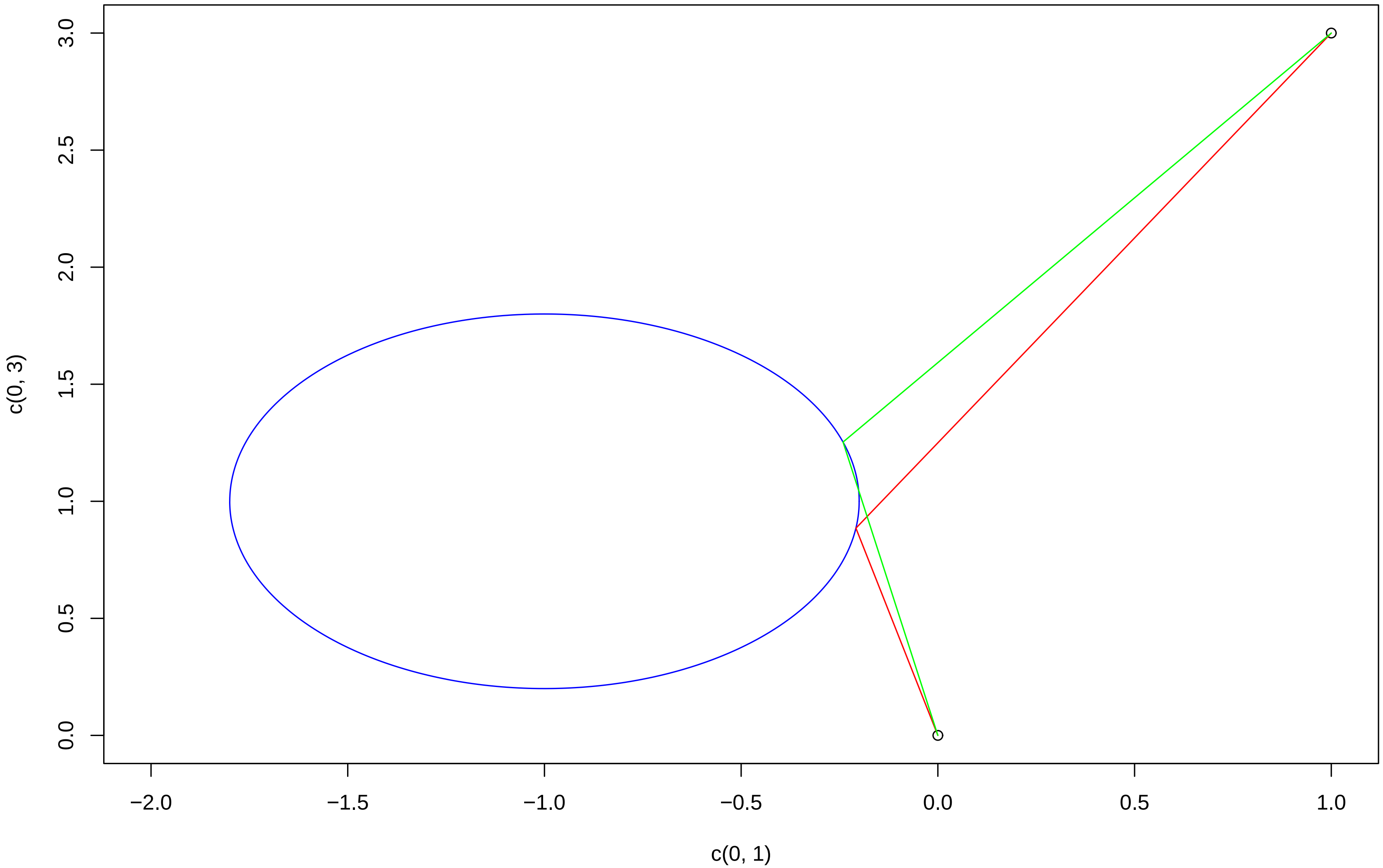
The wanderer in the desert problem



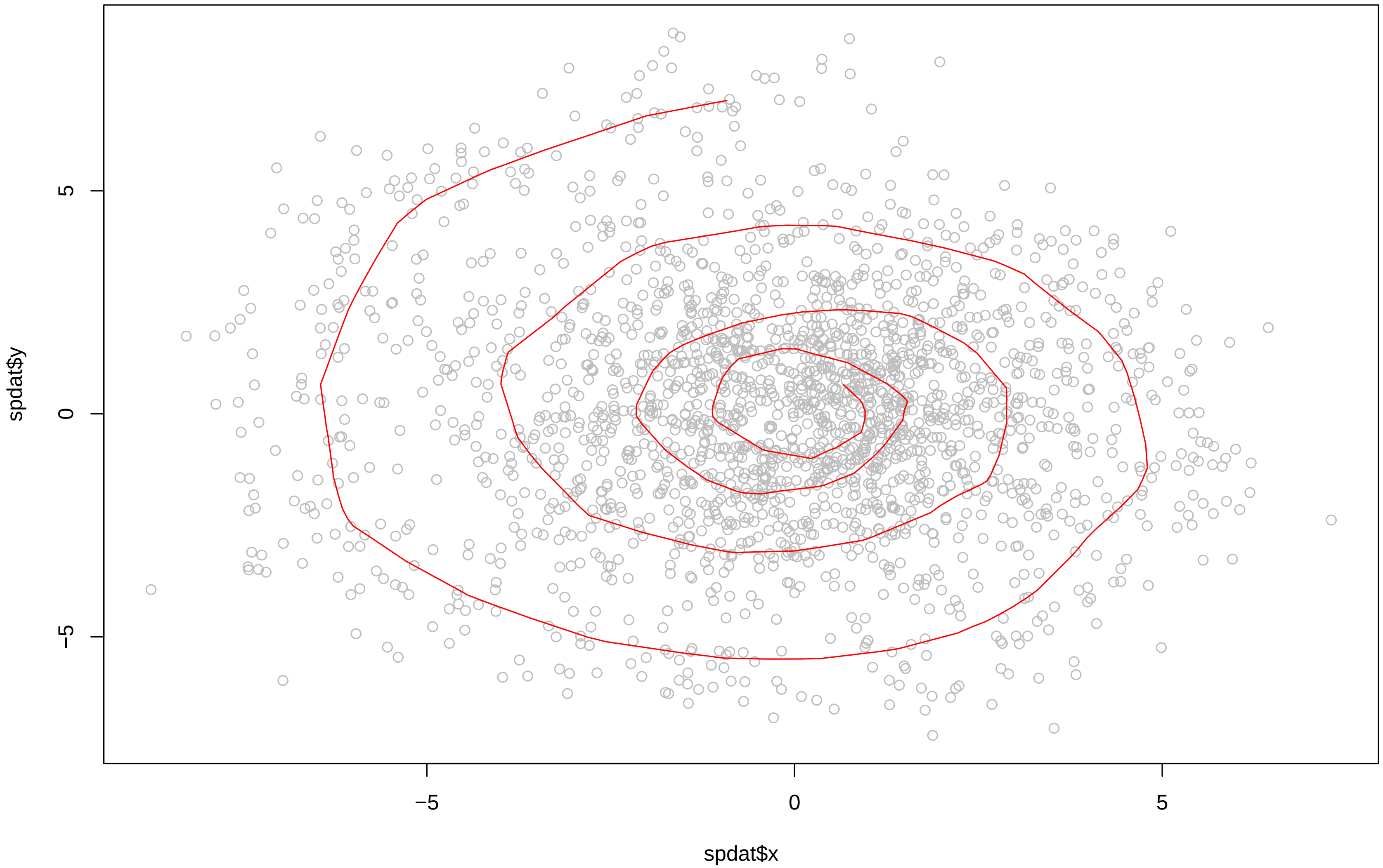
The wanderer in the desert problem



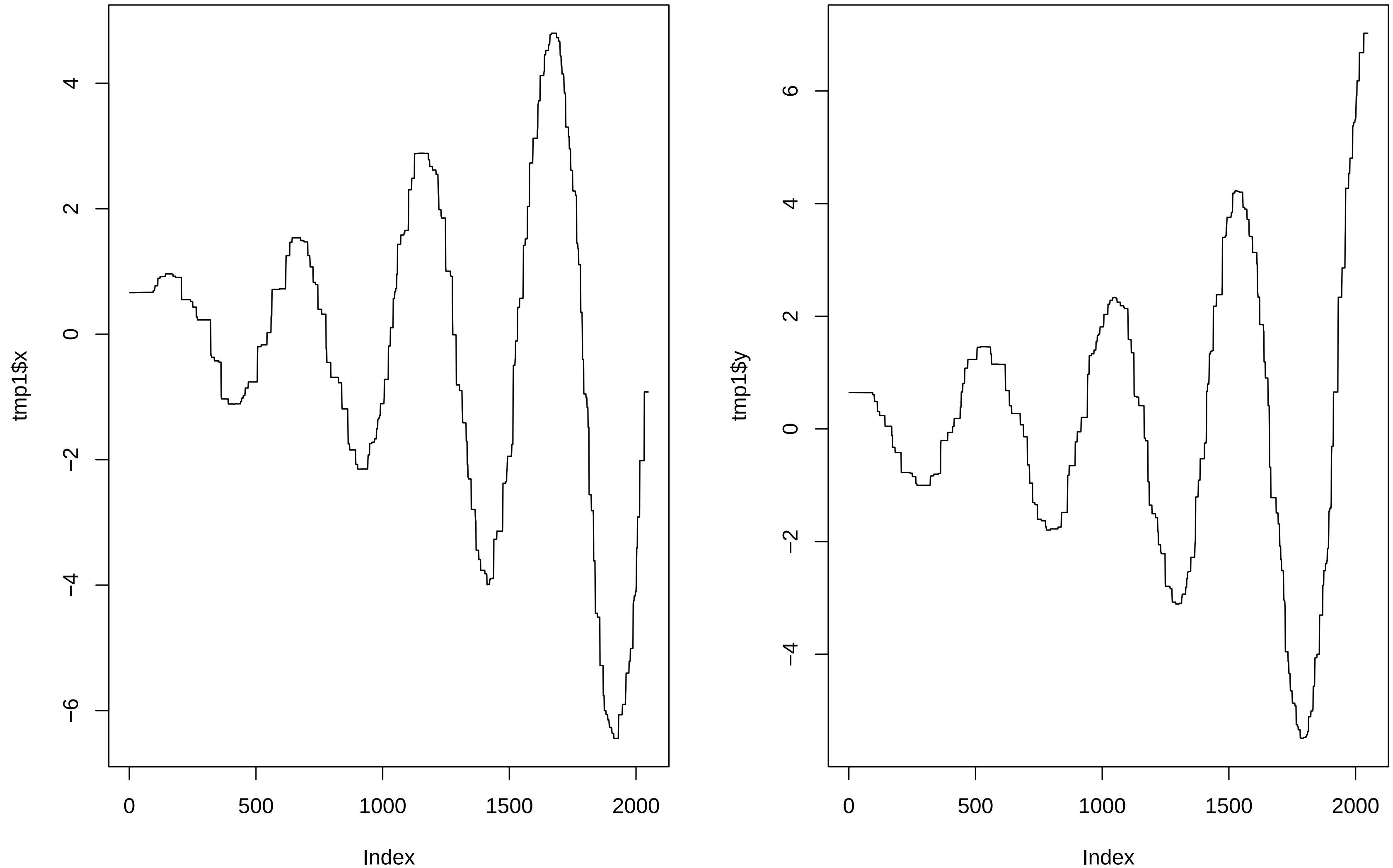
Taut string vs minimiser of functional



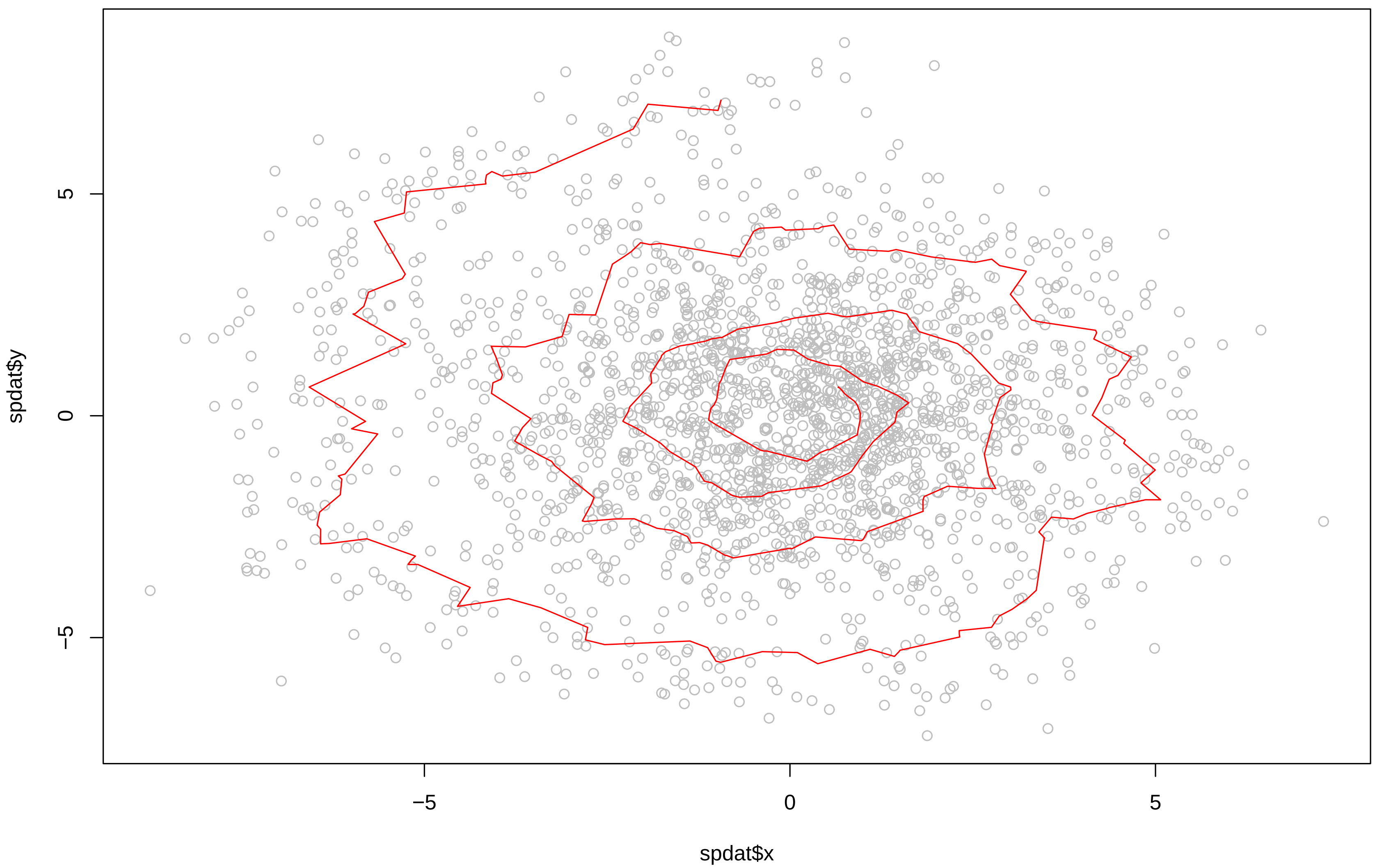
Example revisited



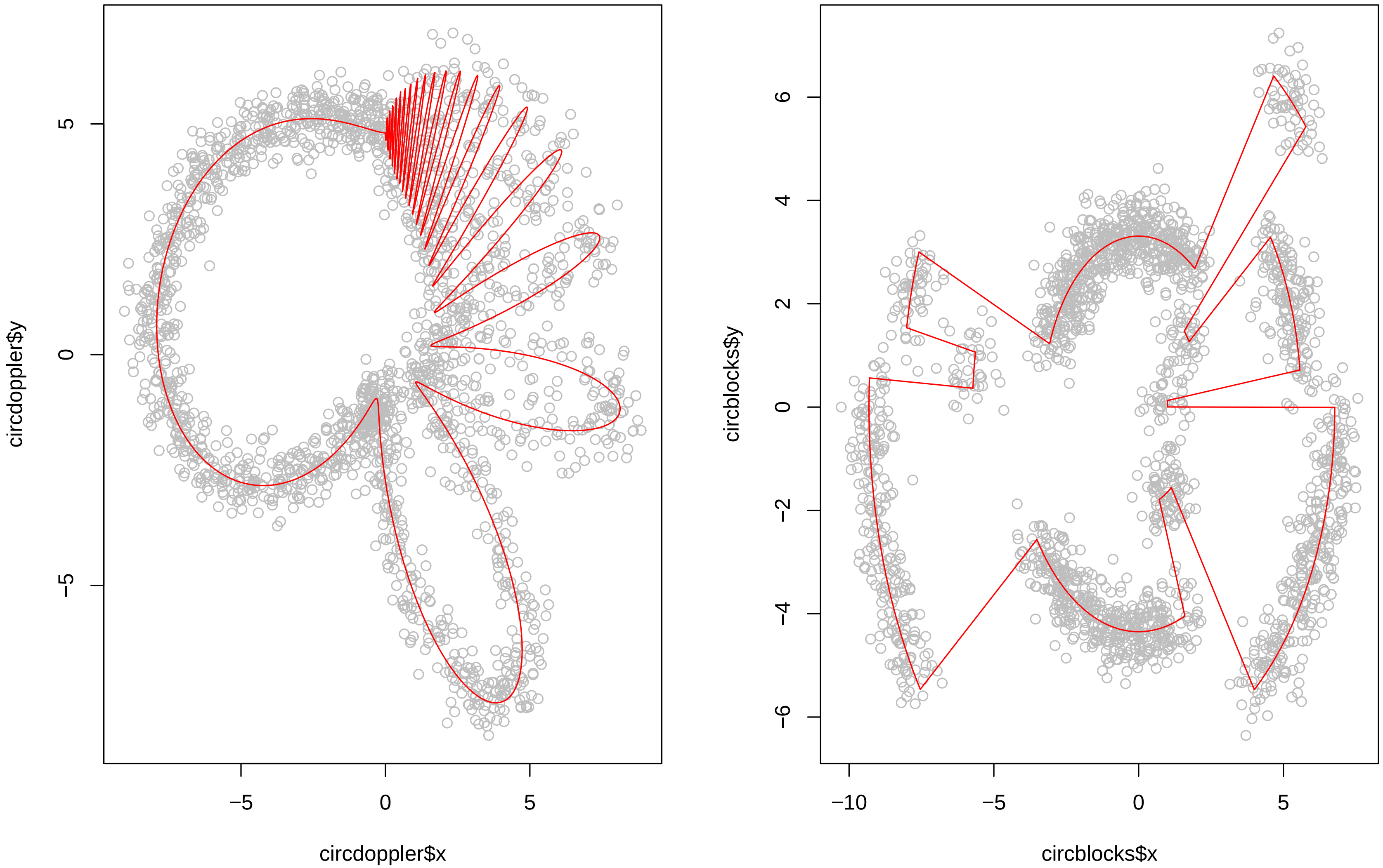
Example revisited



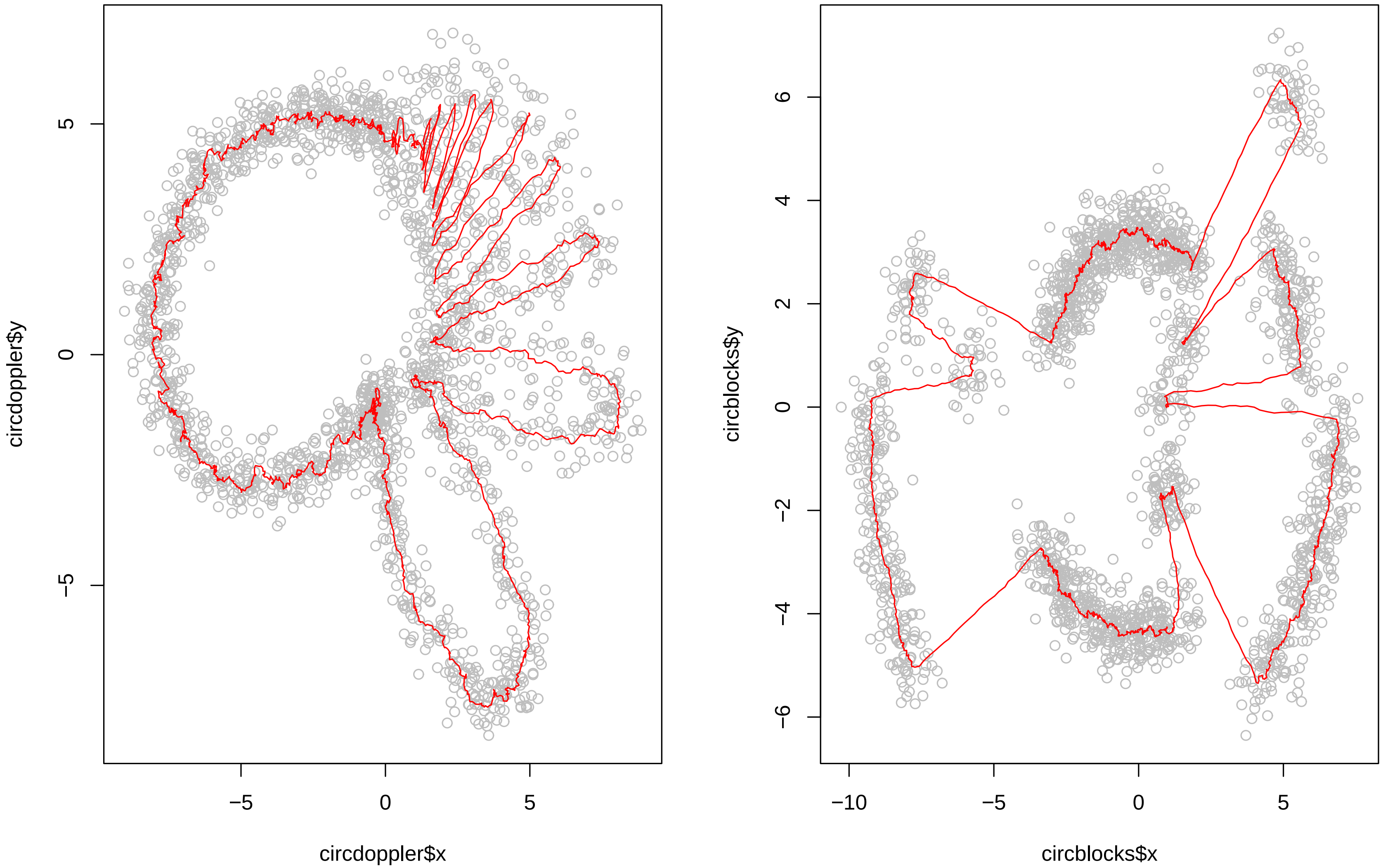
Example revisited



Donoho and Johnstone signals



Donoho and Johnstone signals



Donoho and Johnstone signals

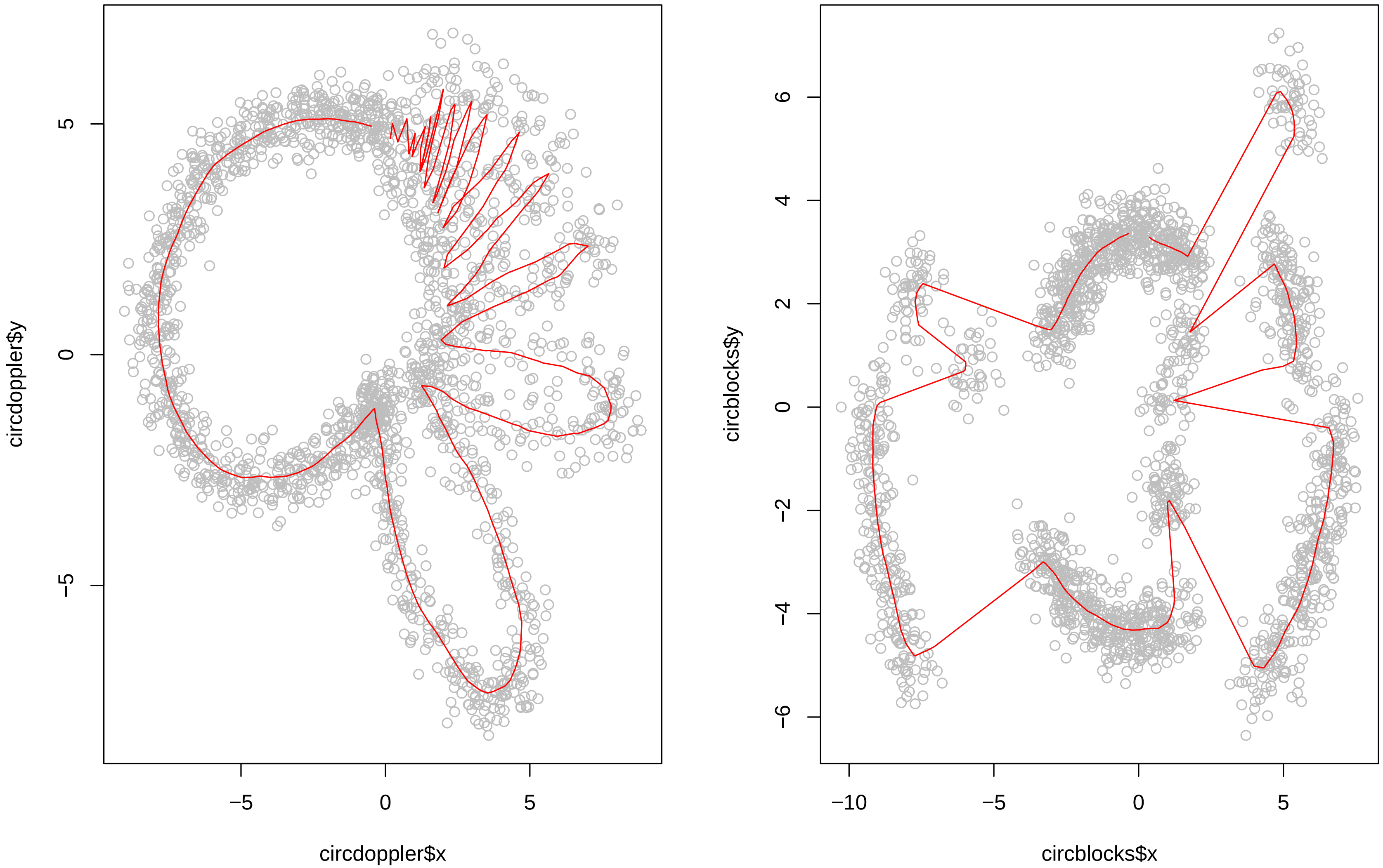
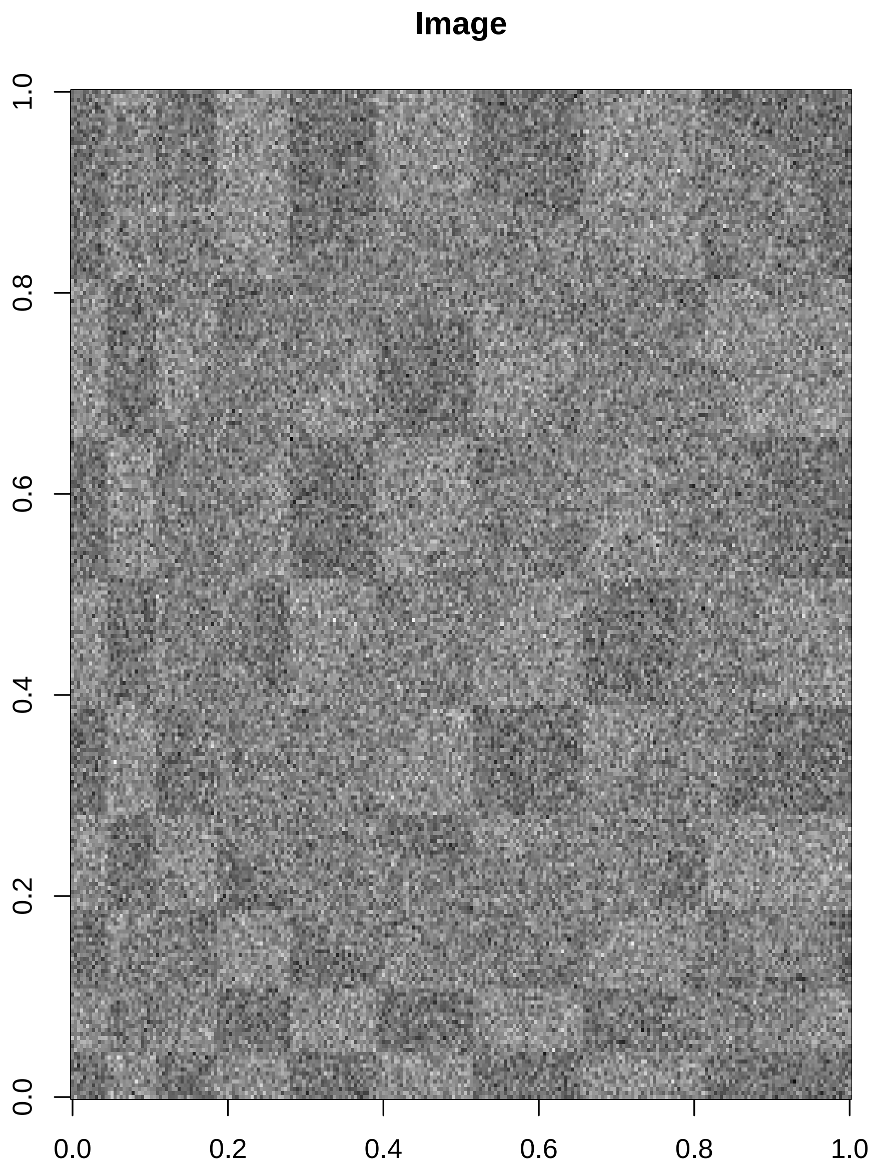


Image analysis



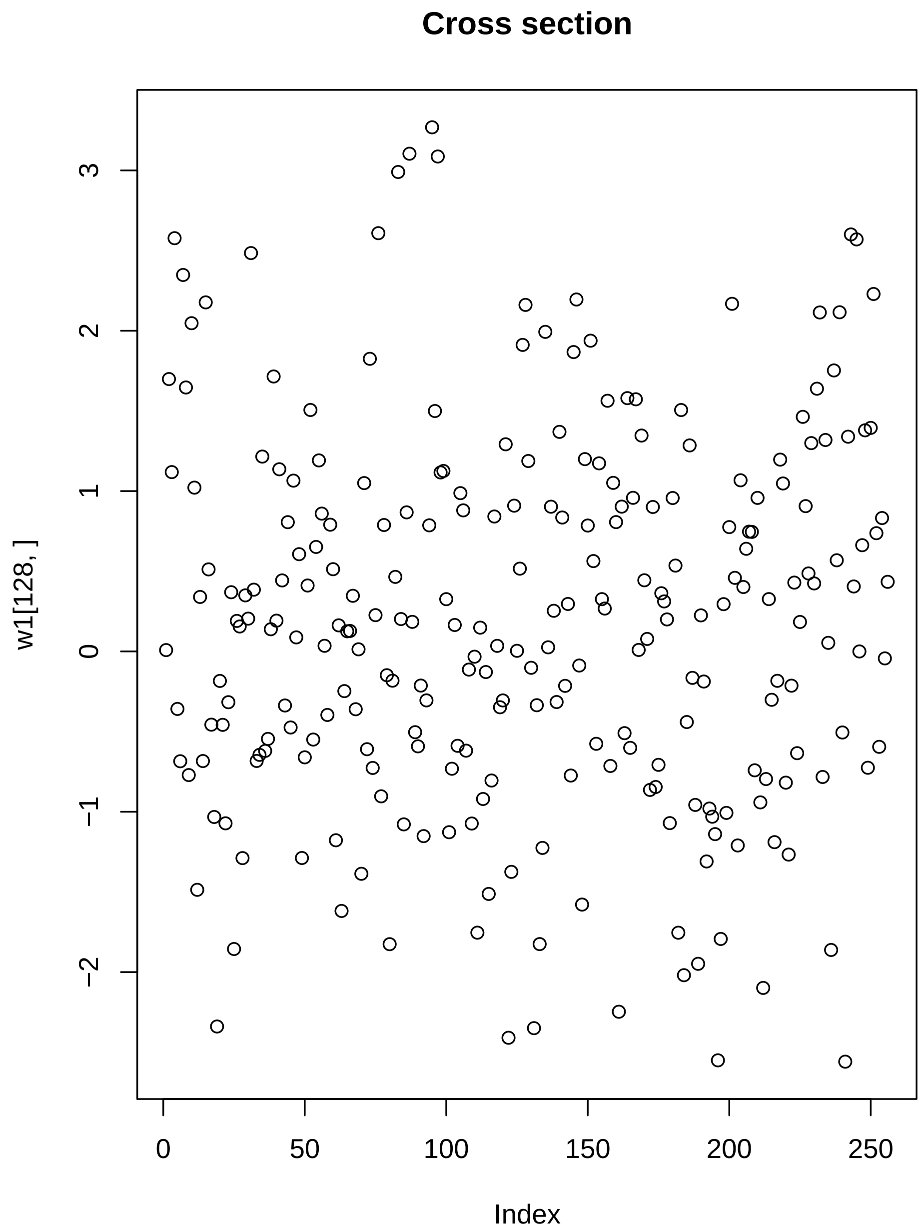


Image analysis

One-dimensional functional:

$$\sum_{i=1}^{n}(y_i - \hat{f}_i)^2 + \sum_{i=1}^{n-1} \lambda_i |\hat{f}_{i+1} - \hat{f}_i|$$

Two-dimensional functional:

$$\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{i,j} - \hat{f}_{i,j})^2 + \sum_{i=1}^{N-1}\sum_{j=1}^{M} \lambda_{i,j} |\hat{f}_{i+1,j} - \hat{f}_{i,j}| + \sum_{i=1}^{N}\sum_{j=1}^{M-1} \mu_{i,j} |\hat{f}_{i,j+1} - \hat{f}_{i,j}|.$$

Image analysis

Penalizing on diagonals:

$$\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{i,j}-\hat{f}_{i,j})^2+\sum_{i=1}^{N-1}\sum_{j=1}^{M}\lambda_{i,j}|\hat{f}_{i+1,j}-\hat{f}_{i,j}|$$

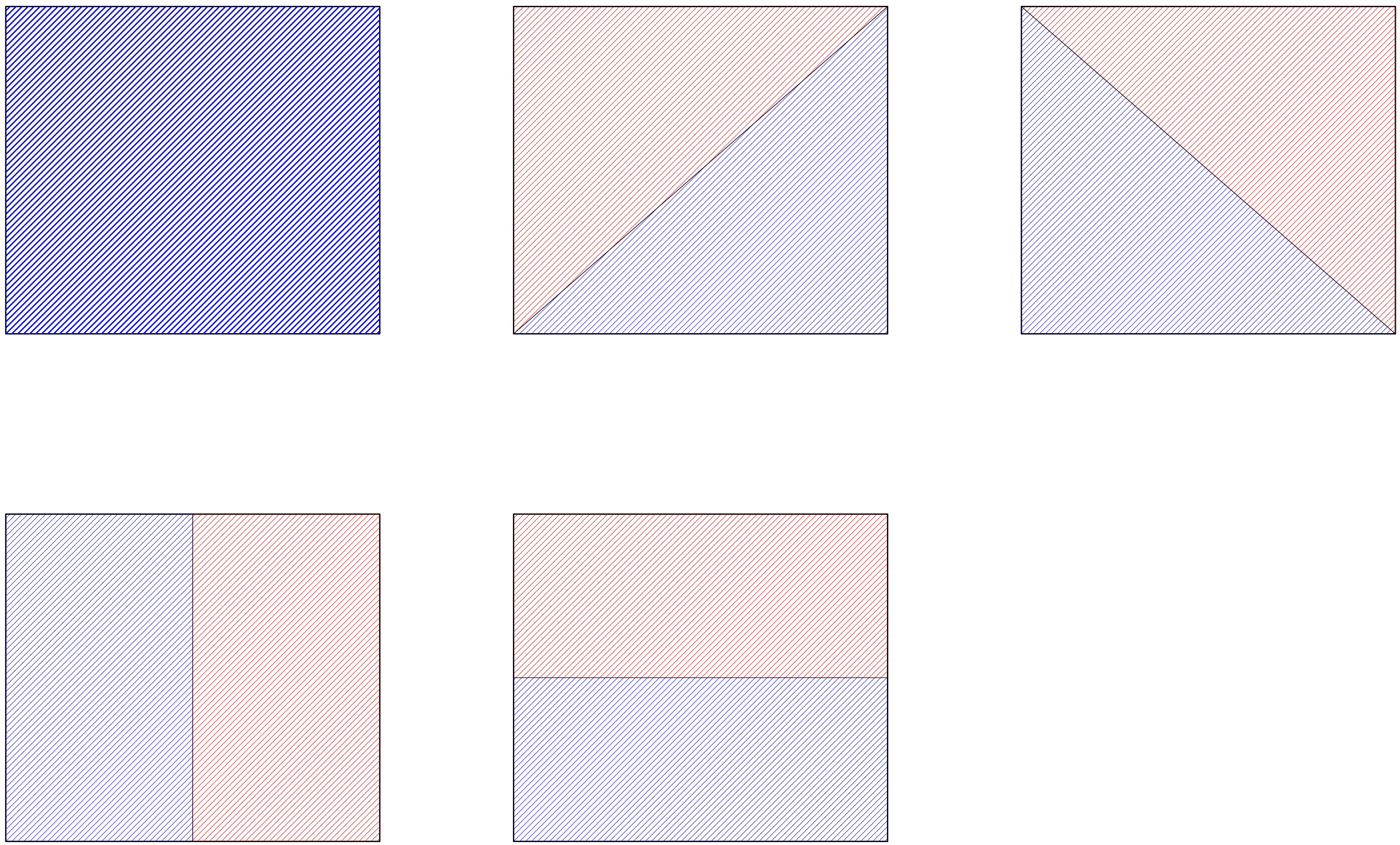
$$+\sum_{i=1}^{N}\sum_{j=1}^{M-1}\mu_{i,j}|\hat{f}_{i,j+1}-\hat{f}_{i,j}|$$

$$+\sum_{i=1}^{N-1}\sum_{j=1}^{M-1}\nu_{i,j}|\hat{f}_{i+1,j+1}-\hat{f}_{i,j}|$$

$$+\sum_{i=1}^{N-1}\sum_{j=2}^{M}\eta_{i,j}|\hat{f}_{i+1,j-1}-\hat{f}_{i,j}|.$$

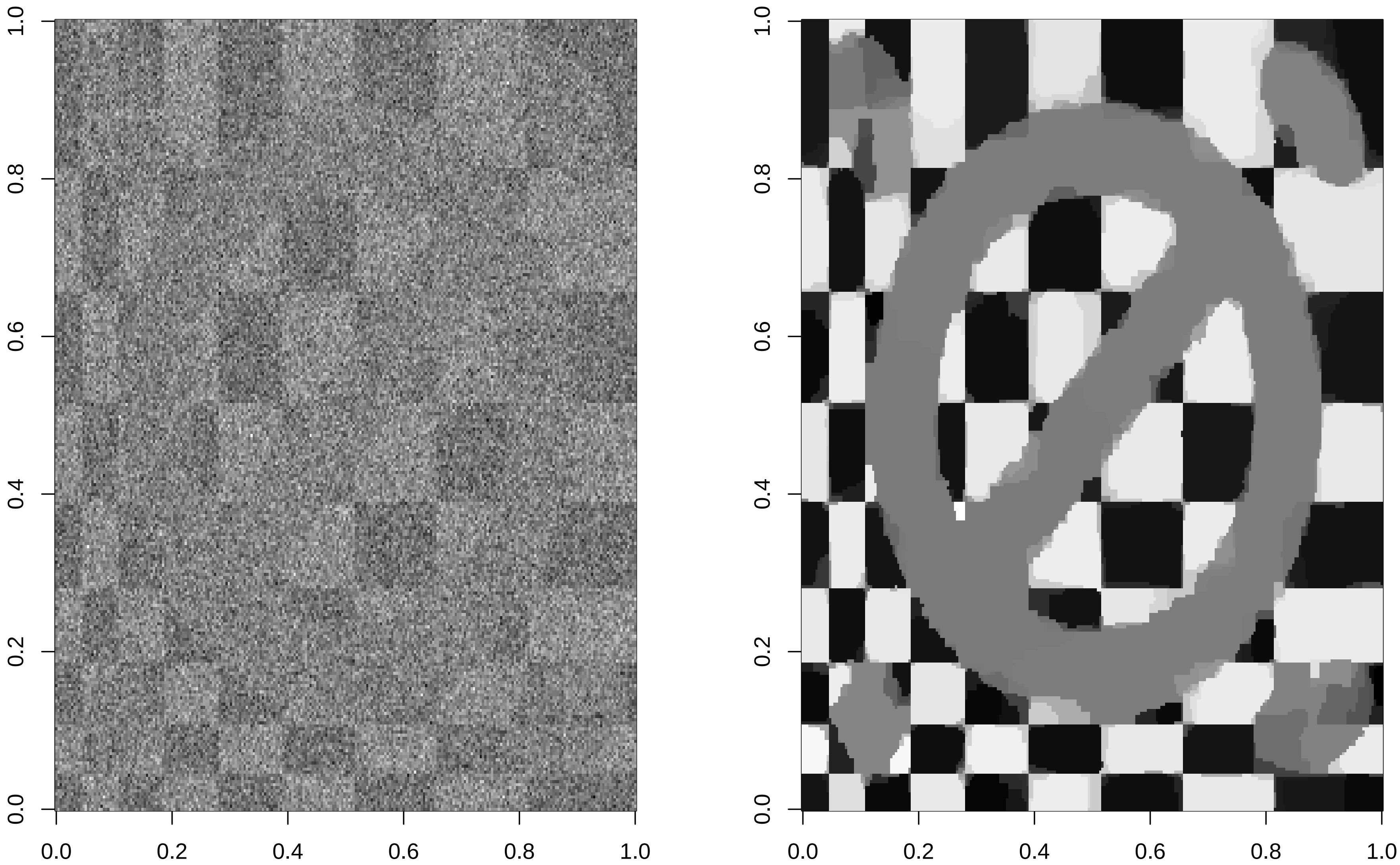
Approximation for images

Check residuals on different scales and locations:

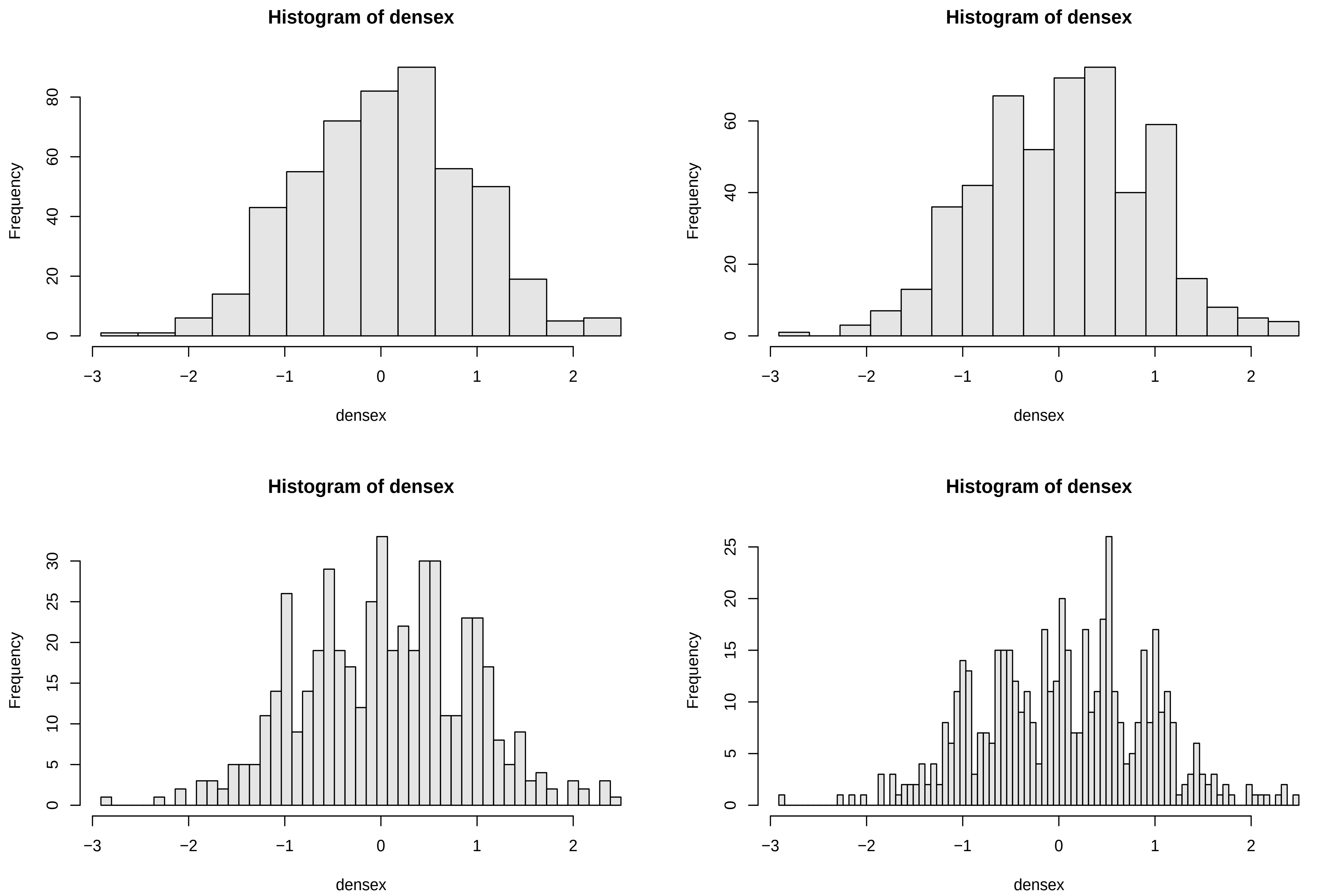


Polzehl and Spokoiny (2001)

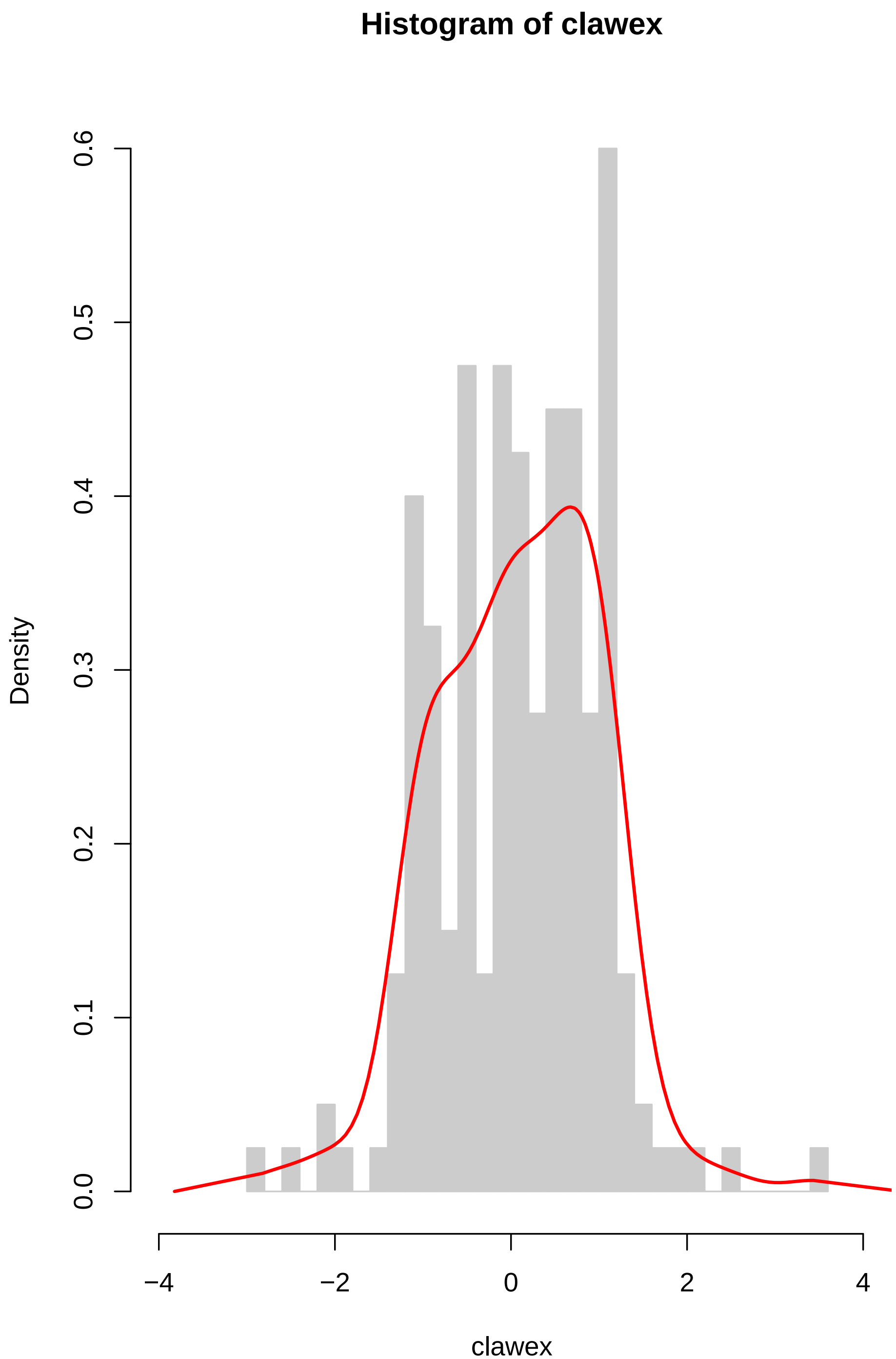
Image analysis

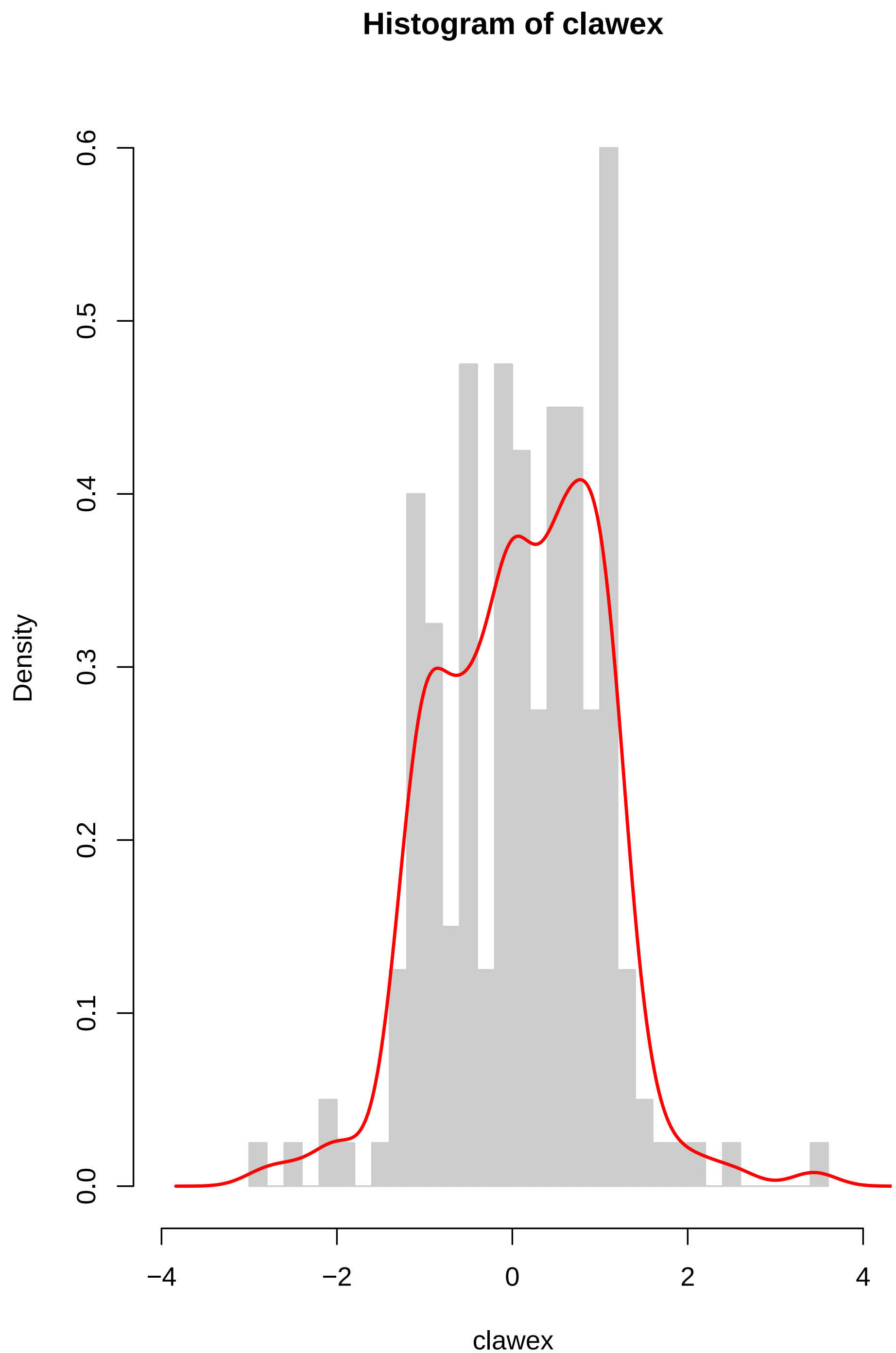


Density estimation

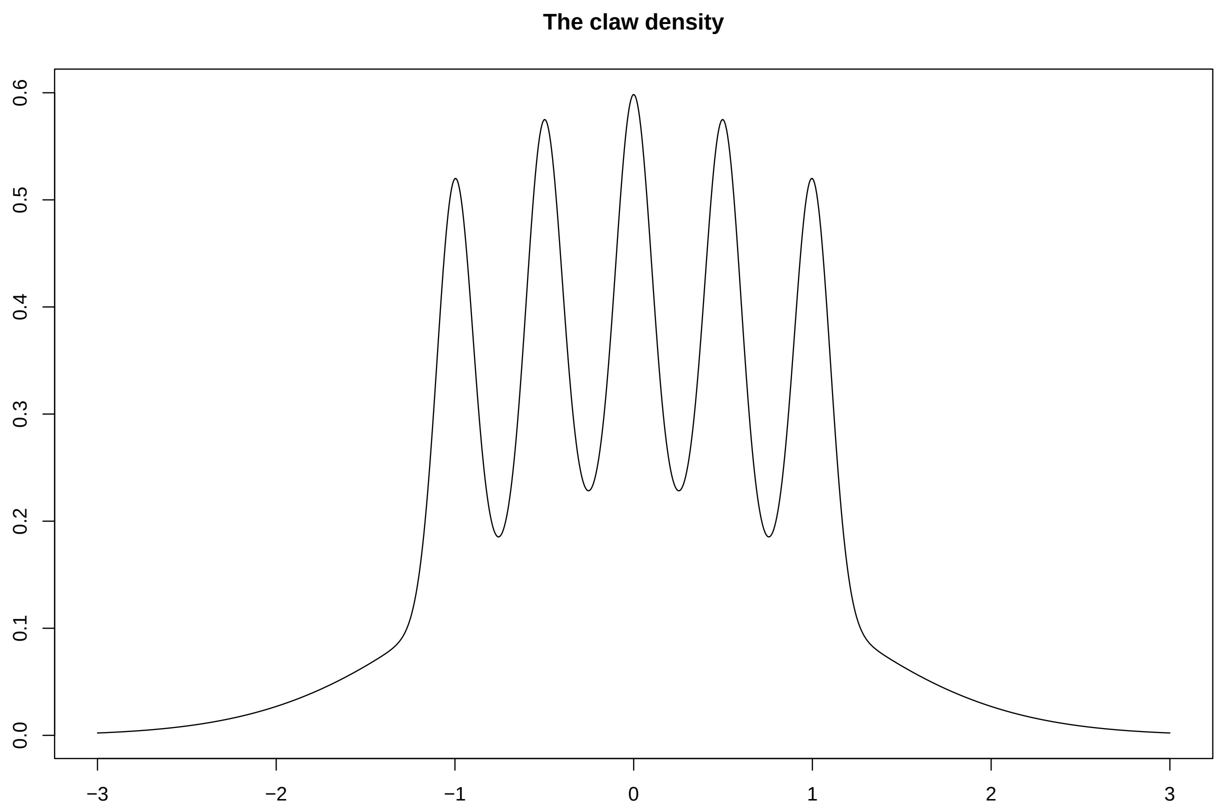


Density estimation





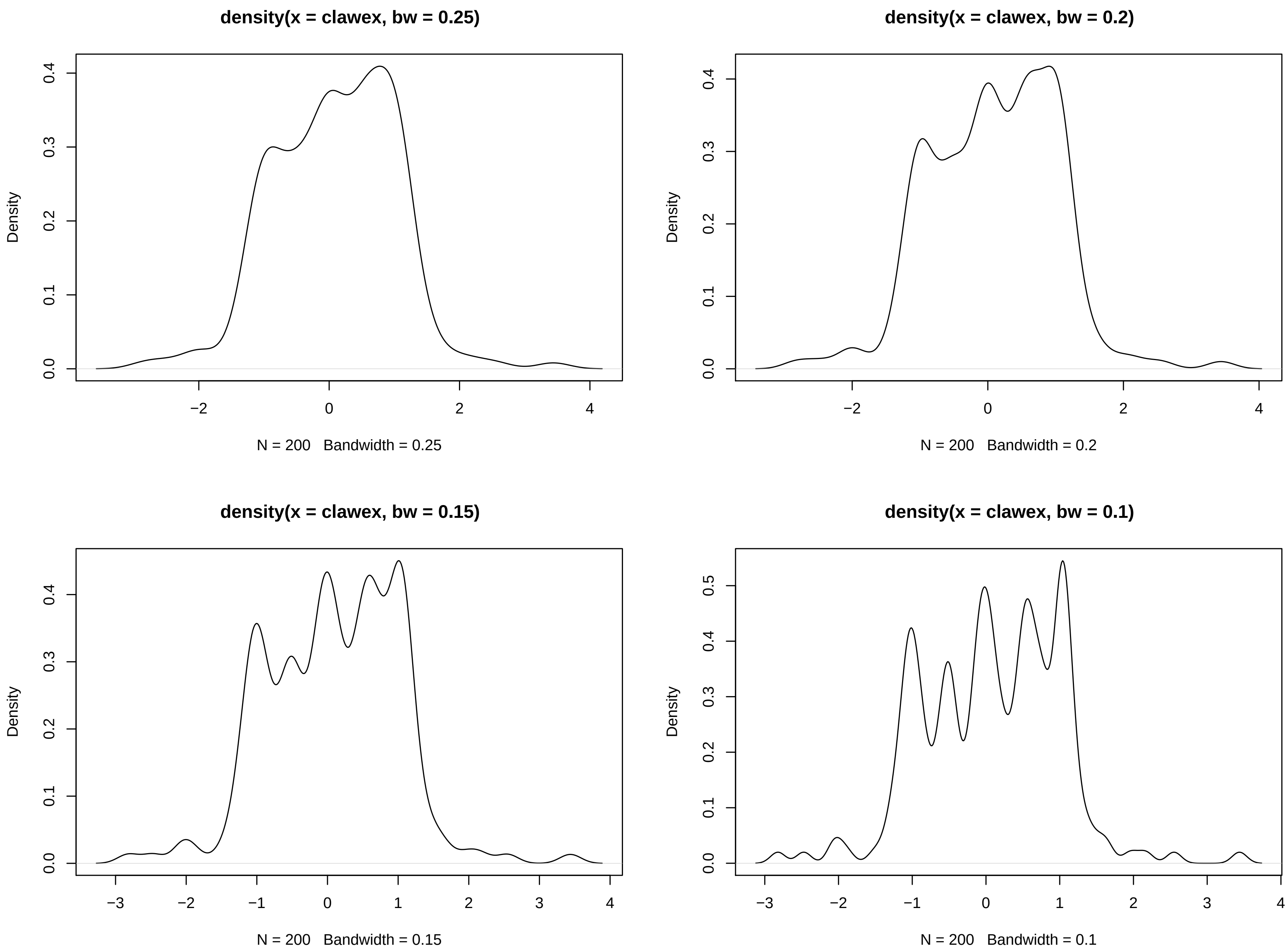
The claw density



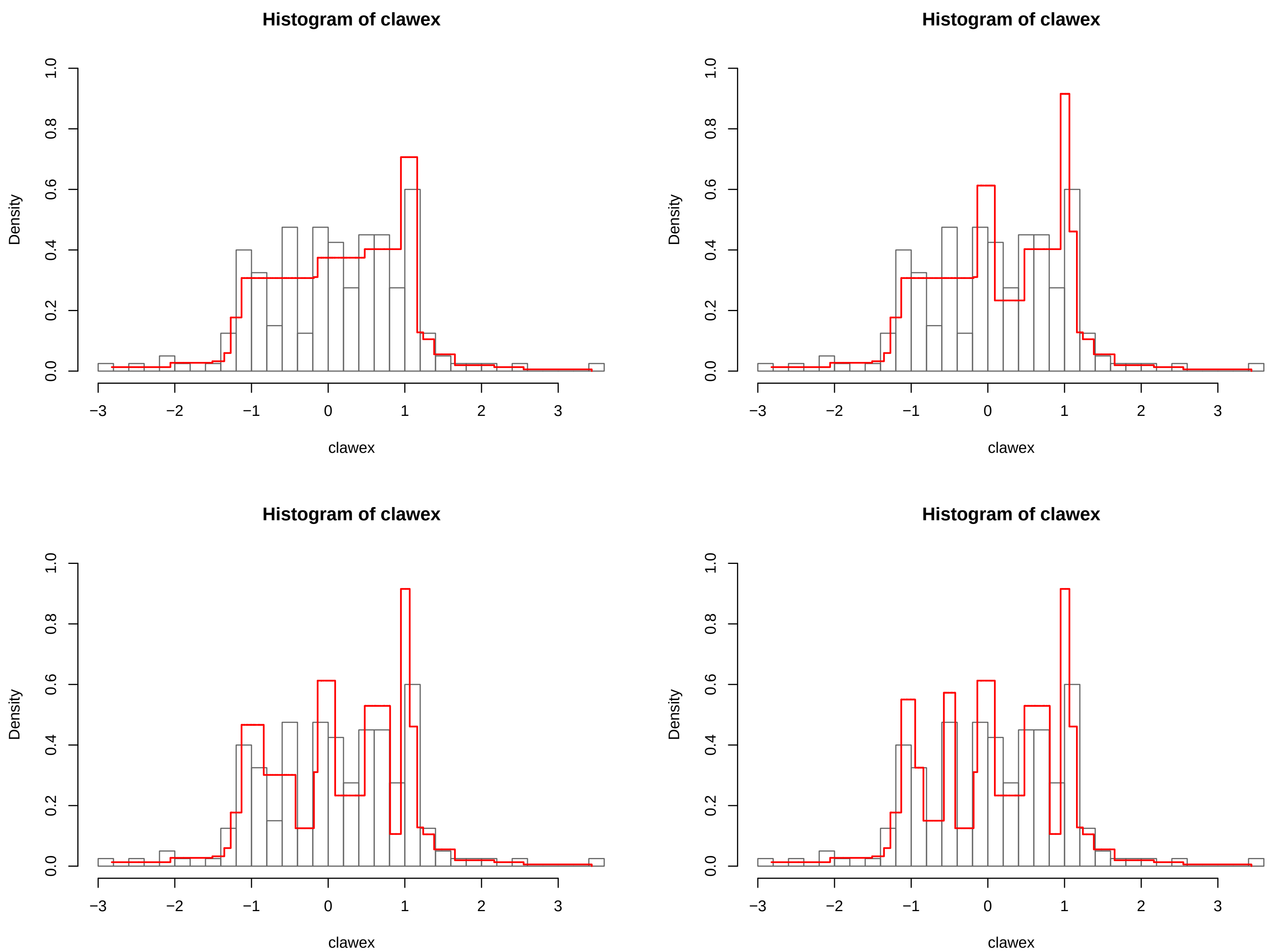
$$\mathbb{Q} = 0.5 * \mathcal{N}(0,1) + 0.1 * \sum_{i=0}^{4} \mathcal{N}(i/2 - 1, 0.1).$$

(Marron and Wand, 1992)

The claw density and kernel estimators

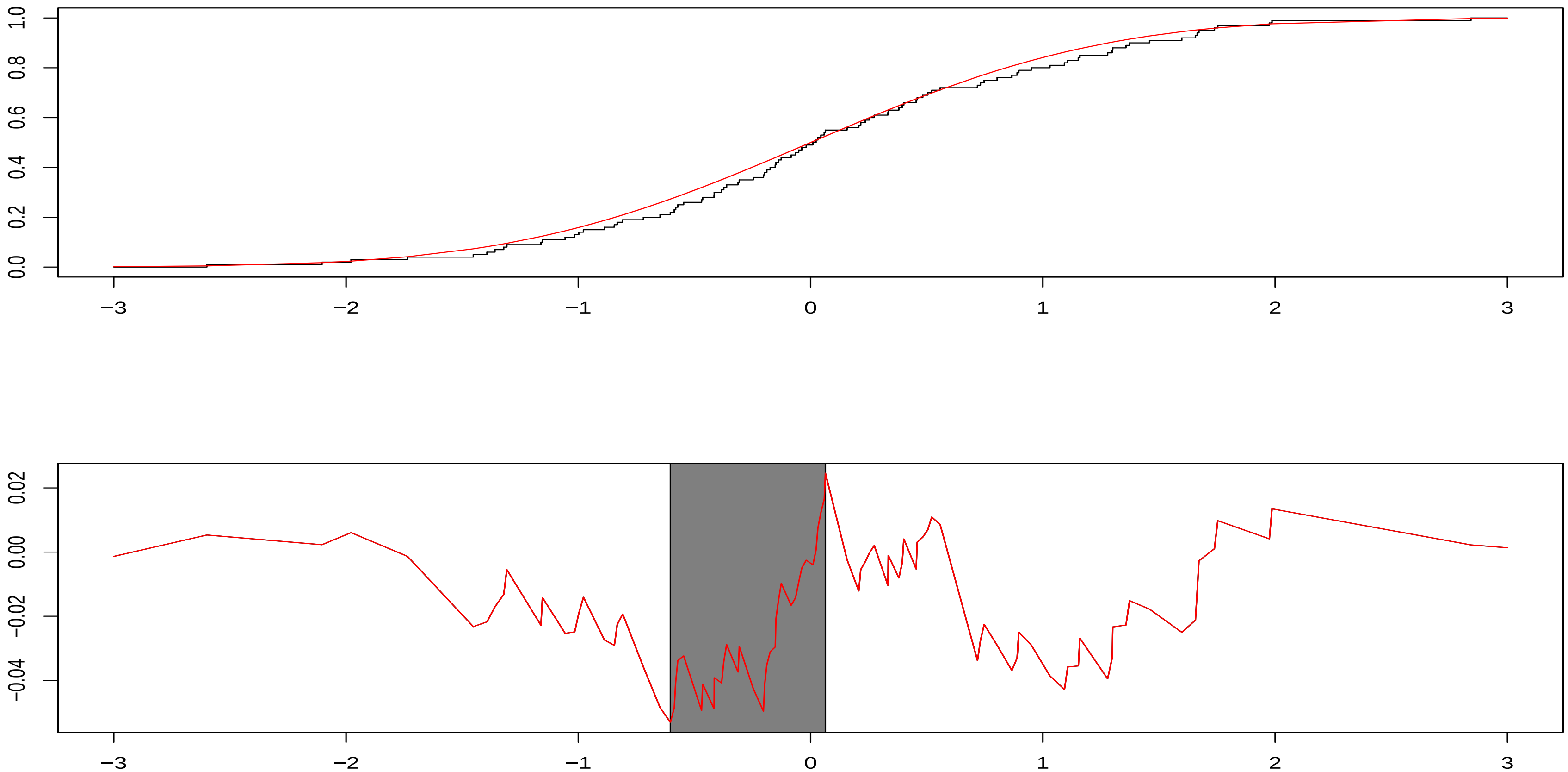


The claw density and taut strings



Taut string for densities is able to detect correct peaks.
Adequacy?

Usual Kuiper metric



Usual Kuiper metric:

$$d_{ku}(F, F_n) = \max_{a<b}\{|E_n(b) - E_n(a)|\},$$

Generalized Kuiper metrics

We define the generalized Kuiper metric d_{ku}^{κ} of order κ by

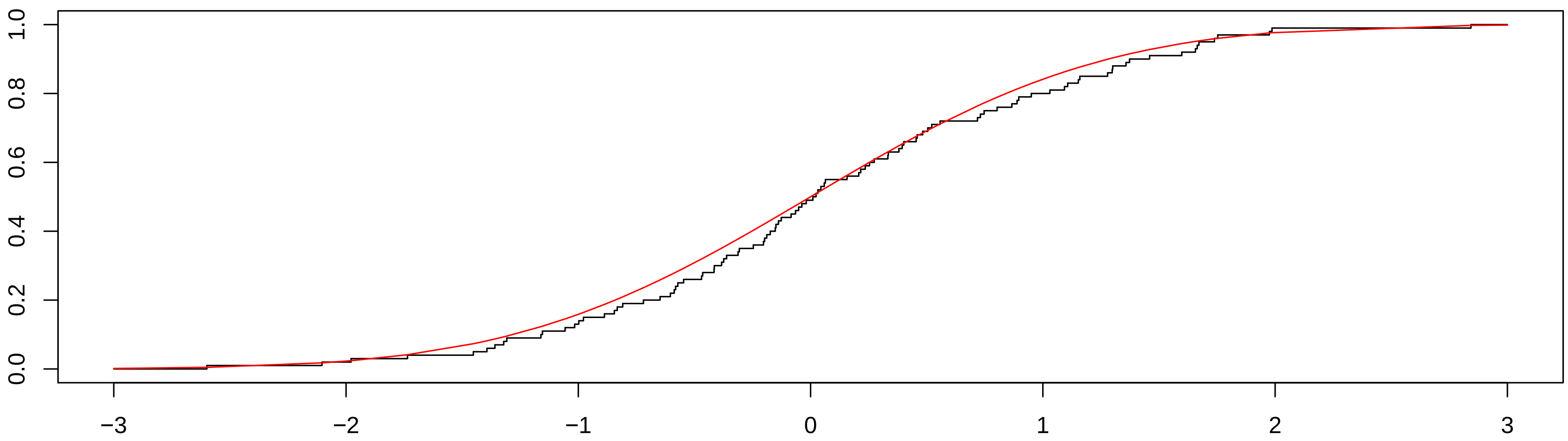
$$d_{ku}^{\kappa}(F,G) = \max\{\sum_{1}^{\kappa} |(F(b_j) - F(a_j)) - (G(b_j) - G(a_j))|\},$$

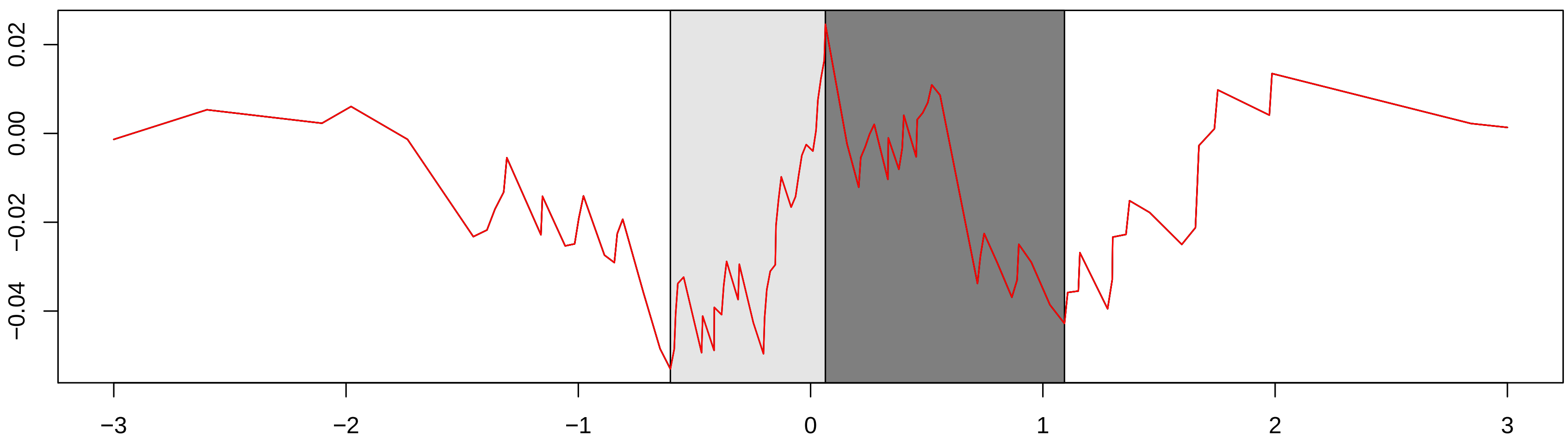
where the maximum is taken over all a_j, b_j with

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

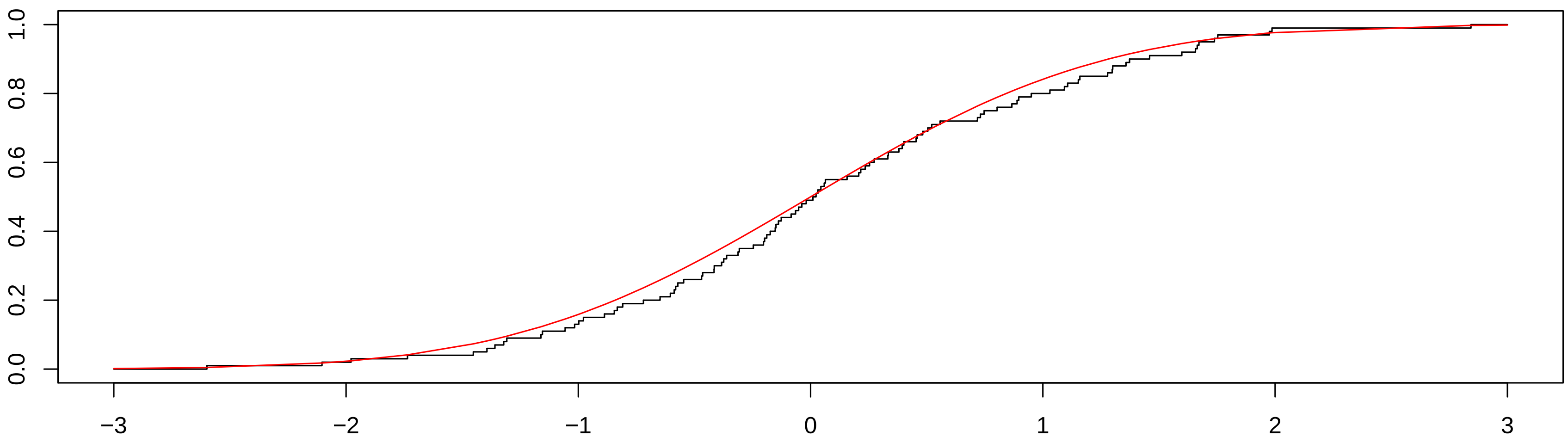
(Davies and Kovac, 2003)

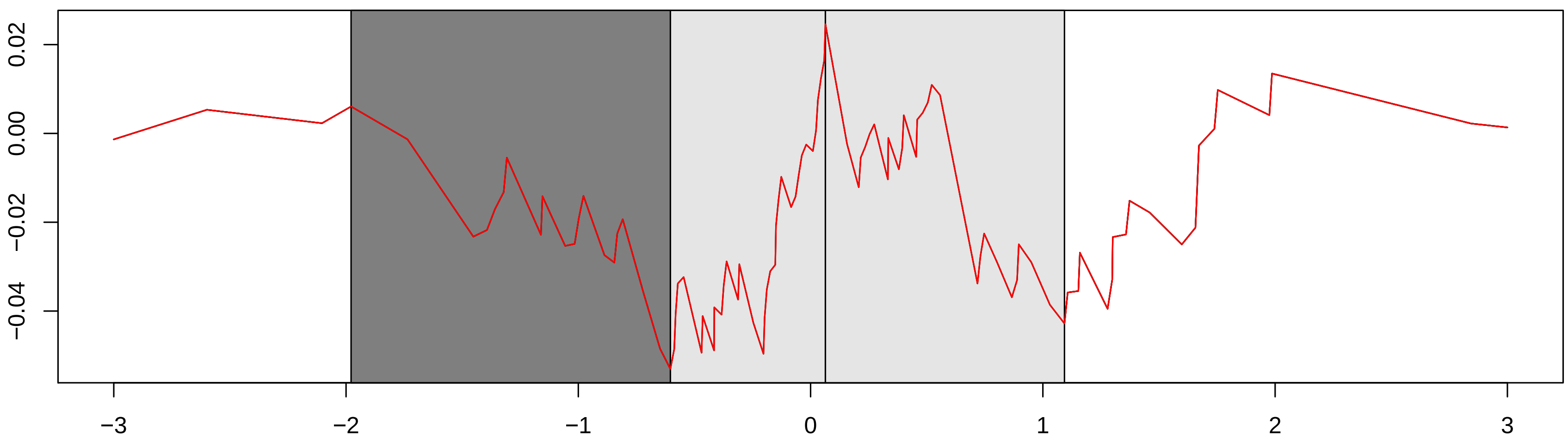
Generalized Kuiper metrics



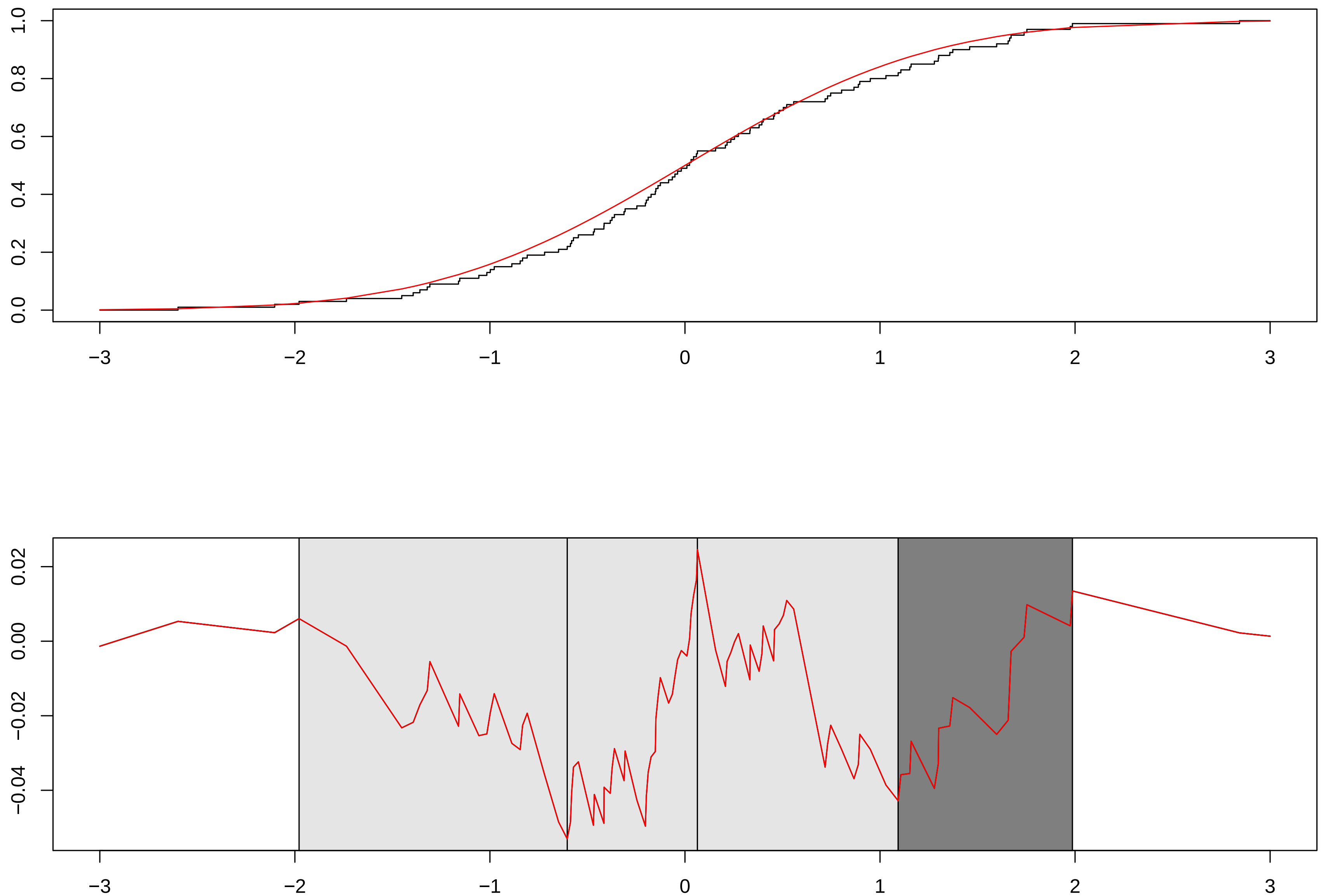


Generalized Kuiper metrics

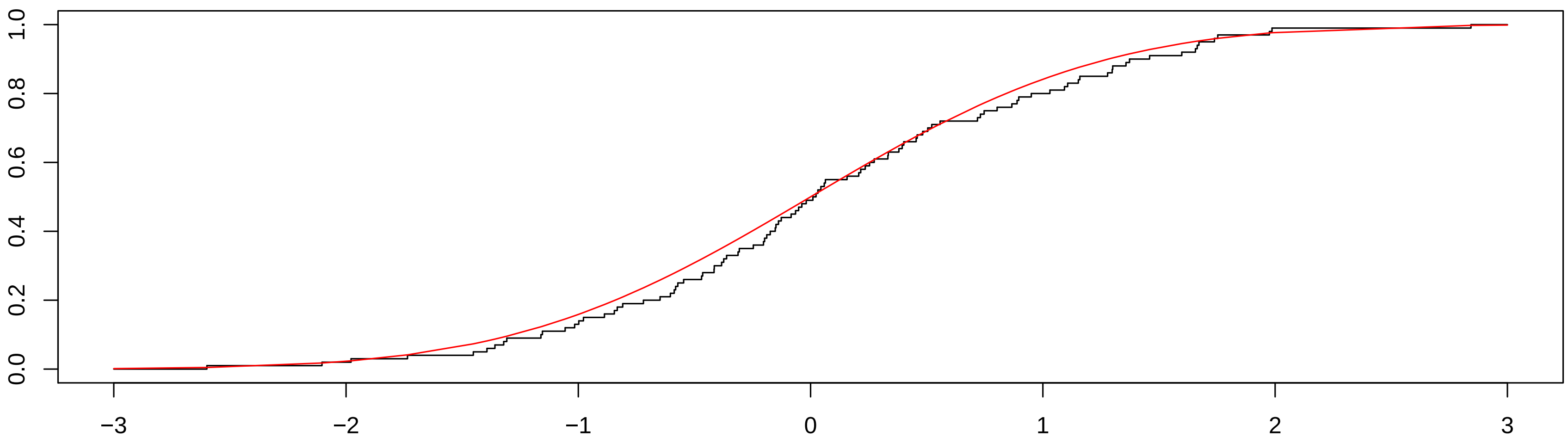


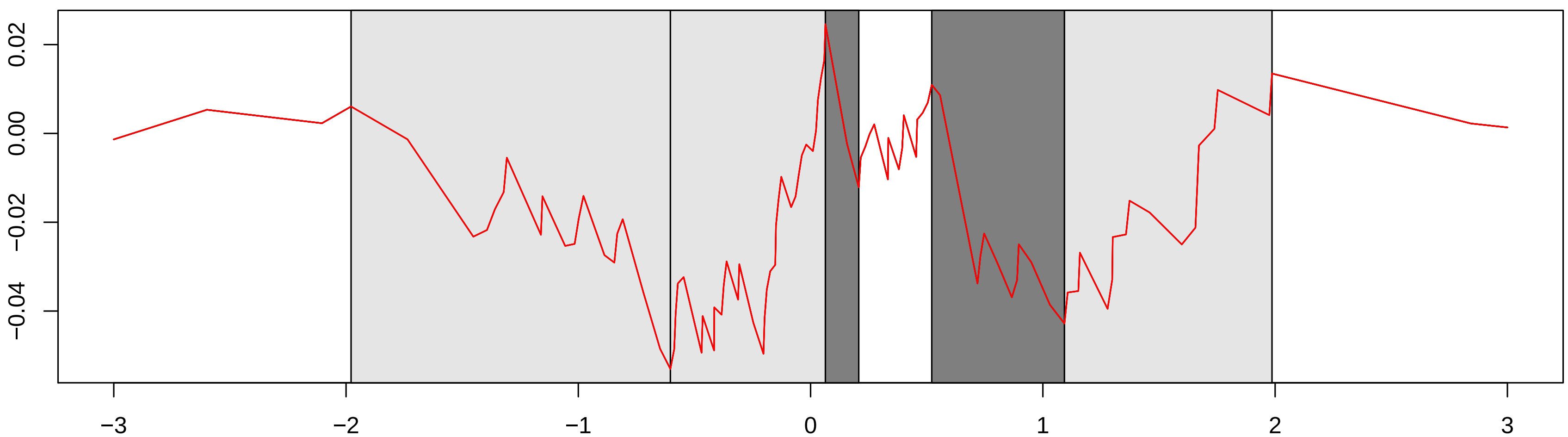


Generalized Kuiper metrics



Generalized Kuiper metrics





Adequate densities

Differences between successive Kuiper metrics of some distribution F and empirical distribution:

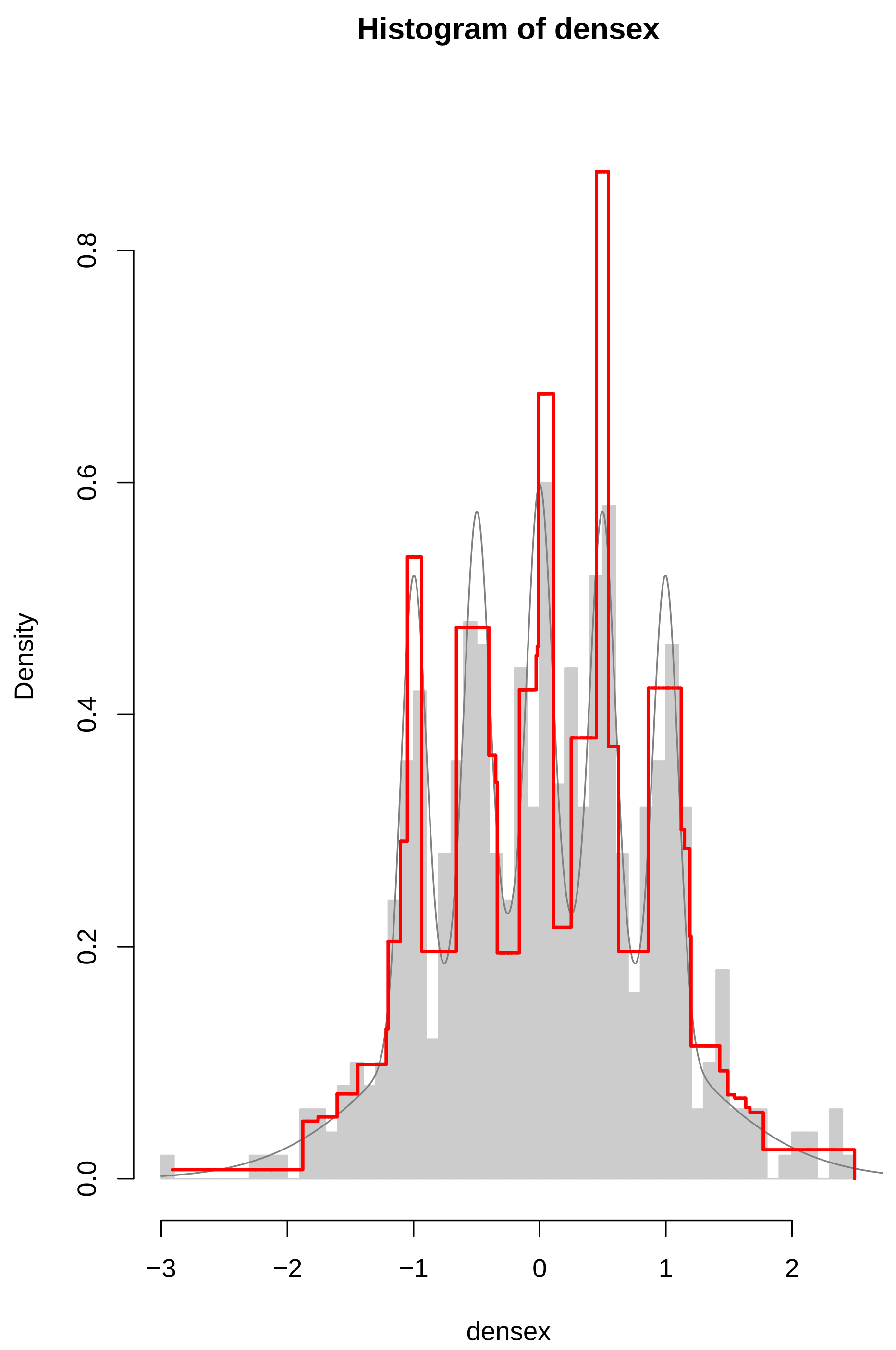
$$\rho_1(F, F_n) = d_{ku}^1(F, F_n)$$
$$\rho_2(F, F_n) = d_{ku}^2(F, F_n) - d_{ku}^1(F, F_n)$$
$$\dots$$
$$\rho_\kappa(F, F_n) = d_{ku}^\kappa(F, F_n) - d_{ku}^{\kappa-1}(F, F_n)$$

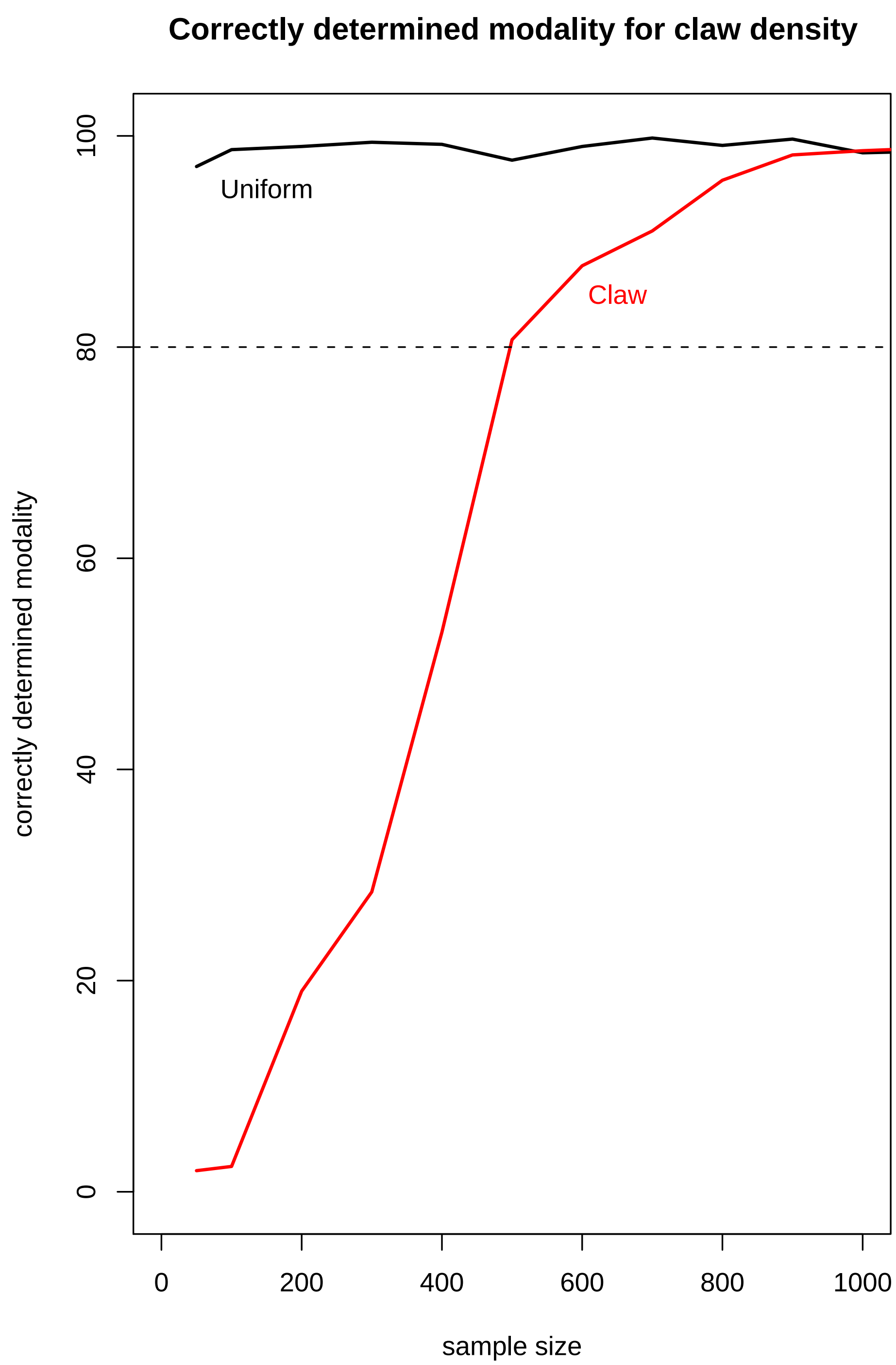
Distribution F is adequate if simultaneously for $i = 1, \dots, \kappa$

$$\rho_i(F, F_n) \leq q_i$$

where q_i is the 0.999-quantile of ρ_i.

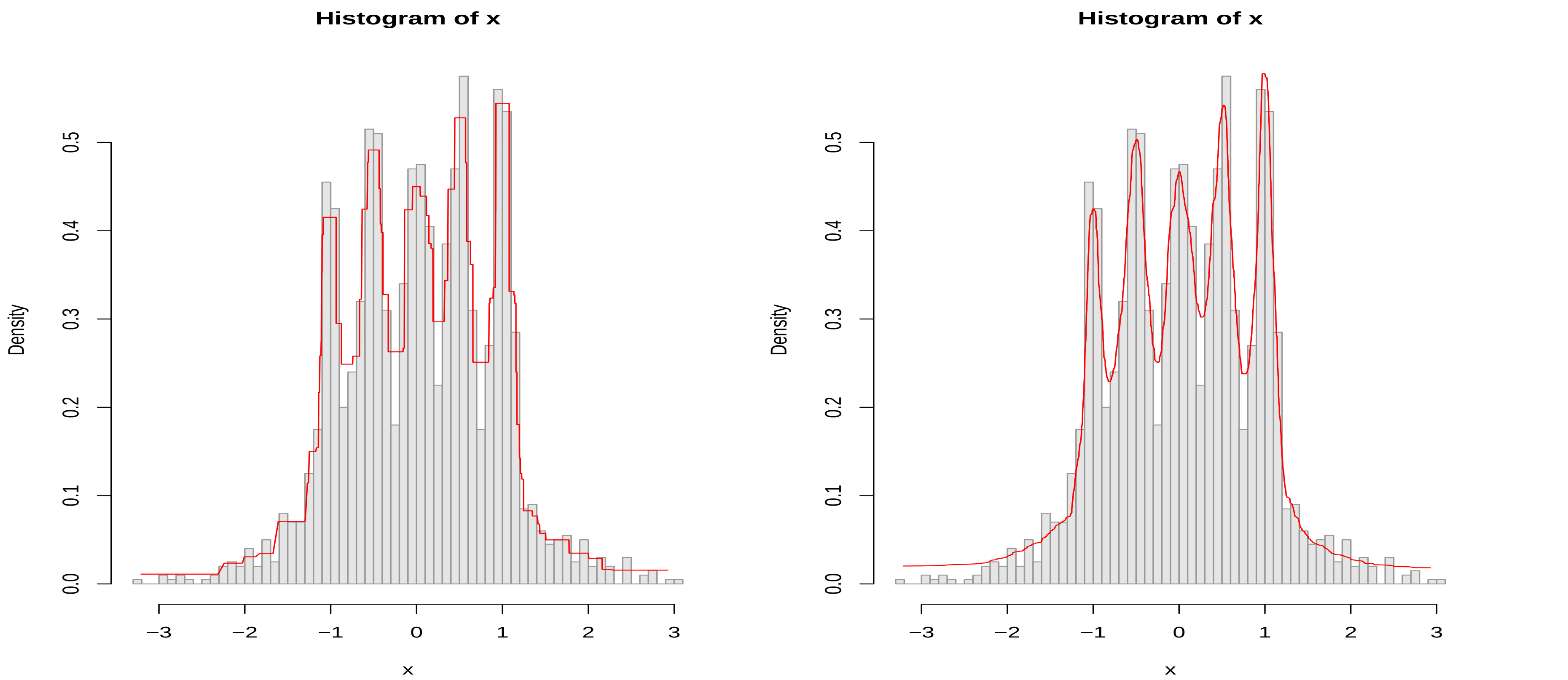
The Claw Density





Density estimation and smoothness

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



Articles and Software

Our web server:

`http://www.maths.bris.ac.uk/~maxak`

- Articles
- Software (R package: ftnonpar)