Regularization

Laurie Davies

Fakultät Mathematik, Universität Duisburg-Essen

Bristol, Friday 14th October 2011

Location-scale

Concentration of copper in a sample of drinking water (milligrams per litre)

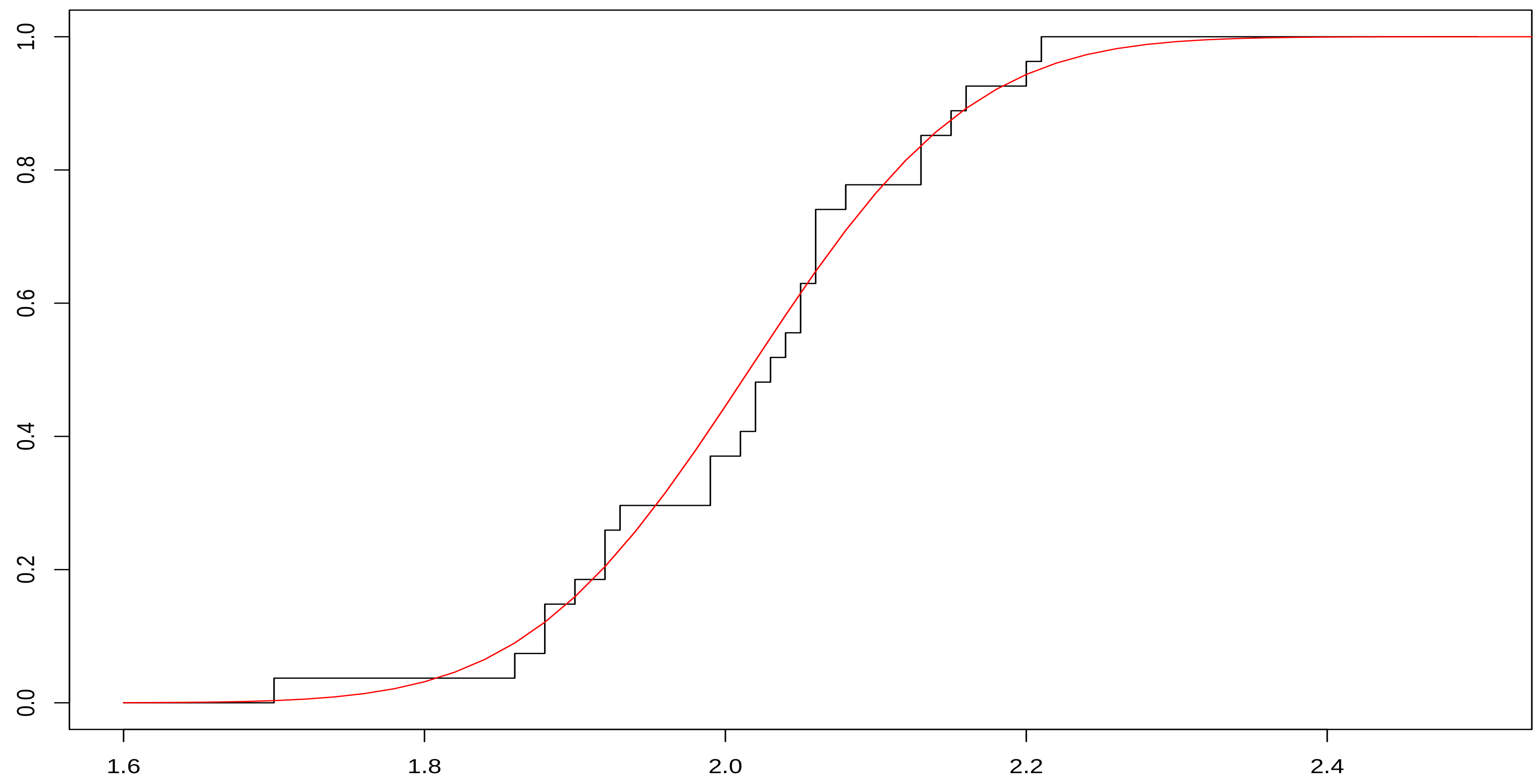
2.16	2.21	2.15	2.05	2.06	2.04	1.90	2.03	2.06
2.02	2.06	1.92	2.08	2.05	1.88	1.99	2.01	1.86
1.70	1.88	1.99	1.93	2.20	2.02	1.92	2.13	2.13

Require a point estimate and a range of plausible values.

Model the data. What is a plausible model?

Location-scale

Normal model $N(\mu, \sigma^2)$.



Location-scale

Identify μ with the concentration of copper.

Behave as if the model were true.

How to estimate μ?

Use the mean as this is the most efficient estimator.

For any α the confidence region of size α is the shortest.

Maximum likelihood $\hat{\mu} = 2.016, \hat{\sigma} = 0.114$. 0.95 confidence interval

$$[1.970, 2.062]$$

The result is certainly plausible, but can we do better?

Location-scale

The rationale for the mean in the Gaussian model is that it is the most efficient estimator.

That is it gives the shortest confidence interval for any α.

General location-scale model $F((\cdot - \mu)/\sigma)$, Gauss $F = \Phi$.

Can we choose F to make the confidence interval shorter?

Location-scale

Kuiper distances, log-likelihoods and 95% confidence intervals with their lengths for four different models for the copper data

Model	Kuiper	log–like.	95%–conf. int.	length
Gauss	0.171	20.31	$[1.970, 2.062]$	0.092
t3	0.153	19.66	$[1.983, 2.067]$	0.084
Laplace	0.163	20.09	$[1.989, 2.071]$	0.082
Comb, k=100	0.161	23.11	$[1.984, 2.036]$	0.052

Location-scale

For $k \in \mathbb{N}$ define the numbers $\iota_k(j)$ by

$$\iota_k(j) = \begin{cases} -4 + 2j/k & j = 0, \ldots, 2k, \\ (2(j-2k)+1)/k & j = 2k+1, \ldots, 4k \end{cases},$$

put

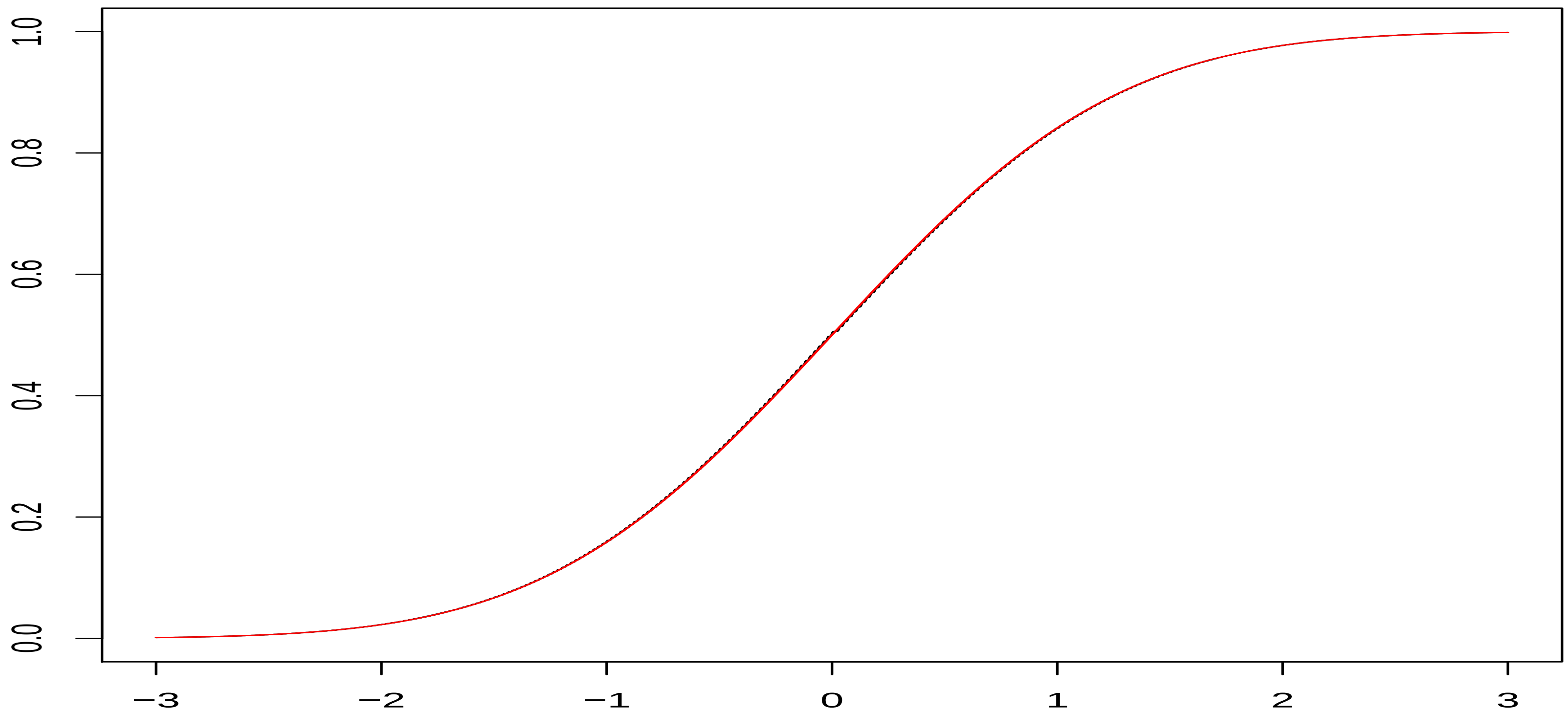
$$w(j) = \frac{\varphi(\iota(j))}{\sum_{i=0}^{4k} \varphi(\iota(i))}, \; j = 0, \ldots, 4k,$$

and finally define the Gaussian comb of order k by

$$F_{\mathsf{comb},k}(x) = \sum_{j=0}^{4k} w(j)\Phi(3k(x - \iota_k(j)))\,.$$

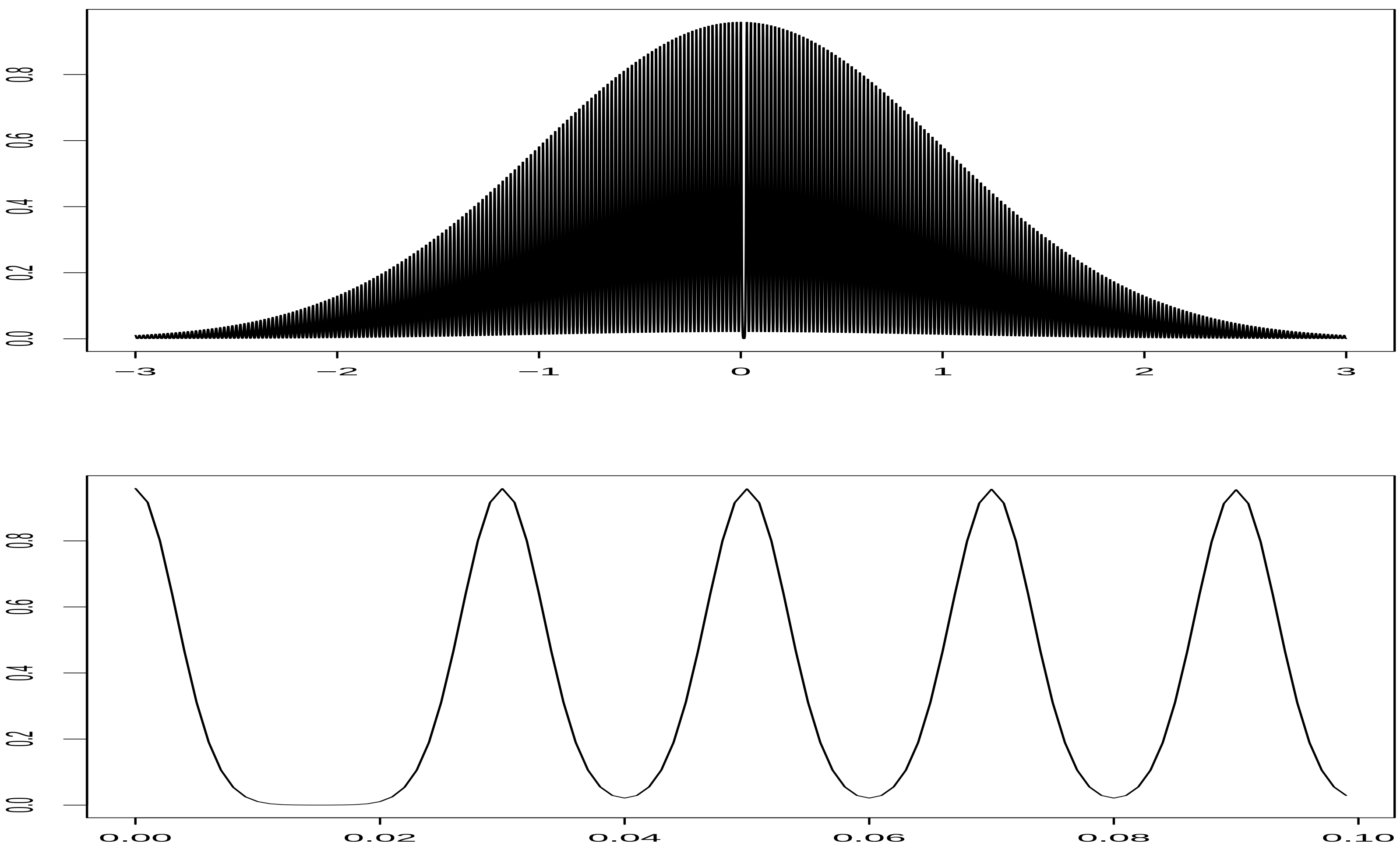
Location-scale

Distribution function of Comb_{100} distribution in black with $N(0, 1)$ superimposed.



Location-scale

Density function of Comb_{100} distribution.



Location-scale

Can one do even better?

Yes, in principle one can drive the length of the confidence interval down to zero.

Can one do even worse?

Not really, the normal model gives essentially the longest optimal confidence interval.

The normal distribution minimizes the Fisher information amongst all distributions with a given variance.

Location-scale

Fisher information

$$\mathcal{I}(F) = \int \frac{f^{(1)}(x)^2}{f(x)}\,dx$$

if F has Lebesgue density f and ∞ otherwise.
In the location model $F(\cdot - \mu)$ any unbiased estimator of μ

$$\boldsymbol{E}(T(\boldsymbol{X}_n(\mu))) = \mu$$

has variance

$$\boldsymbol{V}(T(\boldsymbol{X}_n(\mu))) \geq \frac{1}{n\mathcal{I}(F)}$$

Asymptotically the maximum likelihood estimator satisfies

$$\lim_{n\to\infty} n\boldsymbol{V}(T_{ml}(\boldsymbol{X}_n(\mu))) = \frac{1}{\mathcal{I}(F)}$$

Location-scale

Minimize $\mathcal{I}(F)$ over an ε contamination neighbourhood of the $\mathfrak{N}(0,1)$ distribution

$$\mathcal{P}(\mathfrak{N}(0,1),\varepsilon) = \{P : P = (1-\varepsilon)\mathfrak{N}(0,1) + \varepsilon Q, Q \in \mathcal{P}(\mathbb{R})\}\,.$$

Huber distributions with densities of the form

$$f_0(x) = \begin{cases} \frac{1-\varepsilon}{\sqrt{2\pi}} \exp\left(-x^2/2\right) & |x| \le k \\ \\ \frac{1-\varepsilon}{\sqrt{2\pi}} \exp\left(k^2/2 - k|x|\right) & |x| > k \end{cases}$$

where

$$\frac{2\varphi(k)}{k} - 2\Phi(k) = \frac{\varepsilon}{1-\varepsilon}\,.$$

Location-scale

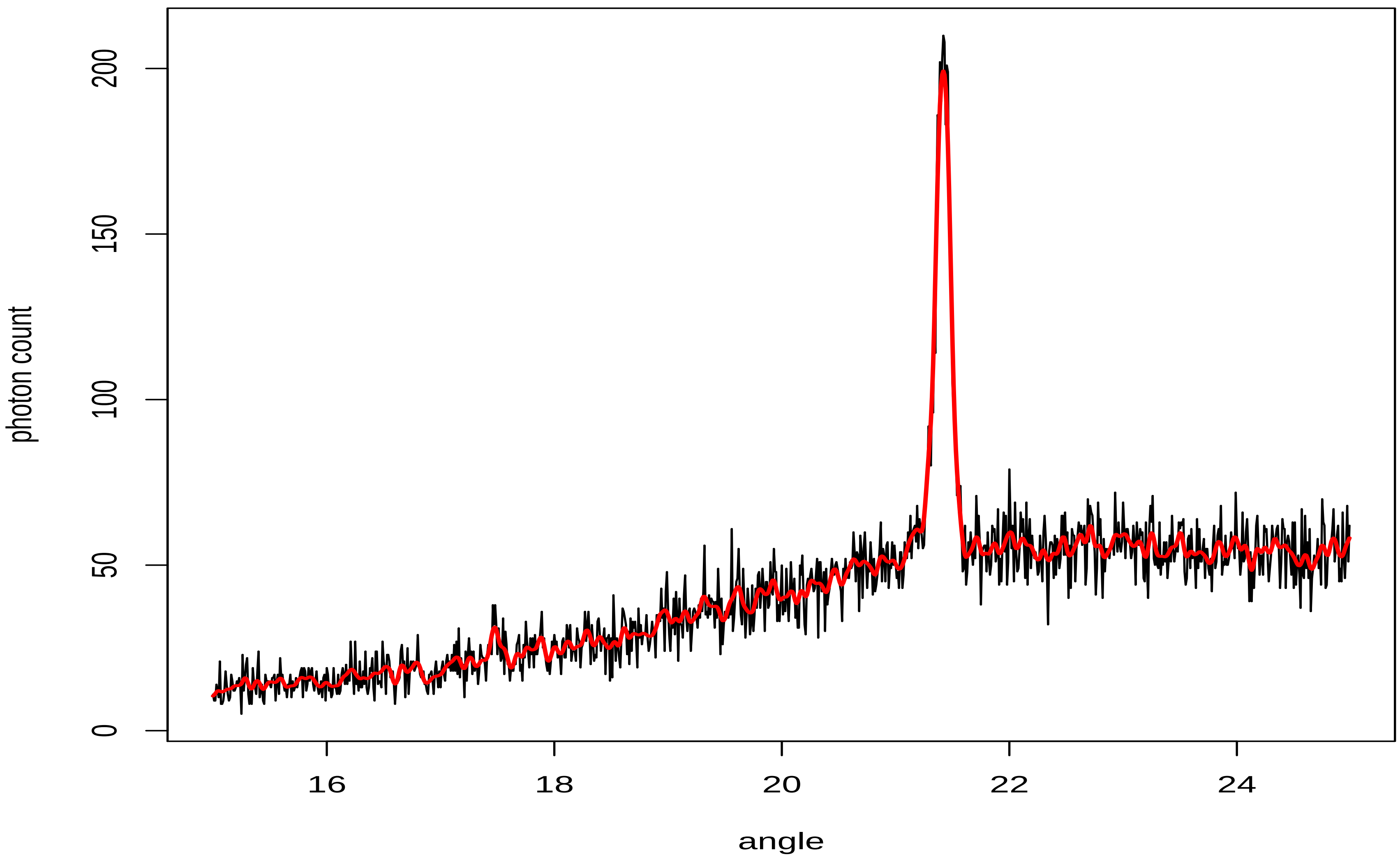
The location-scale problem is ill-posed and must be regularized.

Tukey calls such distributions ‘bland’ or ‘hornless’.

Moral: ‘efficiency’ can be imported from the model at no cost.

TINSTAAFL

Non-parametric regression



Non-parametric regression

Approximation region

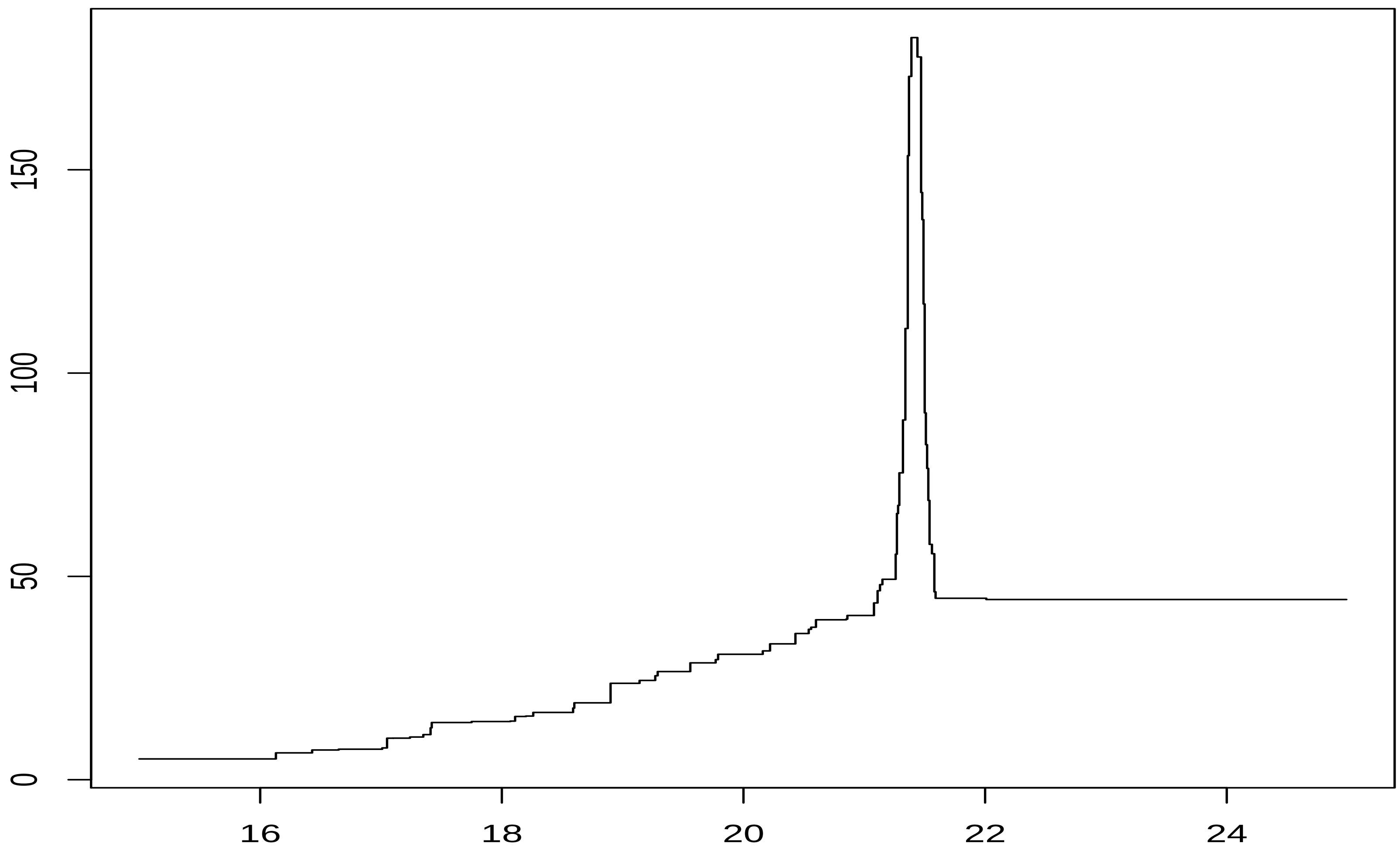
$$\mathcal{A}(\boldsymbol{x}_n, \alpha, \mathcal{P}) = \left\{ g : \max_I |w_n(\boldsymbol{x}_n, g, I)| \leq \sigma \sqrt{\tau_n(\alpha) \log n} \right\}$$

Regularize

Minimize the number of local extreme values of g subject to $g \in \mathcal{A}(\boldsymbol{x}_n, \alpha, \mathcal{P})$

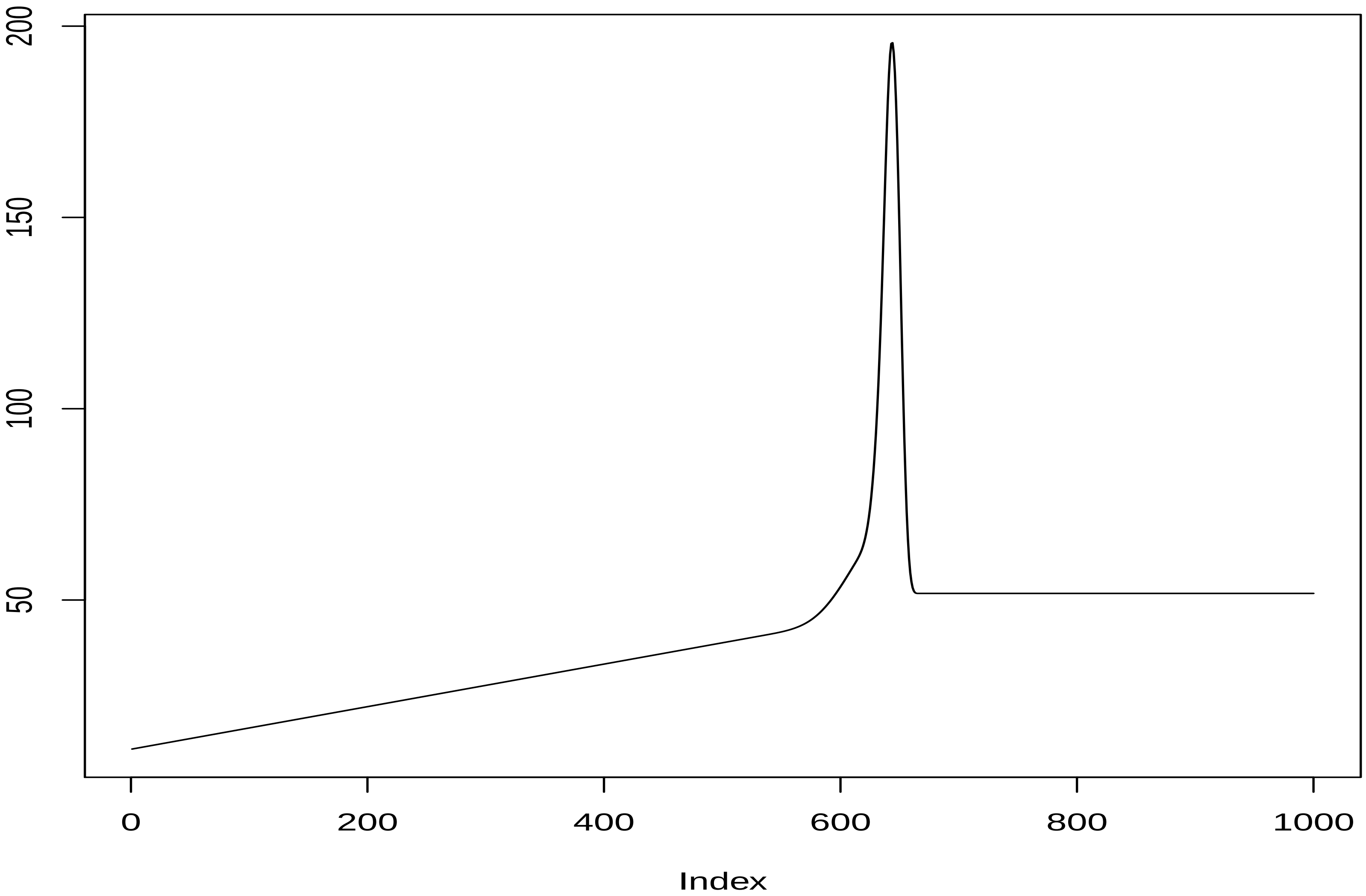
Davies and Kovac (2001) - taut string

Non-parametric regression



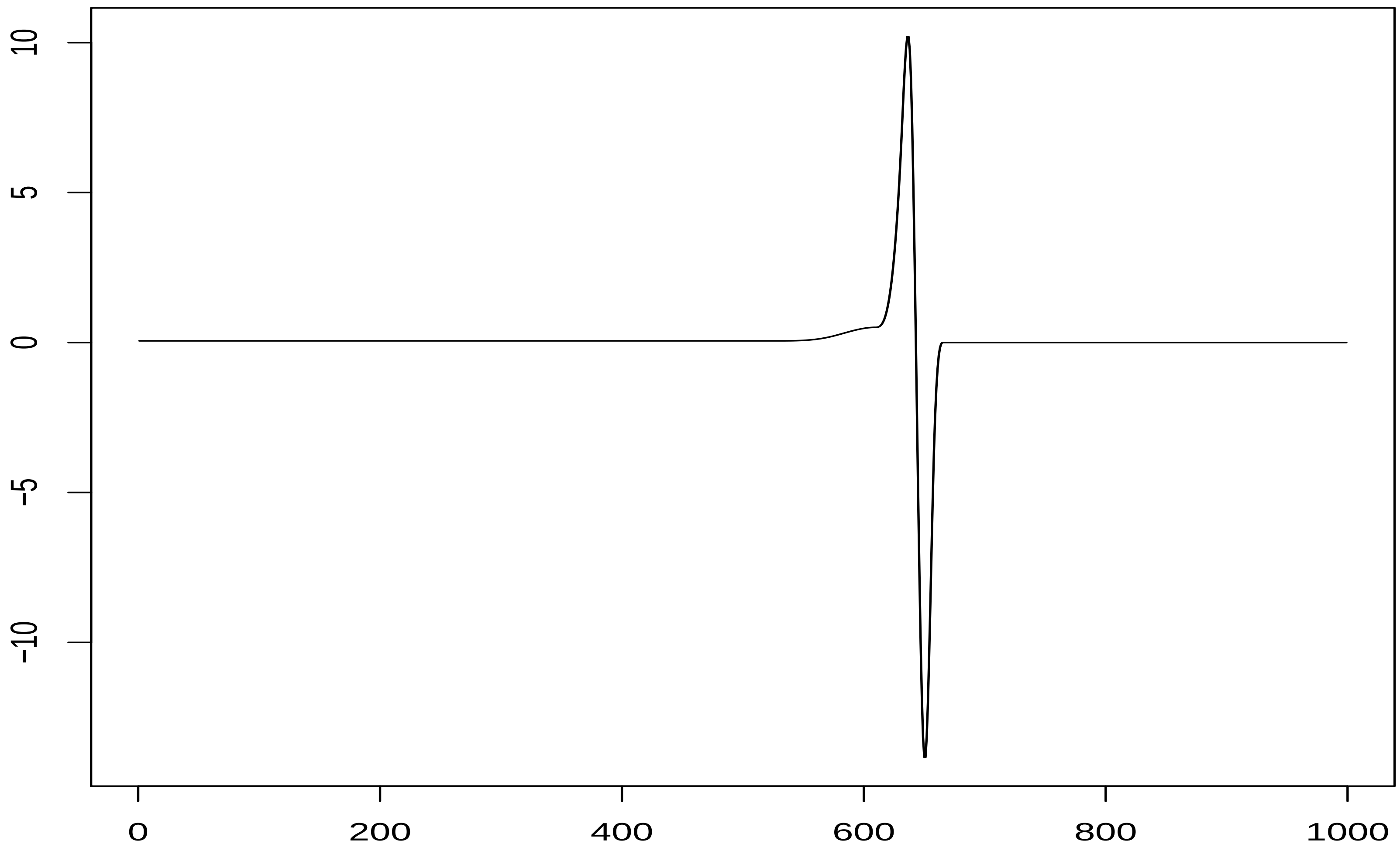
Non-parametric regression

Minimze total variation of fourth derivative.



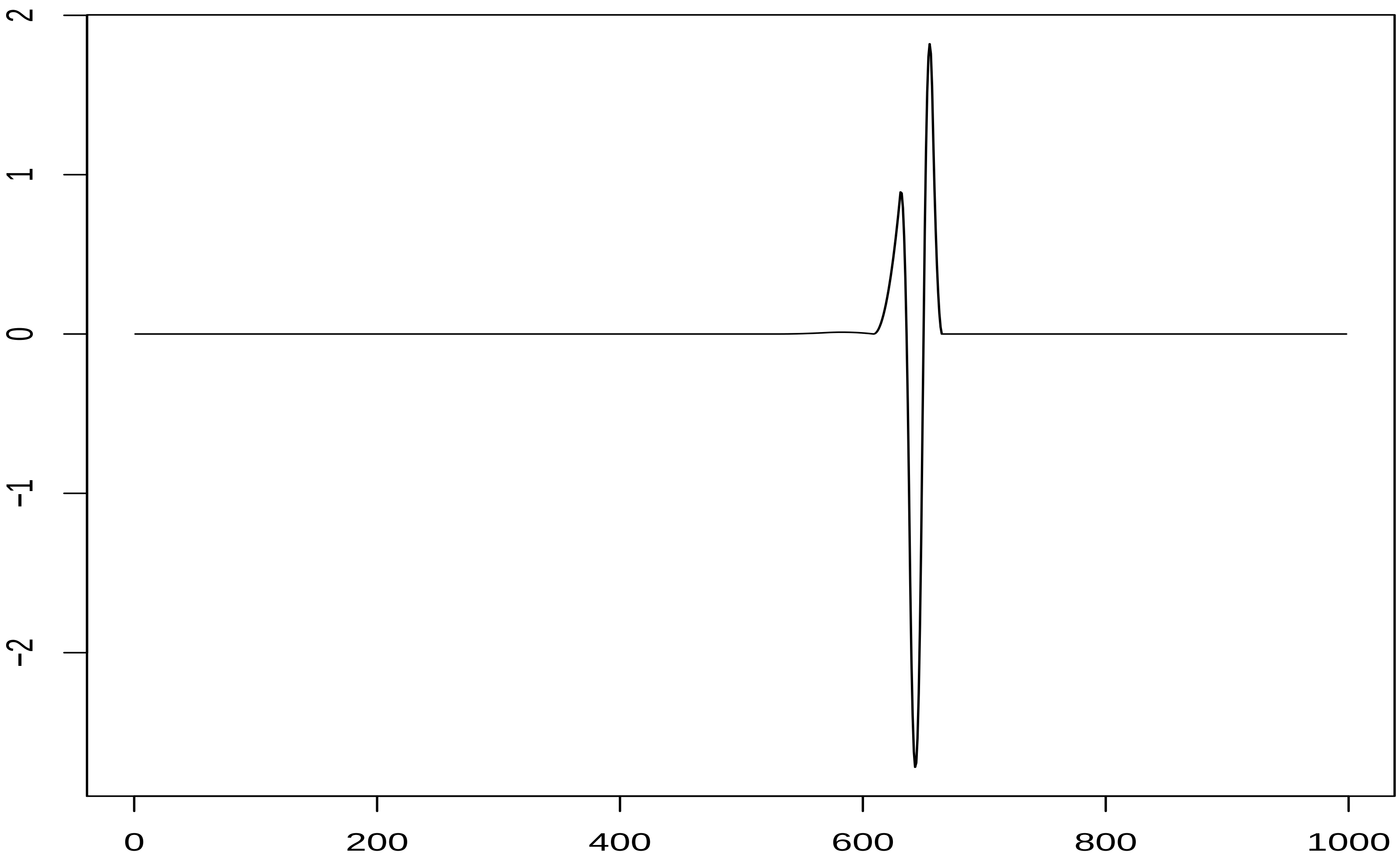
Non-parametric regression

First derivative



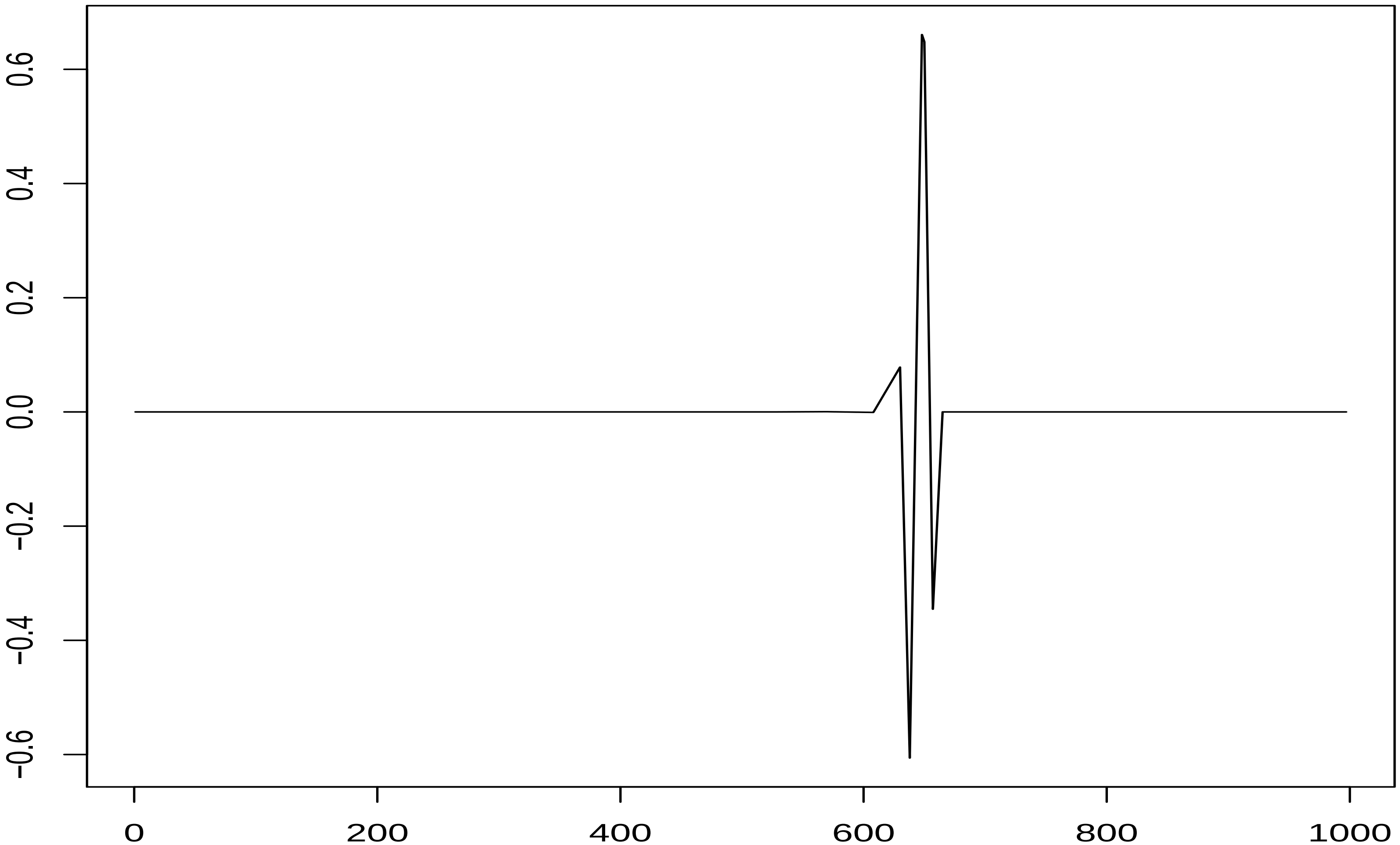
Non-parametric regression

Second derivative



Non-parametric regression

Third derivative



Non-parametric regression

Fourth derivative

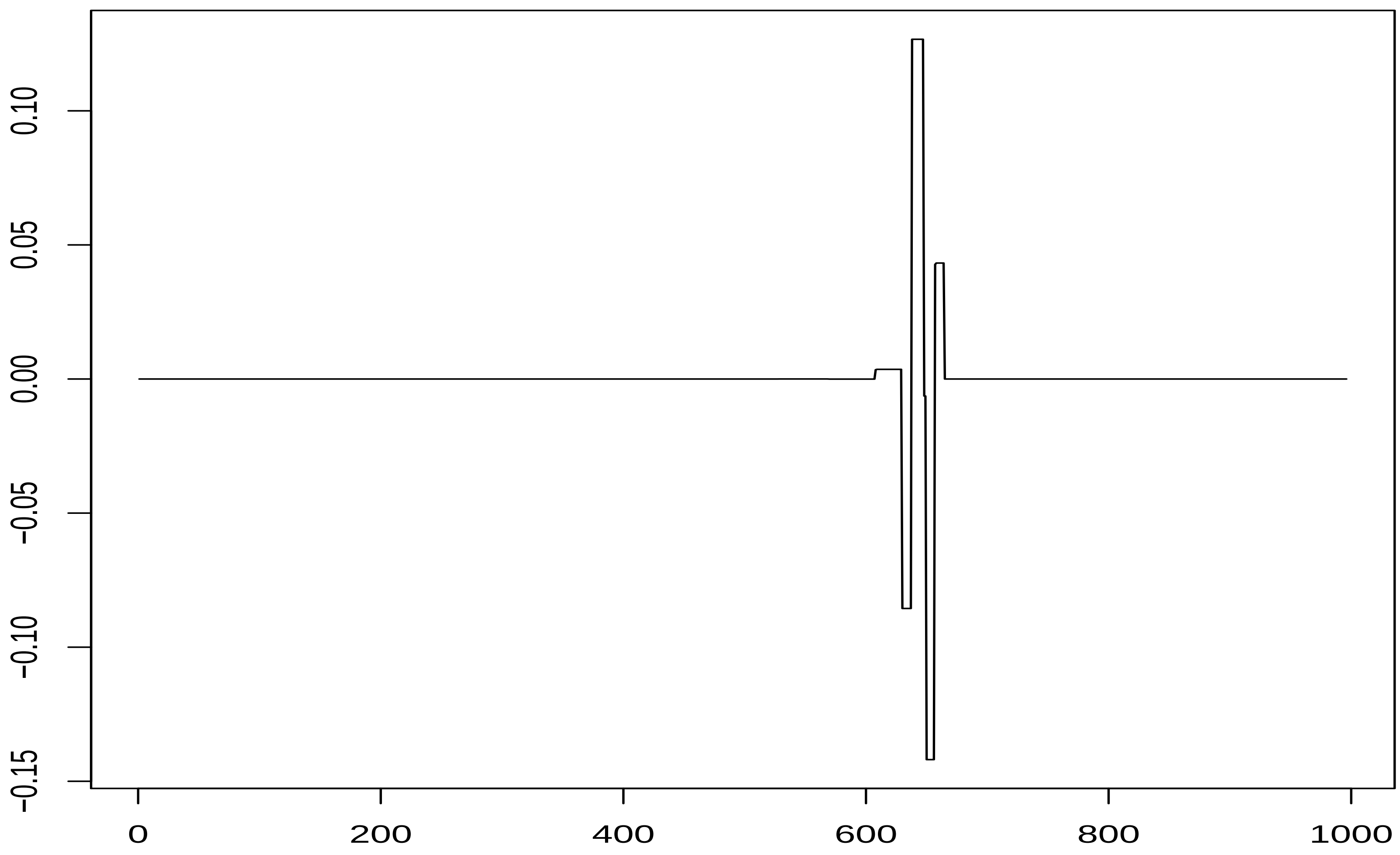


Image analysis

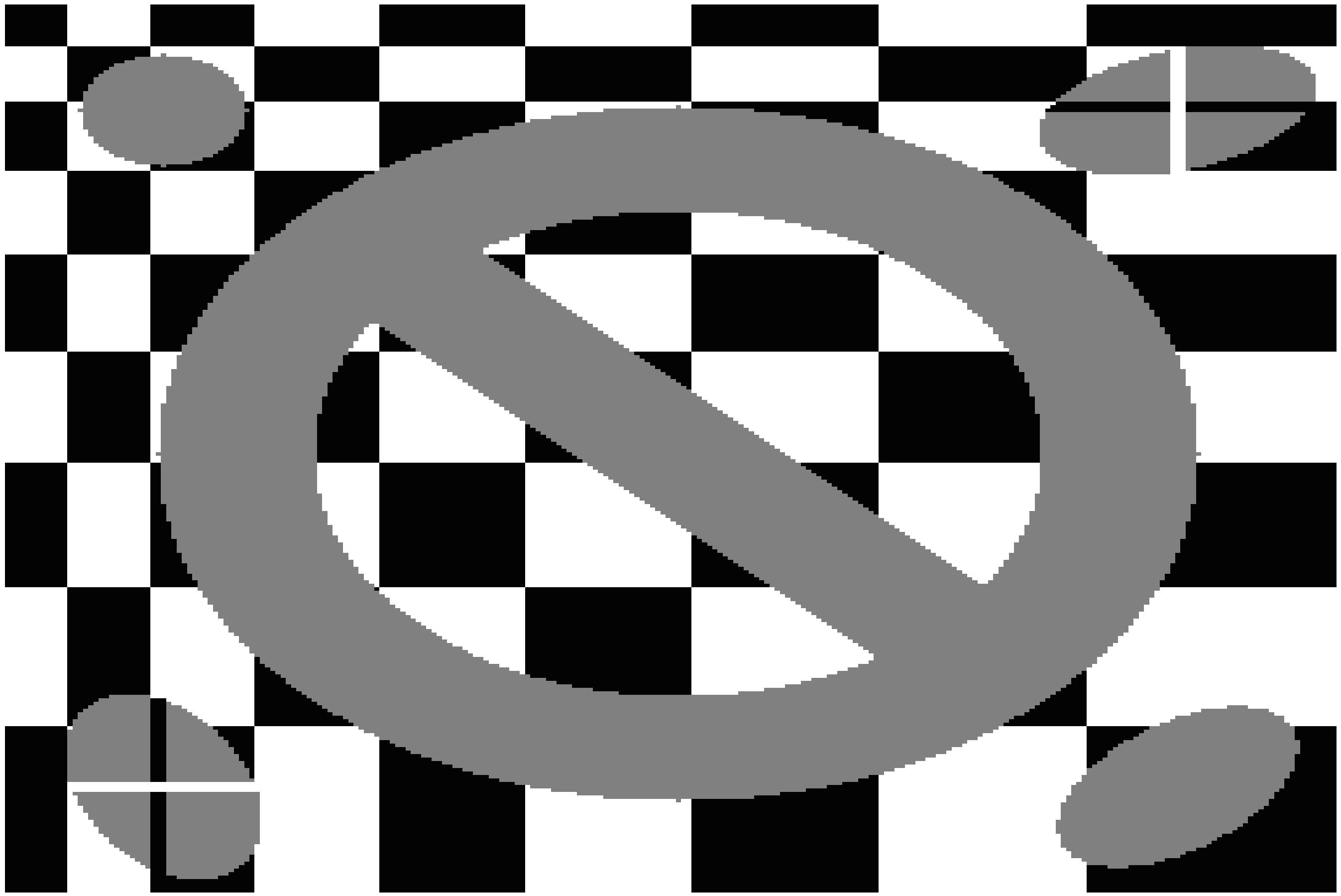


Image analysis

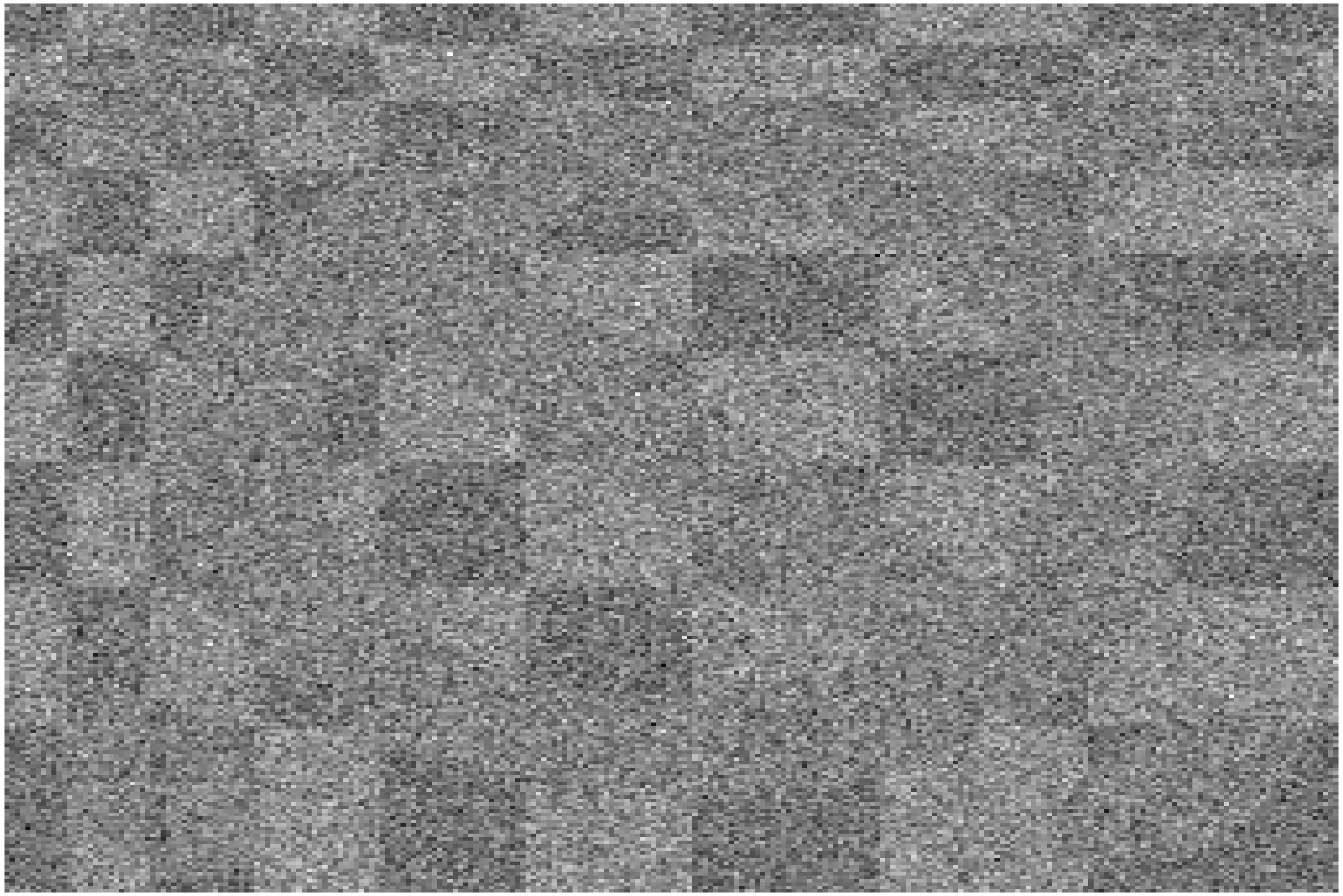


Image analysis