Statistical signal processing for the detection of small/rare features with applications in astrophysics

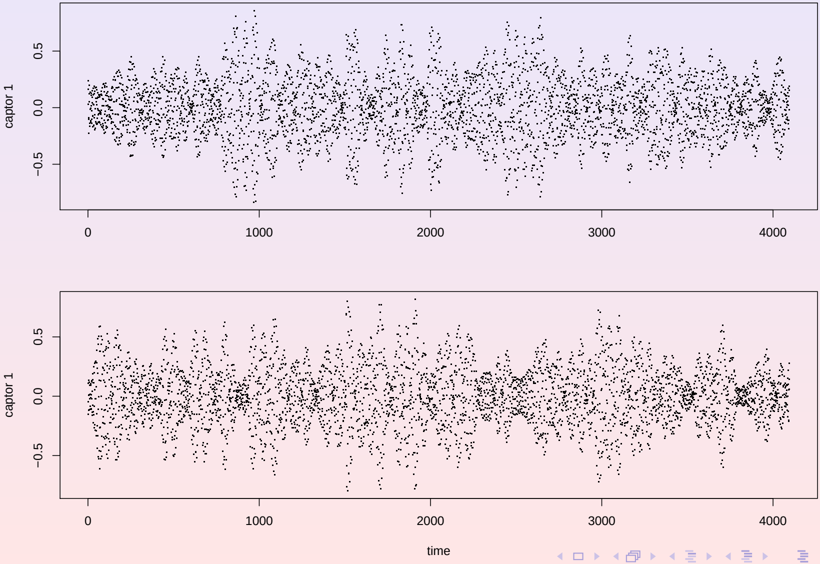
Sylvain Sardy

Mathematics department, University of Geneva

CERN, PhyStat
9 January 2011

Joint work with Caroline Giacobino, Stefano Foffa, Roberto Terenzi

Gravitational wave detection



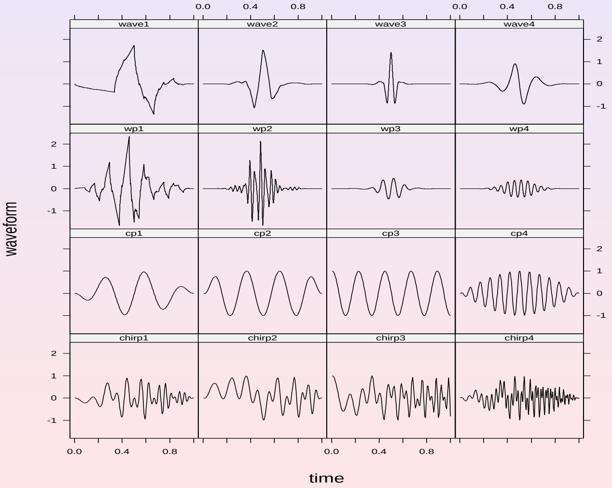
Time series $\{(S_{1,t}, S_{2,t})\}_{t=1,\dots,T}$ for gravitational wave detection:

- 2 detectors: concomitant ($\neq$ repeated) measurements
- poor signal f to noise ϵ ratio
- 5 kHz: 600'000 observations per minute
- colored electronic noise
- Astrophysics predicts rare arrivals of small wavebursts:

$$f(t) = 0 \quad \text{for most } t.$$

Wavelet representation:

$$f(t) = \sum_i \gamma_i \phi_i(t) + \sum_j^J \sum_{k=1}^{n_j} \alpha_{j,k} \psi_{j,k}(t).$$



Sparse wavelet representation:

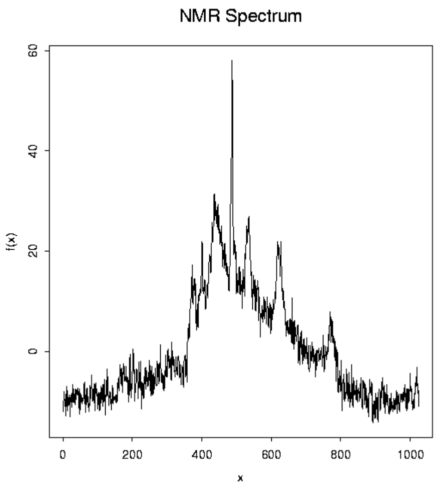
With only a few **non-zero** *wavelet coefficients* $\alpha_{j,k}$, *we can approximate many signals.*

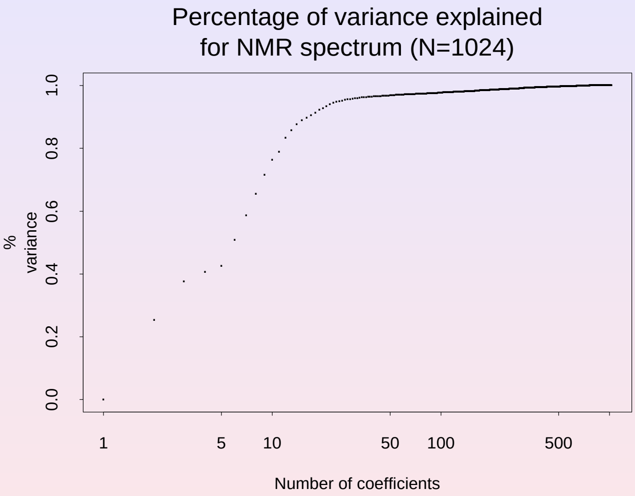
Illustration with an Nuclear Magnetic Resonance noisy signal.

Sparse wavelet representation:

With only a few **non-zero** *wavelet coefficients* $\alpha_{j,k}$*,*
we can approximate many signals.

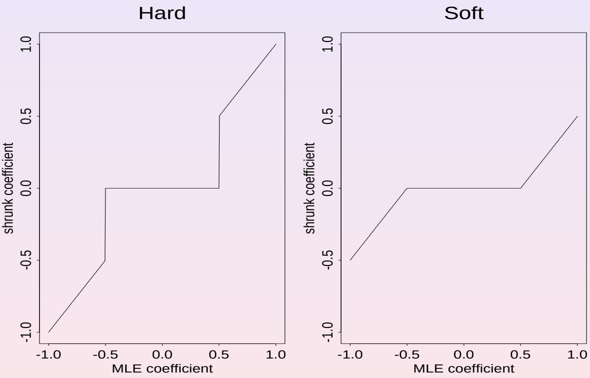
Illustration with an Nuclear Magnetic Resonance noisy signal.


NMR Spectrum
f(x)
60
40
20
0
0
200
400
600
800
1000
x


Percentage of variance explained
for NMR spectrum (N=1024)
% variance
0.0
0.2
0.4
0.6
0.8
1.0
1
5
10
50
100
500
Number of coefficients

How to enforce sparsity?

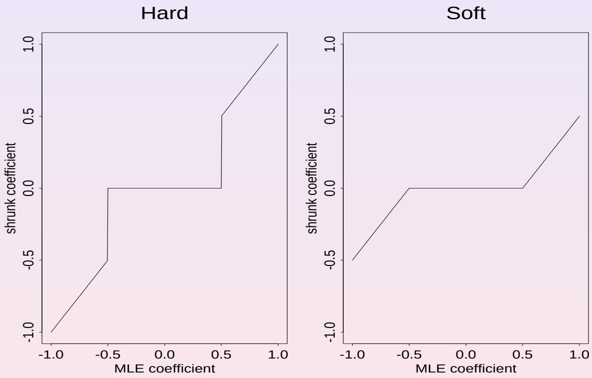
With a non-linear thresholding function with threshold λ:



Anything smaller than λ is treated as noise, and killed.

How to enforce sparsity?

With a non-linear thresholding function with threshold λ:



Anything smaller than λ is treated as noise, and killed.

1 captor: Donoho and Johnstone's 3 easy steps:

1. Analyze

$$\hat{\alpha}^{\mathrm{MLE}} =: \mathbf{Y} = \mathrm{DWT}(\mathbf{S}) \quad (\text{ like FFT})$$

2. Threshold

$$\begin{aligned} \hat{\alpha}_{j,k} &= \eta_{\lambda_j}^{(\mathrm{soft})}(Y_{j,k}) \\ &= \mathrm{sign}(Y_{j,k})(|Y_{j,k}| - \lambda_j)_+ \\ &= (1 - \frac{\lambda_j}{|Y_{j,k}|})_+ Y_{j,k} \end{aligned}$$

for all levels $j = 1, \ldots, J$.

3. Reconstruct $\rightsquigarrow \hat{f}_\lambda$

1 captor: Donoho and Johnstone's 3 easy steps:

1. Analyze

$$\hat{\alpha}^{\mathrm{MLE}} =: \mathbf{Y} = \mathrm{DWT}(\mathbf{S}) \quad (\text{ like FFT})$$

2. Threshold

$$\begin{aligned} \hat{\alpha}_{j,k} &= \eta^{(\mathrm{soft})}_{\lambda_j}(Y_{j,k}) \\ &= \mathrm{sign}(Y_{j,k})(|Y_{j,k}| - \lambda_j)_+ \\ &= (1 - \frac{\lambda_j}{|Y_{j,k}|})_+ Y_{j,k} \end{aligned}$$

for all levels $j = 1, \ldots, J$.

3. Reconstruct $\rightsquigarrow \hat{f}_\lambda$

1 captor: Donoho and Johnstone's 3 easy steps:

1. Analyze

$$\hat{\alpha}^{\mathrm{MLE}} =: \mathbf{Y} = \mathrm{DWT}(\mathbf{S}) \quad \text{(like FFT)}$$

2. Threshold

$$\begin{aligned}\hat{\alpha}_{j,k} &= \eta^{(\mathrm{soft})}_{\lambda_j}(Y_{j,k}) \\ &= \mathrm{sign}(Y_{j,k})(|Y_{j,k}| - \lambda_j)_+ \\ &= (1 - \frac{\lambda_j}{|Y_{j,k}|})_+ Y_{j,k}\end{aligned}$$

for all levels $j = 1, \ldots, J$.

3. Reconstruct $\rightsquigarrow \hat{f}_\lambda$

2 captors: enforce concomitant/block sparsity with

$$\begin{pmatrix} \hat{\alpha}_{j,k}^{(1)} \\ \hat{\alpha}_{j,k}^{(2)} \end{pmatrix} = (1 - \frac{\lambda_j}{\|\mathbf{Y}_{j,k}\|_2})_+ \begin{pmatrix} \mathbf{Y}_{j,k}^{(1)} \\ \mathbf{Y}_{j,k}^{(2)} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \text{or not} \end{cases} .$$

Reminiscent of the truncated James-Stein estimator

$$\hat{\alpha}^{\mathrm{JS}+} = (1 - \frac{P-2}{\|\mathbf{Y}\|_2^2})_+ \mathbf{Y} \tag{1}$$

2 captors: enforce concomitant/block sparsity with

$$\begin{pmatrix} \hat{\alpha}_{j,k}^{(1)} \\ \hat{\alpha}_{j,k}^{(2)} \end{pmatrix} = (1 - \frac{\lambda_j}{\|\mathbf{Y}_{j,k}\|_2})_+ \begin{pmatrix} \mathbf{Y}_{j,k}^{(1)} \\ \mathbf{Y}_{j,k}^{(2)} \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \text{or not} \end{cases} .$$

Reminiscent of the truncated James-Stein estimator

$$\hat{\alpha}^{\mathrm{JS}+} = (1 - \frac{P-2}{\|\mathbf{Y}\|_2^2})_+ \mathbf{Y} \qquad (1)$$

$\rightsquigarrow$ Generalized James-Stein estimator

$$\hat{\alpha}_{j,k}^{\mathrm{GJS},\lambda_j,\nu_j} = \left(1 - \frac{\lambda_j^{\nu_j}}{\|\mathbf{Y}_{j,k}\|_2^{\nu_j}}\right)_+ \mathbf{Y}_{j,k}, \quad \nu_j > 0$$

is a continuous bridge between soft ($\nu_j = 1$) and hard ($\nu_j \to \infty$) block-thresholding.

Challenge: selection of λ and ν.

One possibility is to minimize the risk

$$\mathrm{R}(\lambda, \nu) = \mathrm{E}\|\hat{f}_{\lambda,\nu} - f\|_2^2$$

or an unbiased estimate.

$\rightsquigarrow$ Generalized James-Stein estimator

$$\hat{\alpha}_{j,k}^{\mathrm{GJS},\lambda_j,\nu_j} = \left(1 - \frac{\lambda_j^{\nu_j}}{\|\mathbf{Y}_{j,k}\|_2^{\nu_j}}\right)_+ \mathbf{Y}_{j,k}, \quad \nu_j > 0$$

is a continuous bridge between soft ($\nu_j = 1$) and hard ($\nu_j \to \infty$) block-thresholding.

Challenge: selection of λ and ν.

One possibility is to minimize the risk

$$\mathrm{R}(\lambda, \nu) = \mathrm{E}\|\hat{f}_{\lambda,\nu} - f\|_2^2$$

or an unbiased estimate.

⇝ Generalized James-Stein estimator

$$\hat{\alpha}_{j,k}^{\mathrm{GJS},\lambda_j,\nu_j} = (1 - \frac{\lambda_j^{\nu_j}}{\|\mathbf{Y}_{j,k}\|_2^{\nu_j}})_+ \mathbf{Y}_{j,k}, \quad \nu_j > 0$$

is a continuous bridge between soft ($\nu_j = 1$) and hard ($\nu_j \to \infty$) block-thresholding.

Challenge: selection of λ and ν.

One possibility is to minimize the risk

$$\mathrm{R}(\lambda, \nu) = \mathrm{E}\|\hat{f}_{\lambda,\nu} - f\|_2^2$$

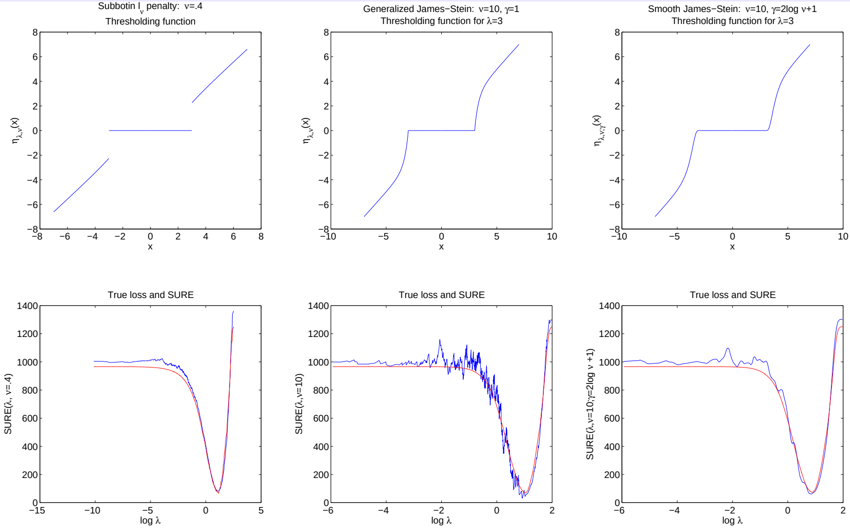
or an unbiased estimate.

Selection of λ_j and ν_j with Stein Unbiased Risk Estimate SURE

The GJS thresholding function is weakly differentiable in Stein's sense, so we can estimate unbiasedly the risk with

$$\begin{aligned}\hat{\mathrm{R}}(\lambda_j,\nu_j) &= \sum_{k=1}^{2^j} \|\mathbf{Y}_{j,k}\|_2^2 1(\|\mathbf{Y}_{j,k}\|_2 \le \lambda_j) + \sum_{k=1}^{2^j} \frac{\lambda_j^{2\nu_j}\|\mathbf{Y}_{j,k}\|_2^2}{\|\mathbf{Y}_{j,k}\|_2^{2\nu_j}} 1(\|\mathbf{Y}_{j,k}\|_2 > \lambda_j) \\ &\quad +2\sum_{k=1}^{2^j}\left(Q - \frac{\lambda_j^{\nu_j}(Q-\nu_j)}{\|\mathbf{Y}_{j,k}\|_2^{\nu_j}}\right) 1(\|\mathbf{Y}_{j,k}\|_2 > \lambda_j) - 2^j Q. \\ &= \mathrm{RSS}(\lambda_j,\nu_j) + 2\,\mathrm{d.f.}_j\end{aligned}$$

and minimize it over λ_j and ν_j.


Subbotin l_ν penalty: ν=.4
Thresholding function
$\eta_{\lambda,\nu}(x)$
x
Generalized James−Stein: ν=10, γ=1
Thresholding function for λ=3
$\eta_{\lambda,\nu}(x)$
x
Smooth James−Stein: ν=10, γ=2log ν+1
Thresholding function for λ=3
$\eta_{\lambda,\nu,\gamma}(x)$
x
True loss and SURE
SURE(λ, ν=.4)
log λ
True loss and SURE
SURE(λ,ν=10)
log λ
True loss and SURE
SURE(λ,ν=10;γ=2log ν +1)
log λ

Selection of λ_j by universal rule for Q captors

$$\lambda_{j,N,Q} = \sqrt{2(\log N_j + (Q/2)\log\log N_j - \log\Gamma(Q/2))},$$

e.g., with 2 captors

$$\lambda_{j,N,2} = \sqrt{2(\log N_j + \log\log N_j)}.$$

Oracle inequality:

$$\mathrm{R}(\hat{\alpha}^{\mathrm{SJS}}, \alpha) \leq (Q + \lambda^2_{N,Q})(Q + \mathrm{R}^{\mathrm{Oracle}}(\omega, \alpha)).$$

Selection of λ_j by universal rule for Q captors

$$\lambda_{j,N,Q} = \sqrt{2(\log N_j + (Q/2)\log\log N_j - \log\Gamma(Q/2))},$$

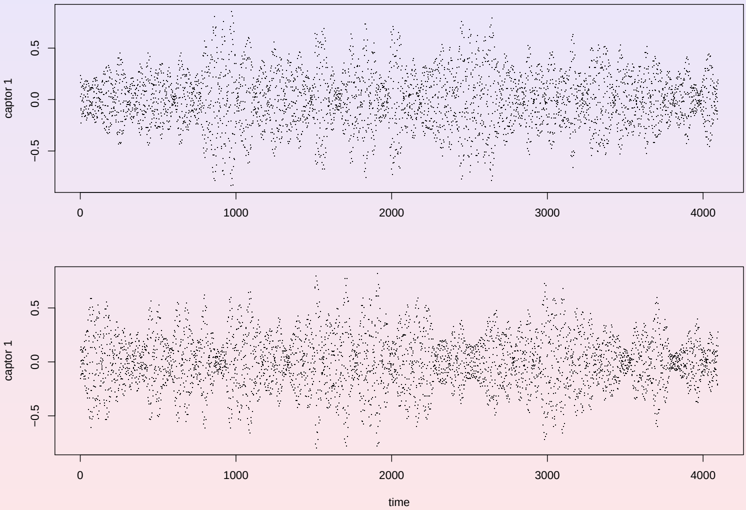
e.g., with 2 captors

$$\lambda_{j,N,2} = \sqrt{2(\log N_j + \log\log N_j)}.$$

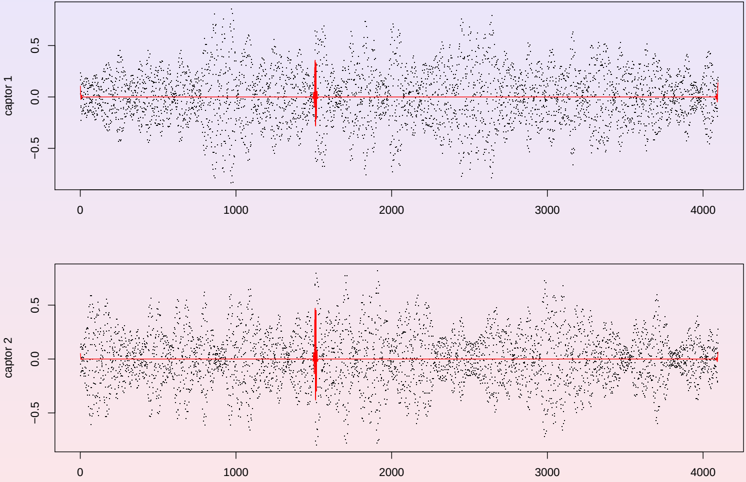
Oracle inequality:

$$\mathrm{R}(\hat{\alpha}^{\mathrm{SJS}}, \alpha) \leq (Q + \lambda_{N,Q}^2)(Q + \mathrm{R}^{\mathrm{Oracle}}(\omega, \alpha)).$$

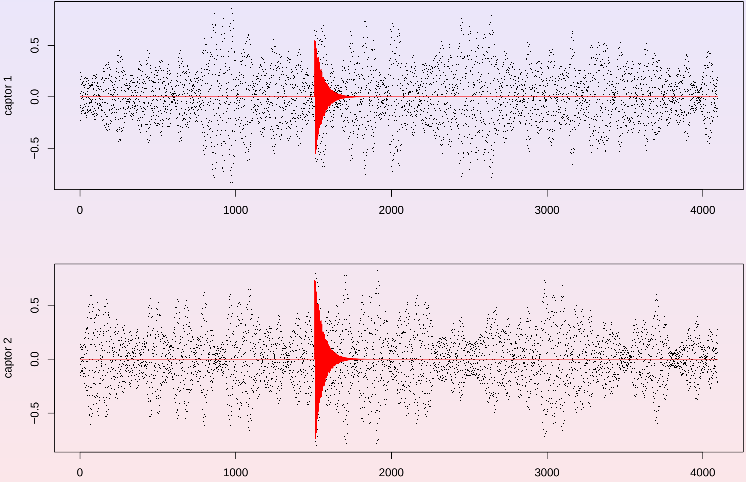
Wavebursts data



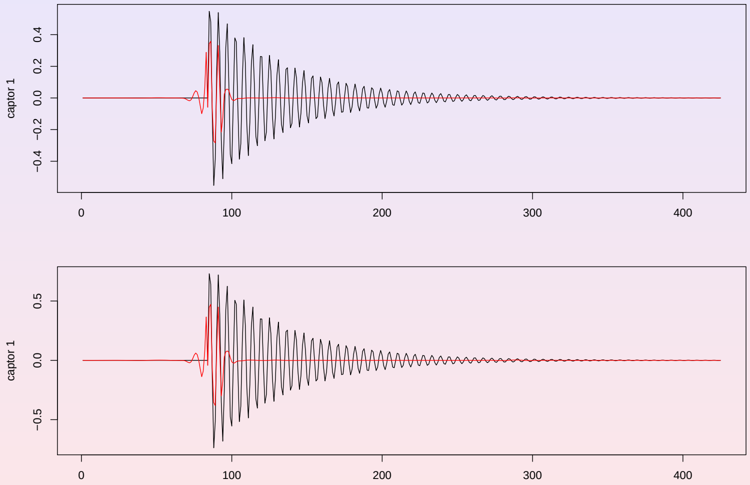
Waveburst detection: SJS with SURE thresholds



Waveburst detection: true “injections”



Zooming on the waveburst



Conclusions

Finding a needle in a haystack is possible.

Future work: compare with existing techniques

- *Performance of the WaveBurst algorithm on LIGO data*, Klimenko et al. (2004)
- *Wavelet entropy filter and Crosscorrelation of gravitational wave data*, Terenzi and Sturani (2009)

by looking at Type I, Type II errors and false descovery rates.

Density estimation/New particle or statistical fluctuation (L. Lyons)?