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

- 2 detectors: concomitant ($\neq$ repeated) measurements
- Poor signal $f$ to noise $\epsilon$ 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:

\[ f(t) = \sum_i \gamma_i \phi_i(t) + \sum_j \sum_{k=1}^{n_j} \alpha_{j,k} \psi_{j,k}(t). \]
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.
Percentage of variance explained for NMR spectrum (N=1024)
How to enforce sparsity?

With a non-linear thresholding function with threshold $\lambda$:

Anything smaller than $\lambda$ is treated as noise, and killed.
How to enforce sparsity?

With a non-linear thresholding function with threshold $\lambda$:

Anything smaller than $\lambda$ is treated as noise, and killed.
1 captor: Donoho and Johnstone’s 3 easy steps:

1. Analyze

\[ \hat{\alpha}^{\text{MLE}} =: Y = \text{DWT}(S) \quad (\text{like FFT}) \]

2. Threshold

\[ \hat{\alpha}_{j,k} = \eta_{\lambda_j}^{(\text{soft})}(Y_{j,k}) \]

\[ = \text{sign}(Y_{j,k})(|Y_{j,k}| - \lambda_j)^+ \]

\[ = (1 - \frac{\lambda_j}{|Y_{j,k}|})^+ Y_{j,k} \]

for all levels \( j = 1, \ldots, J \).

3. Reconstruct \( \sim \hat{f}_\lambda \)
captor: Donoho and Johnstone’s 3 easy steps:

1. Analyze

\[ \hat{\alpha}^{\text{MLE}} =: Y = \text{DWT}(S) \quad (\text{like FFT}) \]

2. Threshold

\[ \hat{\alpha}_{j,k} = \eta_{\lambda_j}^{(\text{soft})}(Y_{j,k}) \]
\[ = \text{sign}(Y_{j,k})(|Y_{j,k}| - \lambda_j)^+ \]
\[ = (1 - \frac{\lambda_j}{|Y_{j,k}|})^+ Y_{j,k} \]
for all levels \( j = 1, \ldots, J \).

3. Reconstruct \( \rightsquigarrow \hat{f}_\lambda \)
1 captor: Donoho and Johnstone’s 3 easy steps:

1. Analyze

\[ \hat{\alpha}^{\text{MLE}} =: Y = \text{DWT}(S) \quad (\text{like FFT}) \]

2. Threshold

\[ \hat{\alpha}_{j,k} = \eta_{\lambda_j}(Y_{j,k}) = \text{sign}(Y_{j,k})(|Y_{j,k}| - \lambda_j)_+ = (1 - \frac{\lambda_j}{|Y_{j,k}|})_+ Y_{j,k} \]

for all levels \( j = 1, \ldots, J \).

3. Reconstruct \( \sim \hat{f}_\lambda \)
2 captors: enforce concomitant/block sparsity with

\[
\left( \begin{array}{c} \hat{\alpha}_{j,k}^{(1)} \\ \hat{\alpha}_{j,k}^{(2)} \end{array} \right) = \left( 1 - \frac{\lambda_j}{\|Y_{j,k}\|_2} \right) + \left( \begin{array}{c} Y_{j,k}^{(1)} \\ Y_{j,k}^{(2)} \end{array} \right) = \left\{ \begin{array}{cc} 0 \\ 0 \end{array} \right\} \quad \text{or not}
\]

Reminiscent of the truncated James-Stein estimator

\[
\hat{\alpha}^{JS+} = (1 - \frac{P - 2}{\|Y\|^2_2}) + Y
\]

(1)
2 captors: enforce concomitant/block sparsity with

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

Reminiscent of the truncated James-Stein estimator

\[
\hat{\alpha}^{JS+} = (1 - \frac{P - 2}{\|Y\|_2^2})^+ Y
\]

(1)
Generalized James-Stein estimator

\[ \hat{\alpha}_{j,k}^{GJS,\lambda_j,\nu_j} = (1 - \frac{\lambda_j^{\nu_j}}{\| Y_{j,k} \|_2^{\nu_j}}) + Y_{j,k}, \quad \nu_j > 0 \]

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

Challenge: selection of \( \lambda \) and \( \nu \).

One possibility is to minimize the risk

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

or an unbiased estimate.
Generalized James-Stein estimator

\[ \hat{\alpha}_{j,k}^{GJS,\lambda_j,\nu_j} = (1 - \frac{\lambda_j^{\nu_j}}{\|Y_{j,k}\|_2}) + Y_{j,k}, \quad \nu_j > 0 \]

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

Challenge: selection of \(\lambda\) and \(\nu\).

One possibility is to minimize the risk

\[ R(\lambda, \nu) = \mathbb{E}\|\hat{f}_{\lambda,\nu} - f\|_2^2 \]

or an unbiased estimate.
Generalized James-Stein estimator

\[ \hat{\alpha}_{GJS, \lambda, \nu} = (1 - \frac{\lambda_j^{\nu_j}}{\|Y_{j,k}\|_2^{\nu_j}}) + Y_{j,k}, \quad \nu_j > 0 \]

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

Challenge: selection of $\lambda$ and $\nu$.

One possibility is to minimize the risk

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

or an unbiased estimate.
Selection of $\lambda_j$ and $\nu_j$ 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

$$
\hat{R}(\lambda_j, \nu_j) = \sum_{k=1}^{2^j} \|Y_{j,k}\|_2^21(\|Y_{j,k}\|_2 \leq \lambda_j) + \sum_{k=1}^{2^j} \frac{\lambda_j^{2\nu_j}}{\|Y_{j,k}\|_2^{2\nu_j}}1(\|Y_{j,k}\|_2 > \lambda_j)
+ 2\sum_{k=1}^{2^j} (Q - \frac{\lambda_j^{\nu_j}(Q - \nu_j)}{\|Y_{j,k}\|_2^{\nu_j}})1(\|Y_{j,k}\|_2 > \lambda_j) - 2^j Q.
$$

$$
= \text{RSS}(\lambda_j, \nu_j) + 2 \text{ d.f.}_j
$$

and minimize it over $\lambda_j$ and $\nu_j$. 
Generalized James–Stein: $\nu = 10$, $\gamma = 1$
Thresholding function for $\lambda = 3$

Smooth James–Stein: $\nu = 10$, $\gamma = 2\log \nu + 1$
Thresholding function for $\lambda = 3$

Subbotin $l_\nu$ penalty: $\nu = 0.4$
Thresholding function

True loss and SURE

\[ \text{SURE}(\lambda, \nu = 0.4) \]

\[ \text{SURE}(\lambda, \nu = 10; \gamma = 2\log \nu + 1) \]
Selection of $\lambda_j$ 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:

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

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

- *Wavelet entropy filter and Crosscorrelation of gravitational wave data*, Terenzi and Sturani (2009)

by looking at Type I, Type II errors and false discovery rates.
Density estimation/New particle or statistical fluctuation (L. Lyons)?