Nonlinear wavelet thresholding: A recursive method to determine the optimal denoising threshold

Kai Schneider

L3M-CNRS & CMI, Université de Provence, Marseille, France

Joint work with :

Alexandre Azzalini, LMD-CNRS, Ecole Normale Supérieure, Paris

Marie Farge, LMD-CNRS, Ecole Normale Supérieure, Paris

Outline

- Motivation
- Denoising by nonlinear wavelet thresholding
- Recursive algorithm
  - Properties of the iteration function
  - Convergence of the algorithm
  - Application to Gaussian white noise
- Numerical application
- Conclusions and perspectives

Denoising by nonlinear wavelet thresholding (I)

Discrete signal $\underline{f} = \{f[k]\}_{k\in[0,...,N-1]}, N = 2^J$ with vanishing mean

Gaussian white noise $\underline{W} = \{W[k]\}_{k\in[0,...,N-1]}$, mean zero, variance σ_W^2

Noisy signal $X[k] = f[k] + W[k]$, $0 \leq k \leq N$.

Orthogonal wavelet decomposition: $\underline{X} = \sum_{\lambda\in\Lambda^J} \tilde{X}_\lambda \underline{\psi}_\lambda$

Index set $\Lambda^J = \left\{\lambda = (j,i), j = 0...J-1, i = 0...2^j - 1\right\}$

Define nonlinear operator $F_T : \underline{X} \mapsto F_T(\underline{X}) = \sum_\lambda \rho_T(\tilde{X}_\lambda)\underline{\psi}_\lambda$

with the thresholding function

$$\rho_T(a) = \begin{cases} a & \text{for} \quad |a| > T, \\ 0 & \text{for} \quad |a| \leq T \end{cases} \tag{1}$$

where T denotes the threshold.

Denoising by nonlinear wavelet thresholding (II)

Index subset $\Lambda_T = \left\{\lambda \in \Lambda^J, |\tilde{X}_\lambda| > T\right\} \subset \Lambda^J$

Relative quadratic error:

$$\mathcal{E}(T) = \frac{\|\underline{f} - F_T(\underline{X})\|^2}{\|\underline{f}\|^2} \tag{2}$$

Donoho and Johnstone ('94):

$\mathcal{E}(T_D)$ with $T_D = \sigma_W (2 \ln N)^{1/2}$ is close to the minimum of $\mathcal{E}(T)$.

T_D depends on the variance of the noise which is unknown in many applications $\longrightarrow$ estimation.

Denoising by nonlinear wavelet thresholding (III)

Dual point of view :

$$F_T^c(\underline{X}) = (Id - F_T)(\underline{X}) = \underline{X} - F_T(\underline{X}) = \sum_{\lambda \in \Lambda^J} \rho_T^c(\tilde{X}_\lambda)\underline{\psi}_\lambda = \sum_{\lambda \in \Lambda_T^c} \tilde{X}_\lambda \underline{\psi}_\lambda \quad (3)$$

where Id denotes the identity
and with the complementary thresholding function $\rho_T^c = Id - \rho_T$
and the complementary index set $\Lambda_T^c = \Lambda^J \backslash \Lambda_T$.

Residual $F_{T_D}^c(\underline{X})$ a quasi optimal estimator of $\underline{W}$, whose relative error is

$$\mathcal{E}'(T) = \frac{\|\underline{X} - F_T(\underline{X}) - \underline{W}\|^2}{\|\underline{W}\|^2} = \frac{\|\underline{f} + \underline{W} - F_T(\underline{X}) - \underline{W}\|^2}{\|\underline{W}\|^2} = \frac{\|\underline{f}\|^2}{\|\underline{W}\|^2}\mathcal{E}(T) \quad (4)$$

Recursive algorithm (I)

Initialization

- given $\underline{X} = \{X[k]\}_{k\in[0,\ldots,N-1]}$, set n=0 and compute the Fast Wavelet Transform of $\underline{X}$ to obtain $\tilde{X}_\lambda$,

- compute the variance σ_0^2 of $\underline{X}$ as rough estimate of the variance of $\underline{W}$ and compute the corresponding threshold $\sigma_0^2 = \frac{1}{N}\sum_{\lambda\in\Lambda^J}|\tilde{X}_\lambda|^2$, $T_0^2 = 2\ln(N)\sigma_0^2$

- set the number of coefficients considered as noise $N_w = Card(\Lambda^J) = N$

Recursive algorithm (II)

Main loop

Do

- set $N'_w = N_w$ and count the wavelet coefficients smaller than T_n: $N_w = Card(\Lambda^c_{T_n})$

- compute the new variance $\sigma^2_{n+1} = \frac{1}{N}\sum_{\lambda\in\Lambda^J} |\rho^c_{T_n}(\tilde{X}_\lambda)|^2$, and the new threshold $T_{n+1} = (2\ln(N)\sigma^2_n)^{1/2}$

- Set n=n+1

until (N'_w==N_w)

Recursive algorithm (III)

Final step

- compute $F_{T_n}(\underline{X})$ from the wavelet coefficients $\{\tilde{X}_\lambda\}_{\lambda \in \Lambda_{T_n}}$ using inverse Fast Wavelet Transform and compute $F^c_{T_n}(\underline{X}) = \underline{X} - F_{T_n}(\underline{X})$

Remark: Sequence of estimated thresholds $(T_n)_{n \in \mathbb{N}}$ and variances $(\sigma_n)_{n \in \mathbb{N}}$.

Iteration function $I_{\underline{X},N} : \mathbb{R}^+ \mapsto \mathbb{R}^+$ such that $T_{n+1} = I_{\underline{X},N}(T_n)$, defined as:

$$I_{\underline{X},N}(T) = \left(\frac{2\ln(N)}{N} \sum_{\lambda \in \Lambda^J} |\rho^c_T(\tilde{X}_\lambda)|^2 \right)^{1/2} = \left(\frac{2\ln(N)}{N} \sum_{\lambda \in \Lambda^c_T} |\tilde{X}_\lambda|^2 \right)^{1/2} \quad (5)$$

Properties of the iteration function

Rewrite the sum as a continuous integral using delta functions:

$$(I_{\underline{X},N}(T))^2 = 2\ln(N)\frac{1}{N}\int_{x=0}^{T} x^2 \sum_{\lambda\in\Lambda^J} \delta(|\tilde{X_\lambda}| - x)dx \qquad (6)$$

Properties:

- $(I_{\underline{X},N}(T))$ is piece-wise constant with a number of discontinuities being bounded from above by N,
- it is monotonously increasing, i.e.

$$I_{\underline{X},N}(T) \le I_{\underline{X},N}(T + \Delta T) \quad \forall\, T, \Delta T \in \mathbb{R}^+$$

Convergence of the algorithm

Theorem 1 *We consider the interval* $[T_a, T_b] \subset \mathbb{R}^+$ *such that* $I_{\underline{X},N}(T_a) \geq T_a$ *and* $I_{\underline{X},N}(T_b) \leq T_b$. *If there exists a step* n_0 *such that* $T_{n_0} \in [T_a, T_b]$, *then* $T_n = I_{\underline{X},N}(T_{n-1})$ *converges to a limit* T_ℓ *within* $[T_a, T_b]$, *such that* $T_\ell = I_{\underline{X},N}(T_\ell)$. *The number of iterations* n_ℓ *is smaller than* N.

Corollary 1 *One has* $\sup_{T \in \mathbb{R}^+} I_{\underline{X},N}(T) = T_0 = (2\ln(N))^{1/2}\sigma_0$ *and* $I_{\underline{X},N}(0) = 0$. *Therefore theorem 1 implies that the sequence* $\{T_n\}_{n \in \mathbb{N}}$ *converges to a limit* $T_\ell \in [0, T_0]$.

Corollary 2 *Let* $\mathcal{A} : \underline{X} \mapsto F_{T_\ell}(\underline{X})$ *be the operator corresponding to the recursive algorithm described above, then* $\mathcal{A}(\mathcal{A}(\underline{X})) = \mathcal{A}(\underline{X}) \quad \forall \quad \underline{X} \in \mathcal{H}$. *This means that* $\mathcal{A}$ *is a non linear projector.*

Application to Gaussian white noise W

Orthonormality of $\{\psi_\lambda\}_{\lambda\in\Lambda^J}$ implies that $\{\tilde{W}_\lambda\}_{\lambda\in\Lambda^J}$ is also a Gaussian white noise.

Analytic expression of the PDF of its wavelet coefficients is known.

Berman (1989): probability that the maximum of the modulus of N values of a Gaussian white noise $\tilde{W}$ is inside the interval $\left[T_D - \frac{\sigma_W \ln(\ln N)}{\ln N}, T_D\right]$, i.e.

$$P(N) = p\left(\max_\lambda(|\tilde{W}_\lambda|) \in \left[T_D - \frac{\sigma_W \ln(\ln N)}{\ln N}, T_D\right]\right) \qquad (7)$$

tends to 1 for large N.

Hence for N large enough, T_D is a good estimator of the expected maximum modulus of the noise.

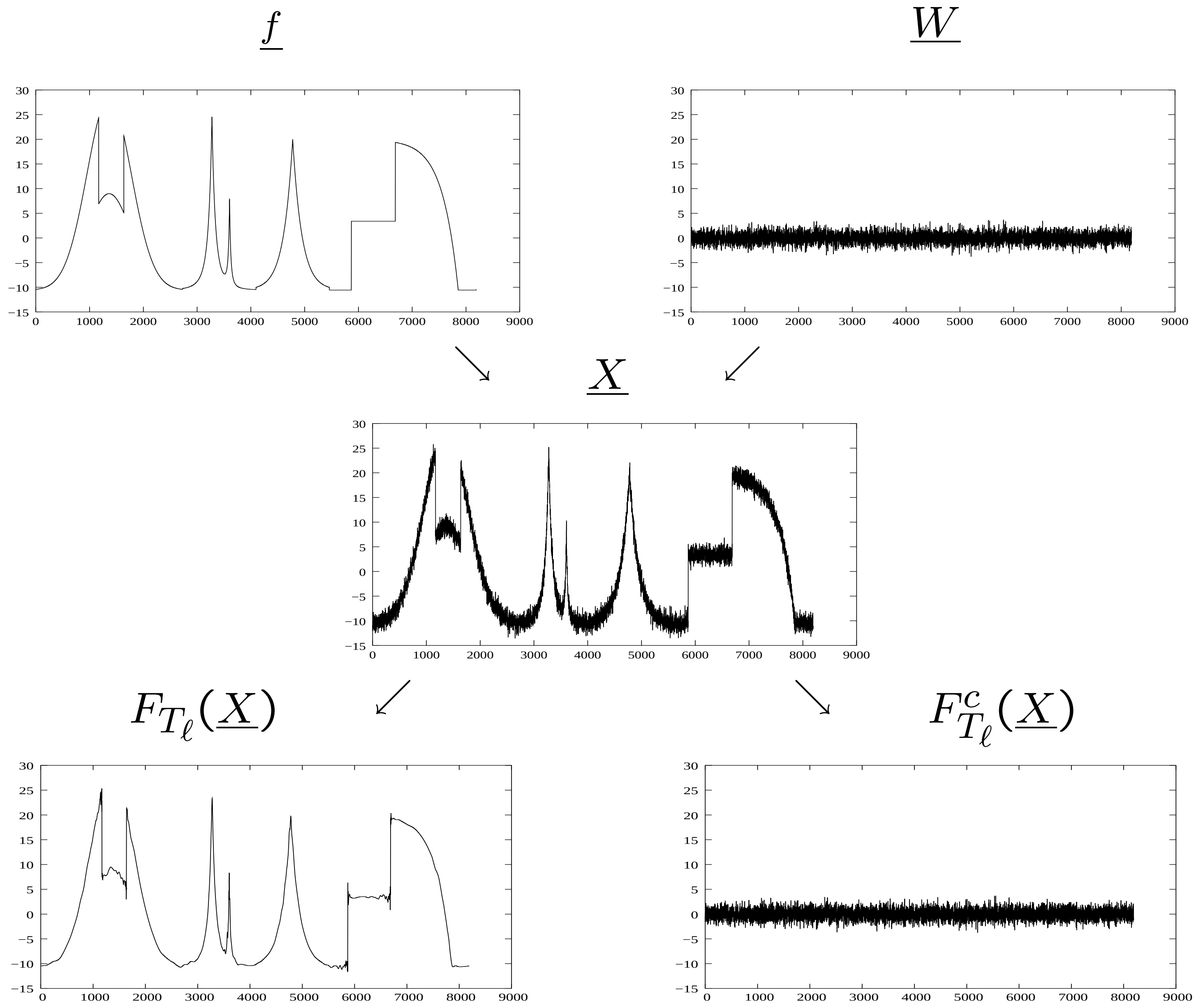
First iteration of the algorithm, we have $T_D = (2 \ln N)^{1/2} \sigma_W = (2 \ln N)^{1/2} \sigma_0 = T_0$, which yields

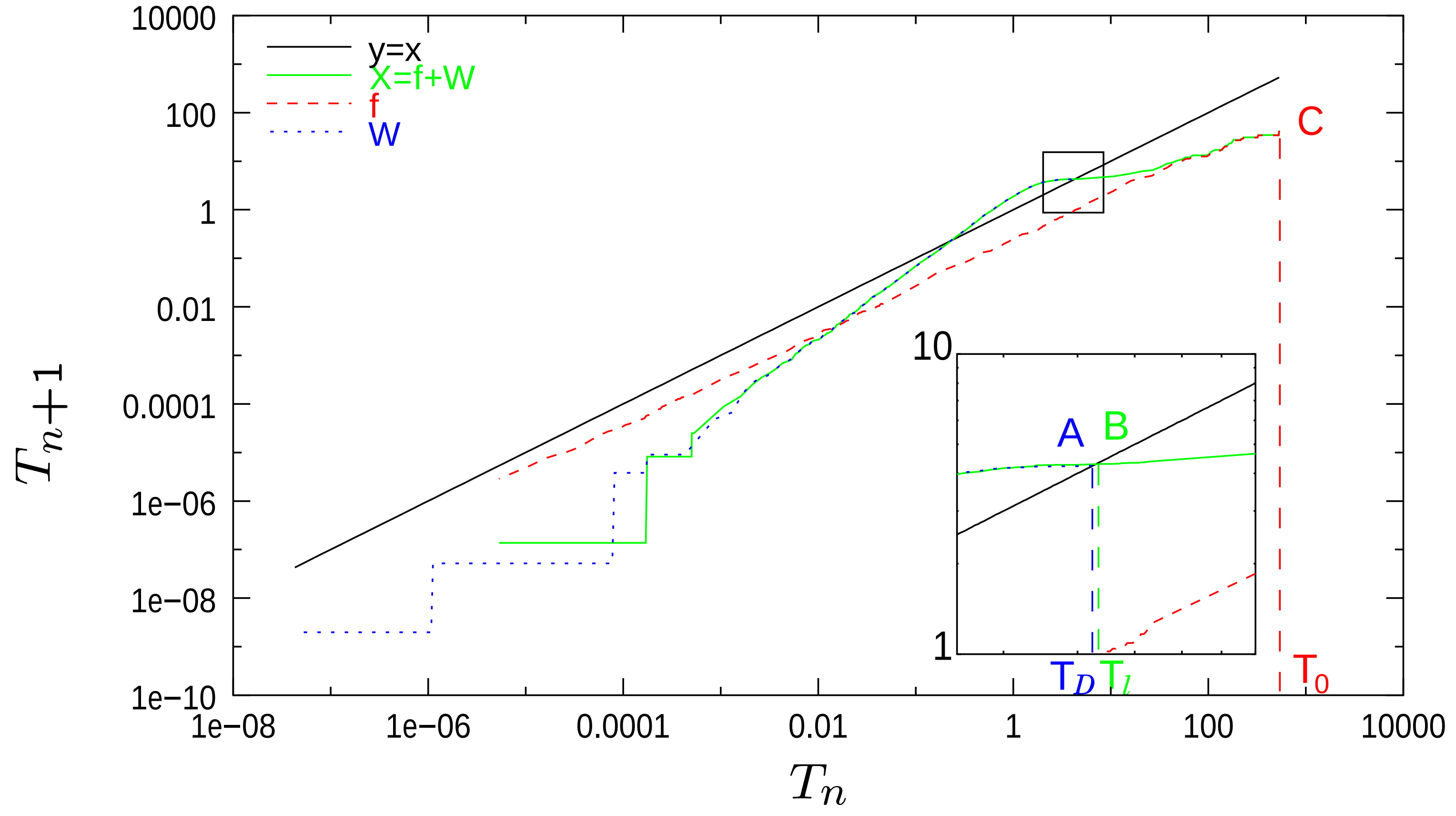
$$\begin{aligned} I_{\underline{W},N}(T_0) = I_{\underline{W},N}(T_D) &= \left(\frac{2 \ln N}{N} \sum_{\lambda \in \Lambda^J} |\rho_{T_D}(\tilde{W}_\lambda)|^2 \right)^{1/2} \\ &\simeq \left(\frac{2 \ln N}{N} \sum_{\lambda \in \Lambda^J} |\tilde{W}_\lambda|^2 \right)^{1/2} = T_0 = T_D \end{aligned} \tag{8}$$

Threshold T_0 (first iteration of the algorithm) is almost a fixed point of the iteration function $I_{\underline{W},N}$.

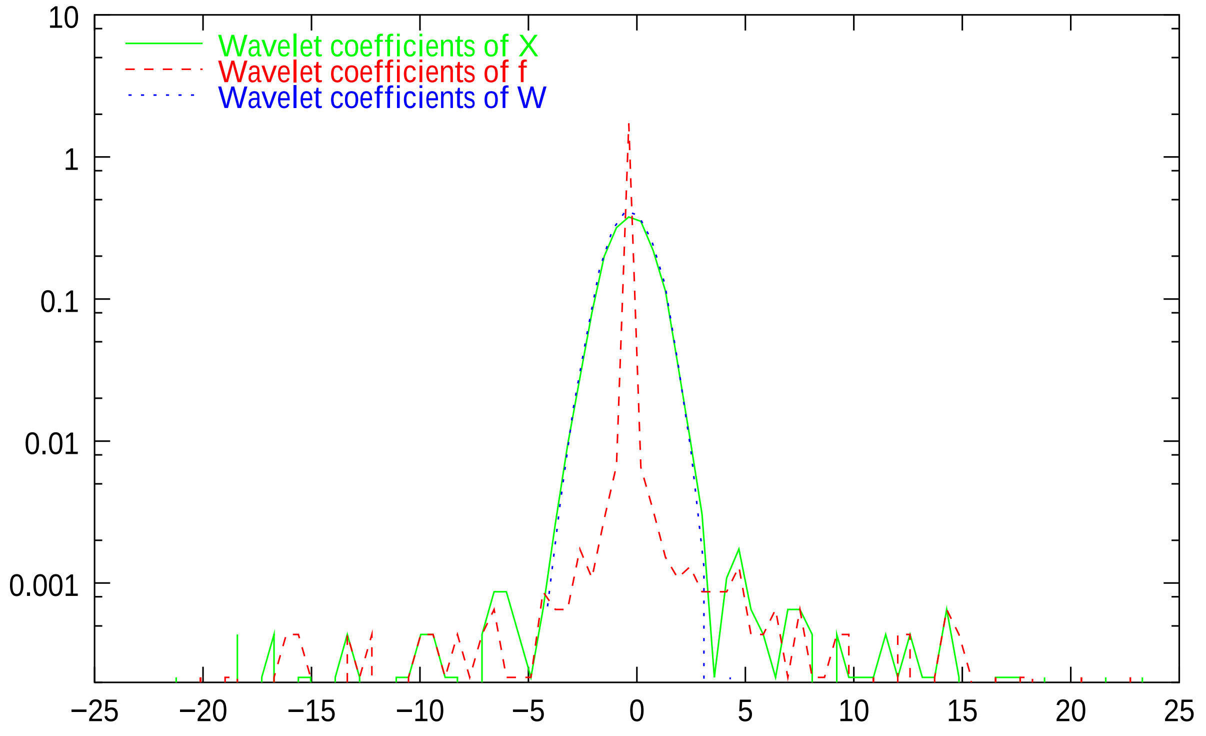
Using the analytical expression of the Gaussian PDF of the noise, one can show that the derivative of the iteration function is almost zero around T_D. This forces the threshold T_ℓ to be close to T_D and the algorithm to converge in one iteration.

Numerical application





Iteration functions $I_{\underline{W},N}$,$I_{\underline{f},N}$,$I_{\underline{X},N}$ for $\underline{W}$, $\underline{f}$ and $\underline{X}$, respectively. The points **A** and **B** correspond to the intersections between the graphs of $I_{\underline{W},N}$ and $I_{\underline{X},N}$ with the line $y = x$, respectively. The point **C** corresponds to the first iteration of the algorithm applied to the noisy signal $\underline{X}$ and its abscissa is T_0.



Histograms of the wavelet coefficients $\tilde{X}_\lambda$, $\tilde{f}_\lambda$, and $\tilde{W}_\lambda$ for the 1D signal.

Signal	n_ℓ	T_ℓ	T_m	T_D	$\mathcal{E}(T_\ell)$	$\mathcal{E}(T_m)$	$\mathcal{E}(T_D)$
X	4	4.34	4.19	4.25	$7.28\,10^{-3}$	$7.06\,10^{-3}$	$7.32\,10^{-3}$
f	21	$1.7\,10^{-6}$	$9.9\,10^{-7}$	0	$4.7\,10^{-14}$	$8.9\,10^{-16}$	0
W	1	4.24	4.19	4.24	$+\infty$	$+\infty$	$+\infty$

Thresholds T_ℓ, T_m and T_D and the corresponding mean square estimation errors.

Conclusions and perspectives

- Recursive algorithm for nonlinear wavelet thresholding

- Mathematical convergence of the algorithm

- Numerical application

- Extension to correlated noise and non Gaussian distributions

http://wavelets.ens.fr

References

- D. Donoho and I. Johnstone.
  Ideal spatial adaptation via wavelet shrinkage.
  Biometrika, **81**, 425-455, 1994.

- A. Azzalini, M. Farge and K. Schneider.
  A recursive algorithm for nonlinear wavelet thresholding : Applications to signal and image processing. Preprint LMD, August 2003

- A. Azzalini, M. Farge and K. Schneider.
  Nonlinear wavelet thresholding :
  A recursive method to determine the optimal denoising threshold.
  Appl. Comput. Harm. Anal. submitted, August 2003

- S. M. Berman. *Sojournes and Extremes of Stochastic Processes.* Wadsworth, Reading, MA, 1989.

- R. von Sachs and J. Neumann. Wavelet Thresholding: Beyond the Gaussian I.I.D. Situation. *In: A. Antoniadis, G. Oppenheim (eds.) "Wavelets and Statistics", Springer Lecture Notes in Statistics* 103, 301-329, 1995

- R. von Sachs and K. Schneider. Wavelet smoothing of evolutionary spectra by non-linear thresholding. *Appl. Comput. Harm. Anal.*, **3**, 268–282, 1996.