A recursive algorithm for wavelet denoising : applications to signal and image processing.

Alexandre Azzalini[1], Marie Farge[1] and Kai Schneider[2]
[1] Laboratoire de Météorologie Dynamique, CNRS, IPSL,
Ecole Normale Supérieure, Paris, France
[2] Laboratoire de Modélisation et de Simulation Numérique en Mécanique
& Centre de Mathématiques et d'Informatique
Université de Provence, Marseille, France
azzalini@lmd.ens.fr

November 9, 2004

Abstract

Nonlinear thresholding of wavelet coefficients has been shown to be an efficient method for denoising signals with isolated singularities corrupted with Gaussian white noise [2]. A quasi optimal value for the threshold can be computed from the noise level using the formula $T_D = \sigma_W(2\ln N)^{1/2}$, where N is the number of available samples of the signal and σ_W is the standard deviation of the noise. However, in most situations the noise level is unknown and has to be estimated. This paper studies an algorithm proposed in [3] which evaluates the value for the threshold. It recursively approximates the standard deviation of the noise with the standard deviation of the noisy signal, computes a threshold value and performs a first split from which it extracts a better estimate of the noise. Then, it iterates this procedure using the new estimate of the noise to compute the new threshold. The iteration stops when the threshold remains unchanged from its previous value. We show that the convergence of the sequence of estimated thresholds depends on a functional of the probability density function (PDF) of the noisy signal. We also find that the sequence converges towards the theoretical value T_D provided that the wavelet representation of the signal is sufficiently sparse. We compare the results obtained for examples in 1D and 2D with the results of a standard method based on the median of the wavelet coefficients of the noisy signal at small scale. Finally, we show that the recursive algorithm gives better results than this method when applied to an experimental signal measuring the atomic density of a Bose-Einstein condensate [8].

1 Introduction

Nonlinear thresholding of the wavelet coefficients has been proposed by Donoho and Johnstone [2] to denoise signals corrupted by a Gaussian white noise, and it has later been generalized to correlated and non Gaussian noises [6]. The noisy signal is transformed into wavelet coefficient space, then only the coefficients whose modulus is above a given threshold value are kept, and the denoised signal is obtained using an inverse wavelet transform. The threshold value only depends on the sample's size and the noise's variance. This method automatically adapts to the local structure of the signal and it has been proven that it minimizes the maximum L^2 risk in a whole class of functions, including Hölder and Besov spaces, without any *a priori* knowledge of the signal. It also turned out that it outperforms linear methods, in particular for functions with an inhomogeneous regularity, *e.g.* signals made of piecewise polynomials or of bounded variation. However, the unknown variance of the noise has to be estimated in order to determine the threshold. The standard method presently used is the Median Absolute Deviation (MAD), which estimates the level of the noise from the median of the modulus of wavelet coefficients at small scales [5].

We propose here a recursive algorithm to estimate the variance of the noise. We study its convergence and stability, and then apply it, first to two academic signals and then to an experimental signal, to illustrate its properties. We compare the threshold computed with the recursive scheme to the theoretical value of Donoho and Johnstone [2] and to the value obtained using the MAD method. The results show that the new algorithm is competitive and efficient.

The paper is organized as follows: after describing the nonlinear wavelet thresholding in section 2 we present the recursive algorithm to determine the threshold by estimating recursively the variance of the noise. We prove the convergence of the algorithm and show that it is a nonlinear projector. In section 4 we present its numerical validation by applying the algorithm to $1D$ and $2D$ academic test signals and to an experimental 2D observation of a Bose-Einstein condensate. The results of the recursive algorithm are compared to the results given using the MAD method. Finally, we conclude and give some perspectives for future work.

2 Denoising by nonlinear wavelet thresholding

A classical problem of signal processing consists in estimating a good approximation of a signal f from noisy samples $\underline{X}$. Donoho and Johnstone [2] proposed to use nonlinear wavelet thresholding. They have shown that this is particularly attractive for the case when f has isolated singularities.

Here we consider a discrete signal $\underline{f} = \{f[k]\}_{k\in[0,\ldots,N-1]}$ of size $N = 2^J$

with vanishing mean. The values $f[k]$ are samples of a function f. We observe noisy data of size N, *i.e.* $\underline{X} = \{X[k]\}_{k\in[0,\ldots,N-1]}$ such that

$$X[k] = f[k] + W[k] \tag{1}$$

where $\underline{W} = \{W[k]\}_{k\in[0,\ldots,N-1]}$ are N samples of a Gaussian white noise with variance σ_w^2, *i.e.* $W \in \mathcal{N}(0, \sigma_w)$.

We decompose the observed data $\underline{X}$ into an orthogonal wavelet series

$$\underline{X} = \sum_{\lambda\in\Lambda^J} \tilde{X}_\lambda \underline{\psi}_\lambda \tag{2}$$

with the wavelet coefficients

$$\tilde{X}_\lambda = \langle \underline{X} | \underline{\psi_\lambda} \rangle \tag{3}$$

The multi–index $\lambda = (j, i)$ denotes the scale j and the position i of the wavelets. The corresponding index set Λ^J is given by

$$\Lambda^J = \left\{ \lambda = (j, i), j = 0...J-1, i = 0...2^j - 1 \right\} \tag{4}$$

The family (ψ_λ) constitutes an orthogonal multi–resolution analysis of $L^2(\mathbb{R})$ [5].

By thresholding the wavelet coefficients $\tilde{X}_\lambda$ and reconstructing the corresponding signal we define a nonlinear operator

$$F_T : \underline{X} \mapsto F_T(\underline{X}) = \sum_\lambda \rho_T(\tilde{X}_\lambda) \underline{\psi}_\lambda \tag{5}$$

with

$$\rho_T(a) = \left\{ \begin{smallmatrix} a \text{ if } |a|>T \\ 0 \text{ if } |a|\leq T \end{smallmatrix} \right. \tag{6}$$

where T denotes the threshold. The operator F_T hence projects the signal $\underline{X}$ onto the orthogonal wavelet basis $(\underline{\psi}_\lambda)$ and uses the thresholding function ρ_T for selecting those wavelet coefficients $\tilde{X}_\lambda$ whose magnitude is larger than the threshold T. Subsequently, it reconstructs $F_T(\underline{X})$ in physical space from the retained coefficients.

For later convenience we introduce the index subset

$$\Lambda_T = \left\{ \lambda \in \Lambda^J, |\tilde{X}_\lambda| > T \right\} \subset \Lambda^J \tag{7}$$

which is the set of wavelet coefficients $\tilde{X}$ that are selected by the thresholding function ρ_{T_D}.

Donoho and Johnstone showed that F_T with the threshold

$$T_D = \sigma_W (2 \ln N)^{1/2} \tag{8}$$

yields minimax estimators for all $f \in \mathcal{H}$ where $\mathcal{H}$ belongs to a wide class of function spaces, including Hölder and Besov spaces. They showed that the maximum mean-square error

$$R(F, H) = \sup_{f \in H} \mathbb{E} \left\{ \| \underline{f} - F(\underline{X}) \|^2 \right\}, \tag{9}$$

which depends on the function space $\mathcal{H}$ *e.g.*$\mathcal{H} = B^{\alpha}_{p,q}$ and on the used operator F, is almost minimized by the nonlinear wavelet estimator F_{T_D}.

More precisely, the relative quadratic error between the signal $\underline{f}$ and its estimator $F_T(\underline{X})$ defined by

$$\mathcal{E}(T) = \frac{\| \underline{f} - F_T(\underline{X}) \|^2}{\| \underline{f} \|^2} \tag{10}$$

has its lower bound $\min_T \mathcal{E}(T)$ close to the minimax error

$$\min_F R(F, H).$$

Moreover, the threshold value T_D in (8), is close to the threshold $T_{\min}$ that minimizes $\mathcal{E}(T)$ but which depends on each particular signal $\underline{f}$. In contrast the threshold T_D depends exclusively on the variance of the noise and therefore it is called universal threshold.

One has

$$T_D \simeq T_{\min}$$

and

$$\mathcal{E}(T_D) \gtrsim \mathcal{E}(T_{\min})$$

An illustration of the quasi-optimality of T_D is given in Fig. 1. We plot the relative error $\mathcal{E}(T)$ for a piece-wise regular signal corrupted with a Gaussian white noise versus the threshold value T. We observe that the universal threshold T_D almost corresponds to the minimum of $\mathcal{E}(T)$, obtained with the optimal threshold. This example also shows that if the threshold value T is chosen above T_D, then the error increases significantly. On the contrary, if T is chosen below T_D, the error tends to the value $\mathcal{E}(0)$ corresponding to no denoising at all.

This sensitivity implies that one has to know the value of the threshold T_D precisely to obtain an accurate estimator F_{T_D} of the signal $\underline{f}$ from the noisy data $\underline{X}$. Hence, the knowledge of the variance of the noise is of primordial interest. As the level of the Gaussian white noise σ_W involved in the expression of T_D is generally unknown, the problem one encounters in practice is to get a good estimation of σ_W from the available noisy data $\underline{X}$.

To address the estimation of the noise, instead of the estimation of the signal, we adopt a dual point of view. Instead of considering $F_{T_D}(\underline{X})$ which

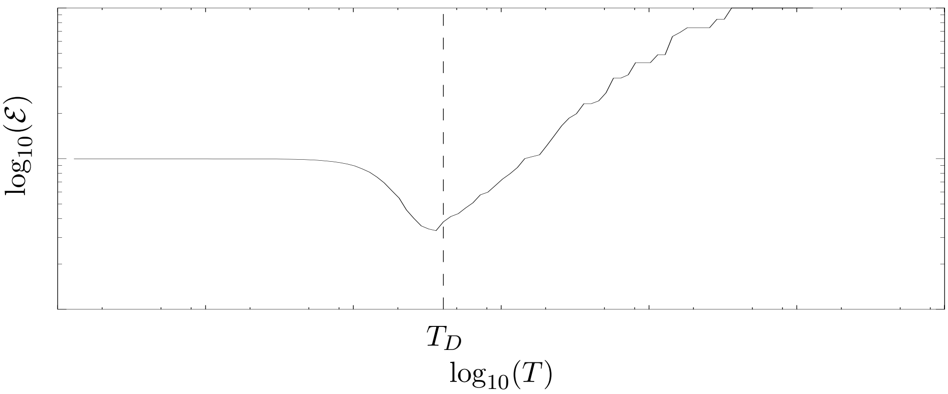

Figure 1: Error $\mathcal{E}(T)$ versus different threshold values T for a piecewise regular signal (see Fig. 2) The vertical line indicates the universal threshold T_D.

is a version of the signal $\underline{X}$ from which a major part of the Gaussian noise has been removed, we focus on the residual of $\underline{X}$ which was not taken into account in $F_{T_D}(\underline{X})$, namely $(\underline{X} - F_{T_D}(\underline{X}))$. It is a quasi optimal estimator of the Gaussian white noise $\underline{W}$, whose relative error is

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

The two points of view are equivalent, as

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

Thus, the value of T minimizing $\mathcal{E}(T)$ also minimizes $\mathcal{E}'(T)$.

Following this dual approach, it is useful to introduce the complementary operators. The operator that estimates the noise from $\underline{X}$ is defined by

$$F_T^c = Id - F_T \tag{11}$$

where Id denotes the identity, and it uses the complementary coefficient selector

$$\rho_T^c = Id - \rho_T \tag{12}$$

The corresponding complementary index set is defined as

$$\Lambda_T^c = \Lambda^J \backslash \Lambda_T \tag{13}$$

Hence it follows for the estimator of the noise that

$$F_T^c(\underline{X}) = (\underline{X} - F_T(\underline{X})) = \sum_{\lambda \in \Lambda^J} \rho_T^c(\tilde{X}_\lambda)\underline{\psi}_\lambda = \sum_{\lambda \in \Lambda_T^c} \tilde{X}_\lambda \underline{\psi}_\lambda \tag{14}$$

3 Recursive algorithm

In [3] we proposed a recursive algorithm to extract coherent vortices from turbulent vorticity fields using nonlinear wavelet thresholding. The principle of the scheme consists in extracting first a rough estimation of the Gaussian part by using the variance of the total signal as estimator for the variance of the noise. In the next step an improved threshold is obtained by using the variance of the noise thus extracted. This improved threshold is used to extract a better estimate of the noise. The above procedure is iterated until the number of wavelet coefficients of the noise is constant.

In the following we present the algorithm and study its mathematical properties.

Algorithm:

Initialization

- given $\underline{X} = \{X[k]\}_{k \in [0,\dots,N-1]}$
- set n=0
- compute the Fast Wavelet Transform of $\underline{X}$ to obtain $\tilde{X}_\lambda$
- compute $\sigma_0^2 = \frac{1}{N} \sum_{\lambda \in \Lambda^J} |\tilde{X}_\lambda|^2$ as rough estimate of the variance of the noise. Note that due to orthonormality of the wavelet basis, the variance of $\underline{X}$ can be calculated from its wavelet coefficients.
- set $N_W = N$, which corresponds to the number of coefficients considered as noise.
- compute the threshold $T_0^2 = 2 \ln(N) \sigma_0^2$.

Main loop

Do

- $N'_W = N_W$
- compute the number of wavelet coefficients smaller than T_n
- $N_W = Card(\Lambda^c_{T_n})$
- compute the new variance

$$\sigma_{n+1}^2 = \frac{1}{N} \sum_{\lambda \in \Lambda^J} |\rho^c_{T_n}(\tilde{X}_\lambda)|^2$$

- compute the new threshold

$$T_{n+1} = (2\ln(N)\sigma_{n+1}^2)^{1/2}$$

- Set n=n+1

until (N'_W==N_W)

Final step

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

End

This algorithm defines the sequence of thresholds $(T_n)_{n\in\mathbb{N}}$ and the corresponding sequence of variances $\left(\sigma_n^2\right)_{n\in\mathbb{N}}$ which are respectively the successive estimates of the threshold T_D and the standard deviation of the noise σ_W given by the algorithm. In the following we show that they converge to limits giving a mean square error $\mathcal{E}$ close to its minimum.

The algorithm has the following properties :

- it uses only one Fast Wavelet Transform to determine the threshold and only one more for computing both $F_{T_n}(\underline{X})$ and $F^c_{T_n}(\underline{X})$.
- the convergence criterion is always satisfied after a finite number of iterations smaller than N (this is shown below, see corollary 1). Therefore no stopping criterion based on some *ad hoc* parameter is needed.
- it is recursive and there exists an *iteration function*

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

which contains all the information about the convergence of the algorithm for a given initial condition. This function allows to study the algorithm using principles of the fixed point theory. It is obtained by merging the definitions of σ_{n+1} and T_{n+1} :

$$I_{\underline{X},N}(T) = \left(\frac{2\ln(N)}{N}\sum_{\lambda\in\Lambda^J}|\rho^c_T(\tilde{X}_\lambda)|^2\right)^{1/2} = \left(\frac{2\ln(N)}{N}\sum_{\lambda\in\Lambda^c_T}|\tilde{X}_\lambda|^2\right)^{1/2} \tag{15}$$

3.1 Properties of the iteration function

Taking the square of (15), it is possible to rewrite the sum as a continuous integral using delta functions,

$$(I_{\underline{X},N}(T))^2 = 2\ln(N)\frac{1}{N}\int_{x=0}^{T} x^2 \sum_{\lambda\in\Lambda^J} \delta(|\tilde{X_\lambda}| - x)dx \tag{16}$$

The function $(I_{\underline{X},N}(T))^2$ has the following properties:

- it is piece-wise constant with a number of discontinuities being bounded from above by N,
- it is monotonically increasing, *i.e.*

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

Furthermore, the iteration function $I_{\underline{X},N}$ is related to the empirical histogram of the wavelet coefficients $|\tilde{X}|$,

$$h(x_0, \Delta x) = \frac{1}{N}\int_{x=x_0-\Delta x/2}^{x=x_0+\Delta x/2} \sum_{\lambda} \delta(|\tilde{X_\lambda}| - x)dx \tag{17}$$

that counts the number of coefficients $|\tilde{X_\lambda}|$ whose value is in the bin of width Δx centered at x_0. The histogram $h(x_0, \Delta x)$ converges to the PDF of $|\tilde{X}|$ for the limits Δx tending to zero and N tending to infinity.

By writing the sum

$$S_{\underline{X},N}(K,T) = \frac{1}{N}\sum_{k=0}^{K} x_k^2 h(x_k, \frac{T}{K}) \tag{18}$$

where $x_k = \frac{T}{K}(k+\frac{1}{2})$, one observes that $2\ln(N)S_{\underline{X},N}(K,T)$ converges to (16) for the limit K tending to infinity. Hence, (16) is an empirical estimator of the 2^{nd} order moment of the PDF of the coefficients $|\tilde{X}|$ smaller than T.

3.2 Convergence

In the following we prove the convergence of the recursive algorithm. Therefore we apply fixed point type arguments to the iteration function $I_{\underline{X},N}$.

Theorem 1. *We consider the interval $[T_a, T_b] \subset \mathbb{R}^+$ with $I_{\underline{X},N}(T_a) \geq T_a$ 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 $T_\ell = I_{\underline{X},N}(T_\ell)$.*

Proof. Suppose that $I_{\underline{X},N}(T_{n_0}) \neq T_{n_0}$

If

$$I_{\underline{X},N}(T_{n_0}) < T_{n_0} \tag{19}$$

it follows that

$$T_{n_0+1} < T_{n_0} \tag{20}$$

as $I_{\underline{X},N}$ is monotonically increasing, we have

$$I_{\underline{X},N}(T_{n_0+1}) \leq I_{\underline{X},N}(T_{n_0}). \tag{21}$$

This leads to

$$T_{n_0+2} < T_{n_0+1} \tag{22}$$

and so, for all $n \geq n_0$, we obtain

$$T_{n+1} < T_n, \tag{23}$$

$$\tag{24}$$

which means that the sequence $\{T_n\}_{n \geq n_0}$ decreases.

As T_{n_0} is in $[T_a, T_b]$, it follows that

$$T_a < T_{n_0} \tag{25}$$

and hence

$$I_{\underline{X},N}(T_a) \leq I_{\underline{X},N}(T_{n_0}). \tag{26}$$

As we assumed

$$T_a \leq I_{\underline{X},N}(T_a) \tag{27}$$

we find

$$T_a \leq T_{n_0+1} \tag{28}$$

therefore, we have for all $n \geq n_0$

$$T_a \leq T_n \tag{29}$$

Hence $\{T_n\}_{n\geq n_0}$ decreases and is bounded from below by T_a. Consequently, it converges to a limit $T_\ell = \inf_{n\geq n_0}(T_n)$ between T_a and T_{n_0}. As the iteration function $I_{\underline{X},N}$ is piece-wise constant with a finite number of discontinuities, its image including the values taken by the sequence $\{T_n\}_{n>n_0}$ is countable and finite. As a consequence, there exists a n_ℓ such that

$$T_{n_\ell} = T_\ell = \inf_{n\geq n_0}(T_n). \tag{30}$$

By definition of the lower bound, we have,

$$T_\ell = \inf_{n\geq n_0}(T_n) \leq T_{n_\ell+1}, \tag{31}$$

on the other hand, the sequence $\{T_n\}_{n\geq n_0}$ decreases, therefore

$$T_{n_\ell+1} = I_{\underline{X},N}(T_{n_\ell}) \leq T_{n_\ell}. \tag{32}$$

Hence

$$T_{n_\ell} \leq I_{\underline{X},N}(T_{n_\ell}) \leq T_{n_\ell} \tag{33}$$

and therewith, we have shown that

$$T_{n_\ell} = I_{\underline{X},N}(T_{n_\ell}) \tag{34}$$

Conversely, if

$$I_{\underline{X},N}(T_{n_0}) > T_{n_0} \tag{35}$$

one can show analogously that $\{T_n\}_{n\geq n_0}$ is increasing and upper-bounded, and therefore converges between T_{n_0} and T_b.

□

Corollary 1. *One has*

$$I_{\underline{X},N}(0) = 0$$

and

$$\sup_{T\in\mathbb{R}^+} I_{\underline{X},N}(T) = T_0 = (2\ln(N))^{1/2}\,\sigma_0.$$

Therefore the sequence $\{T_n\}_{n\in\mathbb{N}}$ converges to a limit $T_\ell \in [0, T_0]$. In addition, the limit T_ℓ is reached after a finite number of iterations n_ℓ bounded from above by N.

Proof. Using (15) to compute $I_{\underline{X},N}(0)$ from those coefficients whose modulus is smaller than 0, it naturally follows that $I_{\underline{X},N}(0) = 0$. On the other hand, for any other threshold T, the value $I_{\underline{X},N}(T)$ is maximum when all the coefficients $\tilde{X}_\lambda$ are taken in the sum of expression (15). This maximum is equal to T_0 and it is obtained for any threshold T larger than

$$T_{\max} = \sup_{\lambda \in \Lambda^J} |\tilde{X}_\lambda|.$$

Therefore, there exists a threshold $T_b \geq \max(T_0, T_{\max})$ such that $I_{\underline{X},N}(T_b) = T_0 \leq T_b$.

Now we apply theorem 1 to $I_{\underline{X},N}$ on the interval $[0, T_b]$ to show that the sequence $\{T_n\}_{n\in\mathbb{N}}$ converges to a limit $T_\ell \in [0, T_b]$. As $I_{\underline{X},N}(T_b) = T_0$, the limit T_ℓ is actually in $[0, T_0]$.

In addition, the proof of theorem 1 shows that $\{T_n\}_{n\in\mathbb{N}}$ reaches the limit T_ℓ after a finite number of iterations n_ℓ. This number is bounded by the finite number of discontinuities of $I_{\underline{X},N}$ and is smaller than the number N of wavelet coefficients $\tilde{X}_\lambda$. □

Another consequence of theorem 1 is the stability and self consistency of the recursive algorithm. The following corollary shows that when the noisy part of a signal $\underline{X}$ has been removed by the recursive procedure, a second pass does not change the result previously obtained.

Corollary 2. *Let*

$$\mathcal{A} : \; \underline{X} \mapsto F_{T_\ell}(\underline{X})$$

be the operator corresponding to the recursive algorithm described above, then

$$\mathcal{A}(\mathcal{A}(\underline{X})) = \mathcal{A}(\underline{X}) \qquad \forall \quad \underline{X} \in \mathcal{H}.$$

This means that $\mathcal{A}$ is a non linear projector.

Hence, if one applies the algorithm to the result of a previous estimation, the recursive procedure yields a threshold which is equal to zero. Therefore, the resulting estimation coincides with the previous one.

Proof. This property can be shown by looking at the graph of the iteration function corresponding to $\mathcal{A}(\underline{X})$ defined as

$$I_{\mathcal{A}(\underline{X}),N}(T) = \left(\frac{2\ln(N)}{N} \sum_{\lambda \in \Lambda^J} |\rho_T^c(\rho_{T_\ell}(\tilde{X}_\lambda))|^2 \right)^{1/2}$$

where $T_\ell > 0$ is the threshold obtained with the first recursive procedure. The iteration function $I_{\mathcal{A}(\underline{X}),N}$ has the following properties

$$I_{\mathcal{A}(\underline{X}),N}(T) < I_{\underline{X},N}(T) \qquad \forall \quad T \in \mathbb{R}^+ \tag{36}$$

as it corresponds to a partial sum of the terms in $I_{\underline{X},N}(T)$. Furthermore, the fact that

$$\rho_T^c \circ \rho_{T_\ell} = 0 \qquad \forall \quad T < T_\ell \tag{37}$$

implies

$$I_{\mathcal{A}(\underline{X}),N}(T) = 0 \qquad \forall \quad T < T_\ell. \tag{38}$$

As we have shown that

$$I_{\underline{X},N}(T) \leq T \qquad \forall \quad T \geq T_\ell \tag{39}$$

it follows that

$$I_{\mathcal{A}(\underline{X}),N}(T) < T \qquad \forall \quad T \geq T_\ell. \tag{40}$$

On the other hand, we have for $T < T_\ell$ that

$$I_{\mathcal{A}(\underline{X}),N}(T) = 0 \leq T. \tag{41}$$

The equality holds for $T = 0$, which is thus the only fixed point of $I_{\mathcal{A}(\underline{X}),N}$ and therefore this is the only possible limit for the sequence of thresholds $\{T_n\}_{n\in\mathbb{N}}$. □

3.3 Convergence for Gaussian white noise

In this subsection, we study theoretically the situation when the algorithm is applied to a Gaussian white noise $\underline{W}$. In this case, as the analytic expression of the probability density function of the noise is known, it is possible to derive conclusions on the behavior of the recursive algorithm.

The orthonormality of (ψ_λ) implies that $\{\tilde{W}_\lambda\}_{\lambda\in\Lambda^J}$ is also Gaussian white noise. Therefore, it is possible to compute the probability for a wavelet coefficient of the noise $\underline{W}$ to be above the threshold T_D

$$\begin{aligned} P\left(|\tilde{W}_\lambda| > T_D\right) &= 1 - P\left(|\tilde{W}_\lambda| \leq T_D\right) \\ &= 1 - \frac{2}{\sigma_W\sqrt{2\pi}} \int_{T_D}^{\infty} exp\left(\frac{-\tilde{w}^2}{2\sigma_W^2}\right) d\tilde{w} \\ &= 1 - \mathrm{erf}\left(\frac{T_D}{\sigma_W 2^{1/2}}\right) = 1 - \mathrm{erf}(\sqrt{\ln(N)}) \end{aligned} \tag{42}$$

where $\mathrm{erf}(x) = \frac{2}{\pi^{1/2}}\int_0^x \exp(-t^2)dt$.

An asymptotic expansion of $\mathrm{erf}(x)$ for large x shows that

$$P\left(|\tilde{W}_\lambda| > T_D\right) = \frac{1}{N\sqrt{\pi \ln(N)}} + o\left(\frac{1}{N\sqrt{\ln(N)}}\right) \tag{43}$$

For the probability that the maximum wavelet coefficient of a Gaussian white noise sampled over N values is above T_D one has

$$P\left(\max_{\lambda\in\Lambda^J}\left(|\tilde{W}_\lambda|\right) > T_D\right) = 1 - P\left(\max_{\lambda\in\Lambda^J}\left(|\tilde{W}_\lambda|\right) \leq T_D\right). \tag{44}$$

As the coefficients $\tilde{W}_\lambda$ are independent, one can express the probability of the maximum value of $\tilde{W}_\lambda$ as a function of the probability of the single variable $\tilde{W}_\lambda$

$$\begin{aligned} P\left(\max_{\lambda\in\Lambda^J}\left(|\tilde{W}_\lambda|\right) > T_D\right) &= 1 - \left[P\left(|\tilde{W}_\lambda| \leq T_D\right)\right]^N \\ &= 1 - \left[\mathrm{erf}(\sqrt{\ln(N)})\right]^N. \end{aligned} \tag{45}$$

Using (43) this yields

$$P\left(\max_{\lambda\in\Lambda^J}\left(|\tilde{W}_\lambda|\right) > T_D\right) \sim O\left(\frac{1}{\sqrt{\pi \ln(N)}}\right) \tag{46}$$

These results show that for N large enough, there is almost no chance for any value of $|\tilde{W}|$ to be larger than the value T_D. Hence, for almost all realizations, one has

$$\Lambda^c_{T_D} = \Lambda^J.$$

One can now remark that, following definitions of T_D and T_0, one has

$$T_D = (2\ln(N))^{1/2}\,\sigma_W = (2\ln(N))^{1/2}\,\sigma_X = (2\ln(N))^{1/2}\,\sigma_0 = T_0.$$

As a consequence, the first step of the algorithm yields

$$I_{\underline{W},N}(T_0) = I_{\underline{W},N}(T_D) = \left(\frac{2\ln(N)}{N}\sum_{\lambda\in\Lambda^J}|\tilde{W}_\lambda|^2\right)^{1/2} = T_0 = T_D \tag{47}$$

This shows that the threshold value T_0 corresponding to the initial value of the sequence $\{T_n\}_{n\in\mathbb{N}}$ is a fixed point of the iteration function $I_{\underline{W},N}$. This results in stopping the recursive algorithm after the first step.

In addition, using the analytical expression of the Gaussian probability density function of the noise and the link between the iteration function

$I_{\underline{W},N}$ and the empirical histogram defined in (17), one can show that there exists a threshold value $T_a < T_0$ such that for all T in $[T_a, T_0[$, one has

$$I_{\underline{W},N}(T) \geq T.$$

This confirms that the bell shape of the Gaussian PDF of the wavelet coefficients of the noise, with its large width and its strong decay at the tails, is responsible for the convergence of the algorithm towards the limit $T_\ell = T_D = T_0$. Note that T_0 also corresponds to the value $T_{\max}$ defined in the proof of corollary 1.

As a consequence, the noise $\underline{W}$ is invariant with respect to the recursive noise extractor $\mathcal{A}^c$ defined as

$$\mathcal{A}^c : \ \underline{X} \mapsto F^c_{T_\ell}(\underline{X}) = Id - \mathcal{A}(\underline{X})$$

One has

$$\mathcal{A}^c(\underline{W}) = W$$

In other words, the recursive algorithm is able to perfectly identify $\underline{W}$ as Gaussian white noise rather than a signal.
The remaining question is to determine if T_ℓ is a correct estimator of T_D, for $\underline{X}$ being a noisy signal resulting of the superposition of a given signal $\underline{f}$ and a noise $\underline{W}$. The next section does not prove this result formally, but instead shoes as a first approach that it is verified for a set of various numerical examples.

4 Numerical application

4.1 Application to 1D and 2D test signals.

In the following we validate numerically the above recursive algorithm for for 1D and 2D test signals and illustrate its properties. We construct a noisy signal $\underline{X}$ by superposing to different signals $\underline{f}$ a Gaussian white noise $W \in \mathcal{N}(0, \sigma_W)$ with given variance σ_W^2, produced by a standard random number generator. First we apply the recursive algorithm to the signal $\underline{f}$ without any noise, then to the noise $\underline{W}$ only, and finally to the noisy signal $\underline{X}$ for several signal to noise ratios (SNR), where SNR is defined by :

$$SNR = 10 \log_{10}\left(\frac{\sigma_f^2}{\sigma_W^2}\right) = 20 \log_{10}\left(\frac{\sigma_f}{\sigma_W}\right). \tag{48}$$

The aim is to track the influence of $\underline{f}$ and $\underline{W}$ on the results obtained for the total signal $\underline{X}$ by looking at the influence of $I_{\underline{f},N}$ and $I_{\underline{W},N}$ on the iteration function of the total signal $I_{\underline{X},N}$.

For each signal to noise ratio, we compare the results of non-linear wavelet thresholding using the recursively found threshold T_ℓ and the universal threshold T_D computed with the known variance of the noise σ_W^2. In order to evaluate the performance of our method, we also compare the results with those obtained using the estimator T_m of the universal threshold given by the Median Absolute Deviation method (MAD) [2]. This method relies on the fact that the sparsity of the wavelet coefficients of $\underline{f}$ increases for smaller scales for signals with isolated singularities. Hence, the median of the modulus of the wavelet coefficients of the noisy signal is insensitive to the amplitude of these few strong outliers. Therefore, it is a good estimator of the median of the modulus of the coefficients of the noise. For Gaussian white noise we have,

$$\operatorname*{med}_{\lambda=(j,i)\in\{(j,i),j=J\}}(|\tilde{W}_\lambda|) = 0.6745\,\sigma_W\,.$$

From this formula, one obtains the MAD threshold

$$T_m = \frac{(2\ln(N))^{1/2}}{0.6745}\operatorname*{med}_{\lambda=(j,i)\in\{(j,i),j=J\}}(|\tilde{X}_\lambda|) \tag{49}$$

4.1.1 Application to a 1D signal without noise

This section presents the iteration function $I_{\underline{f},N}$ corresponding to the signal $\underline{f}$ shown on Fig. 2. It is a piece-wise regular signal provided with the WaveLab software package [9]. It has been normalized such that

$$\sigma_f = (\frac{1}{N}\sum_{k=0}^{N-1}|f[k]|^2)^{1/2} = 10$$

We compute the discrete wavelet transform of the discrete signal $\underline{f}$ sampled on $N = 8192$ points using the Coiflet wavelets with four vanishing moments. These wavelets are almost symmetric, and the coefficients corresponding to the scaling function at the smallest scale are almost equal to the samples $\underline{f}$. This reduces the computation of the wavelet coefficients to the simple Mallat's fast wavelet transform and avoid expensive interpolation procedures.

In Fig. 3, we plotted the iteration function $I_{\underline{f},N}(T)$ for values T in the interval $[10^{-8}T_{\max}\,, T_{\max}$, where $T_{\max} = \sup_{\lambda\in\Lambda^J}|\tilde{f}_\lambda|$ is the magnitude of the largest wavelet coefficient of the signal $\underline{f}$. We observe that the graph of $I_{\underline{f},N}$ remains below the line $y = x$ and hence has no fixed point in this interval. According to theorem 1, the sequence of thresholds $\{T_n\}_{n\in\mathbb{N}}$ should therefore converge to a limit T_ℓ below $10^{-8}T_{\max}$.

When applying the recursive algorithm to $\underline{f}$, we actually obtain a limit $T_\ell = 1.7\,10^{-6} = 3.1\,10^{-9}T_{\max}$. This threshold corresponds to a relative mean square error $\mathcal{E}(T_\ell) = 4.7\,10^{-14}$ which is negligible. The corresponding number of iterations was $n_\ell = 21$.

This behavior essentially comes from the piece-wise regularity of $\underline{f}$ which implies the sparsity of the wavelet representation of $\underline{f}$. As a result, most of the wavelet coefficients $\tilde{f}_\lambda$ have values concentrated close to zero. Therefore, the moment of inertia $\left(I_{\underline{f},N}(T)\right)^2$ of the empirical histogram of the coefficients smaller than the threshold T increases slower than T^2.

Consequently, the signal $\underline{f}$ is invariant with respect to the recursive denoising process $\mathcal{A}$ and one has :

$$\mathcal{A}(\underline{f}) = \underline{f}$$

Hence, similarly to what has been shown analytically in section 3.3 for the noise $\underline{W}$, the recursive algorithm was able to perfectly identify $\underline{f}$ as being signal without any noise.

4.1.2 Application to Gaussian white noise

This section validates numerically the conclusions of the theoretical study of section 3.3 where the recursive algorithm was applied to a Gaussian white noise.

We compute the iteration function $I_{\underline{W},N}$ corresponding to a realization $\underline{W}$ of a Gaussian white noise of size $N = 8192$ with a variance $\sigma_W^2 = 1$ provided by a standard random number generator. The graph of $I_{\underline{W},N}$ is displayed in Fig. 4.

We first notice that this graph shows the piece-wise constant nature of the iteration function. Note that this characteristic is more visible for small threshold values. We also observe that the iteration function $I_{\underline{W},N}$ presents two fixed points. The right intersection point of $I_{\underline{W},N}$ and $y = x$ denoted by **A** corresponds exactly to the abscissa $T_D = 4.24$. For values T_n greater than T_D, the curve is flat because $T_D = T_0$ is the maximum value of $I_{\underline{W},N}$. Conversely, as mentioned in section 3.3, the graph of $I_{\underline{W},N}$ bumps over $y = x$ on the left side of point **A**. This is a consequence of $(I_{\underline{W},N})^2$ being the second order moment of the histogram of $\tilde{W}_\lambda$. Due to the fast decay of the Gaussian function, the derivative of $I_{\underline{W},N}$ on the left side of T_D is almost zero and $I_{\underline{W},N}$ is nearly horizontal there.

As expected from these observations the recursive algorithm reaches convergence at the first step. Hence one has $n = n_\ell = 1$ with the threshold $T_\ell = 4.240$ being almost equal to $T_D = 4.245$. Using this threshold, we found that only one of the $N = 8192$ coefficients whose value $\tilde{W}_{\lambda_0} = 4.29$ is larger than the threshold $T_\ell = 4.24$. This yields an almost perfect estimation of $\underline{W}$ by $\mathcal{A}(\underline{X}) = F^c_{T_\ell = T_D}(\underline{W}) \simeq \underline{W}$.

As shown in section 3.3 and as previously observed for signal $\underline{f}$, $\underline{W}$ is invariant with respect to the recursive denoising process $\mathcal{A}$.

In the following, we consider the situation that arises when noise $\underline{W}$ is added to the signal $\underline{f}$.

4.1.3 Application to the signal plus noise

In this case, we apply the recursive algorithm to $\underline{X} = f + W$. We first study the results obtained with $\sigma_f = 10$ and $\sigma_W = 1$, which correspond to signal to noise ratio is equal to 20 db. The number of points is still $N = 8192$. Fig. 5 summarizes the iteration curves computed for $\underline{X}$, $\underline{f}$, and $\underline{W}$. One observes that $I_{\underline{X},N}$ is superposed on $I_{\underline{W},N}$ for small values of T_n whereas it follows $I_{\underline{f},N}$ for large values of T_n, up to the point **C** corresponding to the first iteration of the algorithm.

An explanation of this behavior can be found by looking at the histograms of the wavelet coefficients of $\underline{X}$, $\underline{f}$ and $\underline{W}$ which are respectively related to the iteration functions $I_{\underline{X},N}$, $I_{\underline{f},N}$ and $I_{\underline{W},N}$.
Fig. 6 shows the histograms of the wavelet coefficients corresponding to Fig. 5. As explained in section 4.1.1, the sparsity of the wavelet representation of $\underline{f}$ causes most coefficients $(\tilde{f}_\lambda)$ to be close to zero. Therefore $I_{\underline{f},N}$ remains below the line $y = x$.

One also observes that the histogram of $(\tilde{X}_\lambda)$ and the histogram of $(\tilde{f}_\lambda)$ present the same heavy tails for values larger than the maximum magnitude of the noise $T_D = (2\ln(N))^{1/2}\sigma_W = 4.24$ (cf. section 4.1.2). This coincides with the fact that $I_{\underline{X},N}$ superposes upon $I_{\underline{f},N}$ for values larger than T_D. An interpretation of this superposition is that the heavy tails of the PDF of $(\tilde{f}_\lambda)$ have a strong weight in the second order moment of the histogram of the coefficients $(\tilde{X}_\lambda)$. On the contrary, the coefficients of the noise being concentrated within the range $[-T_D, T_D]$, their contribution to $I_{\underline{X},N}(T)$ for T larger than T_D remains negligible.

At the opposite, when T is smaller than T_D, most of the coefficients $(\tilde{f}_\lambda)$ smaller than T have their value close to zero. Therefore their contribution to the moment of inertia $\left(I_{\underline{X},N}(T)\right)^2$ is dominated by the contribution of the coefficients $(\tilde{W}_\lambda)$ whose distribution far from zero is wider. Thus the noise $\underline{W}$ dominates $\underline{f}$ in the graph of $I_{\underline{X},N}$ for small T. This is still true for T approaching T_D as soon as $\underline{f}$ is sparse enough in wavelet coefficients space.

The consequence is that the intersection **B** of $I_{\underline{X},N}$ with $y = x$ remains close to the intersection **A** of $I_{\underline{W},N}$ with $y = x$. Therefore, the limit T_ℓ of the recursive algorithm applied to $\underline{X}$ is close to the limit obtained for the noise alone which is equal to T_D.

This is true since no other fixed point is present for larger values of the threshold T, thanks to the fact that between **B** and **C**, $I_{\underline{X},N}$ is below $y = x$.

For the considered signal to noise ration of $20db$, the algorithm converges to the value $T_\ell = 4.30$ which is close to the universal threshold $T_D = 4.24$. The resulting estimations $F_{T_\ell}(\underline{X})$ and $F^c_{T_\ell}(\underline{X})$ of $\underline{f}$ and $\underline{W}$, respectively, are shown in Fig. 2.

The iteration functions results for signal to noise ratios $SNR = 10$, $SNR = 40$ and $SNR = 100$ are shown on Fig. 7.

We observe that despite the wide range of different SNR, the global behavior of $I_{\underline{X},N}$ is preserved. The iteration function for the total signal is still superimposed upon $I_{\underline{W},N}$ for threshold values smaller that T_D and switches to $I_{\underline{f},N}$ for larger values of the threshold T. The thresholds T_ℓ obtained are summarized in the table 1. It summarizes the results of wavelet thresholding using different thresholds. We tested the threshold T_ℓ given by the recursive algorithm, the universal threshold T_D and the median based threshold T_m. The values of T_ℓ, T_m and T_D are displayed for different SNR along with the resulting mean square errors defined in (10) of the estimations $\mathcal{E}(T_\ell)$, $\mathcal{E}(T_m)$, $\mathcal{E}(T_D)$ for the $1D$ signal of Fig. 2. These errors are related to enhanced signal to noise ratios resulting of the denoising and defined as

$$SNR'(T) = 20\log_{10}\left(\frac{\|f\|_{L_2}}{\|f - F_T(X)\|_{L_2}}\right) = -10\log_{10}(\mathcal{E}(T)).$$

One can observe that T_ℓ gets closer to T_D as SNR increases.
We remark that the fact that we obtain an better estimation of the level of the noise $\sigma_W = \frac{T_D}{(2\ln(N))^{1/2}}$ for a weaker contribution of the noise to the signal $\underline{X}$ might sound counterintuitive. This phenomena comes from the fact that the value of $I_{\underline{X},N}(T)$ is insensitive to values of $(\tilde{X}_\lambda)$ larger than T. As explained above, due to the sparsity of the wavelet coefficients $(\tilde{f}_\lambda)$, the values of $(\tilde{X}_\lambda)$ weaker than the threshold T_D are strongly dominated by the coefficients of the noise. As a consequence the influence of strong signal to noise ratios on $I_{\underline{X},N}(T)$ is hidden for T being up to T_D. However, this depends on the distribution of the wavelet coefficients of $\underline{f}$ that are smaller than T_D. Therefore, the evolution of the shift observed between T_ℓ and T_D for different SNR can be different for other types of signal.

We also observe that despite the fact that the threshold T_m is usually closer to T_D than T_ℓ and the error $\mathcal{E}(T_m)$ is smaller than the error $\mathcal{E}(T_\ell)$, the performances of the two methods are of same order. Moreover, for all SNR, the threshold T_D results in a larger error than the thresholds T_m and T_ℓ. This surprising result shows the non-optimality of the universal threshold as well as the good performance of both approaches using T_m and T_ℓ, especially for weak signal to noise ratios. However, for increasing SNR, the performance of the estimation with each method gets more uniform.

We also observed that the number of iterations n_ℓ increases with the signal to noise ratio. This is consistent with the fact that only one iteration is

SNR	$-\infty$	10	20	40	100	$+\infty$
σ_f	0	$10^{\frac{1}{2}}$	10	10^2	10^5	$\sigma_W = 0$ $\sigma_f = 10$
n_ℓ	0	4	5	7	12	21
T_ℓ	4.24	4.34	4.30	4.30	4.23	$1.7\,10^{-6}$
T_m	4.19	4.19	4.20	4.20	4.24	$9.9\,10^{-7}$
T_D	4.24	4.25	4.25	4.25	4.25	0
$\mathcal{E}(T_\ell)$	$+\infty$	$7.28\,10^{-3}$	$6.46\,10^{-4}$	$9.72\,10^{-6}$	$2.04\,10^{-11}$	$4.7\,10^{-14}$
$\mathcal{E}(T_m)$	$+\infty$	$7.06\,10^{-3}$	$6.36\,10^{-4}$	$9.77\,10^{-6}$	$2.04\,10^{-11}$	$8.9\,10^{-16}$
$\mathcal{E}(T_D)$	$+\infty$	$7.32\,10^{-3}$	$6.68\,10^{-4}$	$9.77\,10^{-6}$	$2.04\,10^{-11}$	0
$SNR'(T_\ell)$	$-\infty$	21.37	31.90	50.12	106.9	133.3
$SNR'(T_m)$	$-\infty$	21.51	31.96	50.10	106.9	150.5
$SNR'(T_D)$	$-\infty$	21.35	31.75	50.10	106.9	$+\infty$
Flat. of $F^c_{T_\ell}(\underline{X})$	3.05	3.08	3.03	3.08	3.14	5.00
Flat. of $F^c_{T_m}(\underline{X})$	3.05	3.08	3.03	3.08	3.14	4.28
Flat. of $F^c_{T_D}(\underline{X})$	3.05	3.08	3.03	3.08	3.14	undefined

Table 1: Numerical results for the 1D signal : , number of iterations before convergence n_ℓ, thresholds T_ℓ, T_m and T_D, the relative errors $\mathcal{E}(T_\ell)$, $\mathcal{E}(T_m)$ and $\mathcal{E}(T_D)$ and the corresponding enhanced signal to noise ratios SNR' obtained for $SNR = 10\,db$, $20\,db$, $40\,db$ and $100\,db$ and the two limit cases $SNR = \pm\infty$. For all lines columns, $\sigma_W = 1$, except for the last one (c.f. section 4.1.1). The flatness of the estimated noises $F^c_{T_\ell}(\underline{X})$, $F^c_{T_m}(\underline{X})$ and $F^c_{T_D}(\underline{X})$ is also given

needed for the noise alone (which corresponds to $SNR = -\infty$) and that the maximum number of iterations is obtained for the signal without added noise (which corresponds to $SNR = +\infty$).

We interpret this result by saying that the wavelet coefficients of the noise are responsible for deflecting the graph of $I_{\underline{X},N}$ above the line $y = x$. This deflection interrupts the sequence of iteration by forcing the decreasing sequence of thresholds T_n to converge to the intersection point **B**. The stronger the noise level, the sooner the deflection and the convergence.

The histograms of the estimated signals $F_{T_\ell}(\underline{X})$, $F_{T_m}(\underline{X})$ obtained using the thresholds T_ℓ and T_m, and the histogram of the corresponding estimations of the noise $F^c_{T_\ell}(\underline{X})$, $F^c_{T_m}(\underline{X})$ are shown on Fig. 9. One observes that they are very well superposed. This confirms that for this academic case, the performance of the recursive algorithm is comparable to the one of the median based wavelet thresholding.

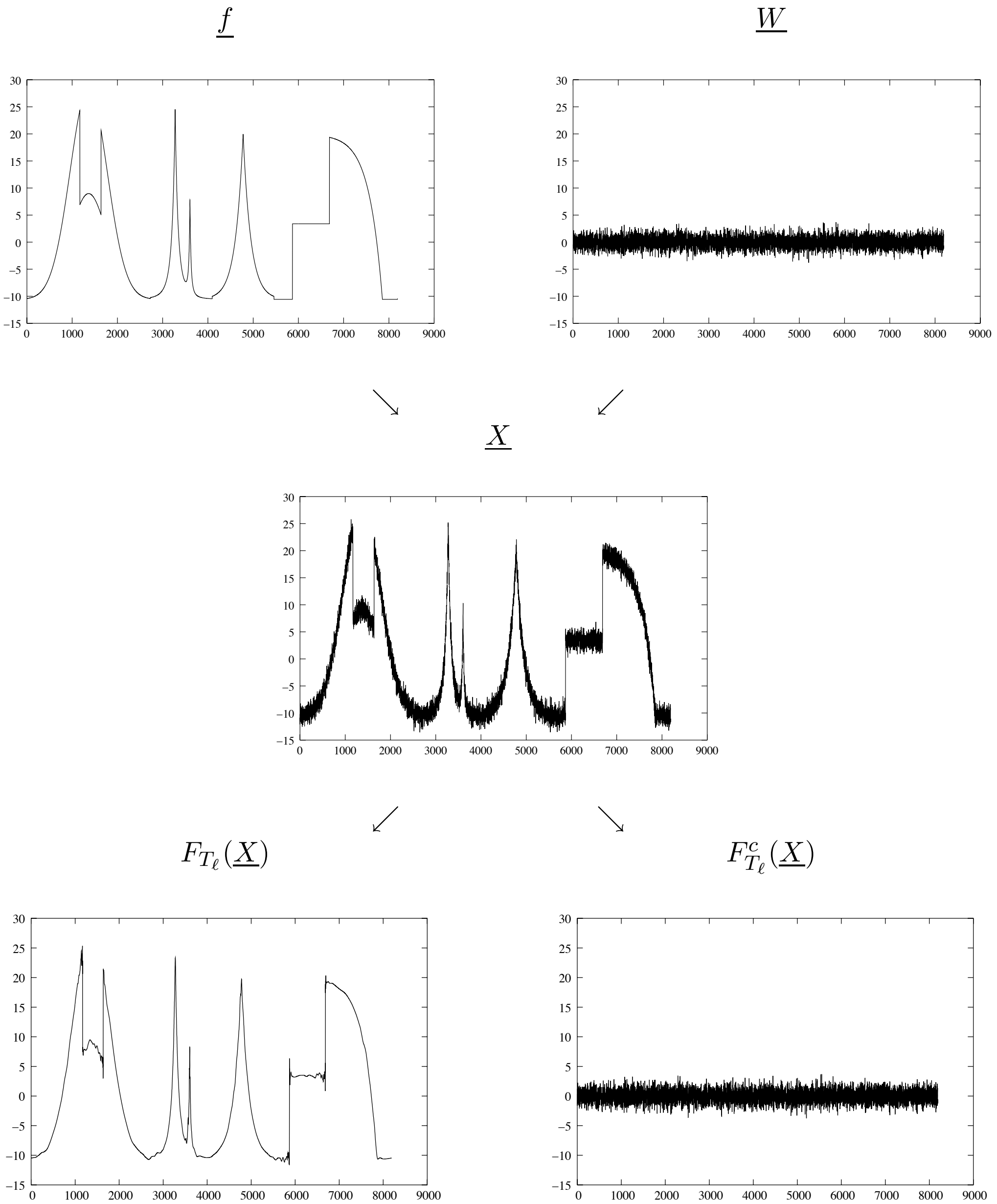

Figure 2: Construction of a 1D noisy signal $\underline{X} = \underline{f} + \underline{W}(SNR = 20\,db)$, and its nonlinear wavelet thresholding using the threshold T_ℓ.

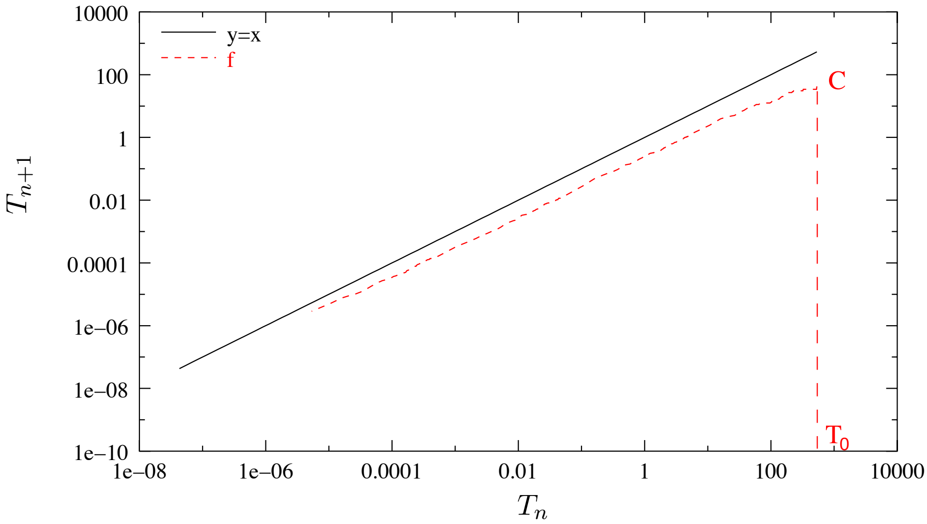

Figure 3: The graph of the iteration function $I_{\underline{f},N}(T)$ for the signal (Fig. 2 top, left). The point **C** corresponds to the first iteration of the algorithm.

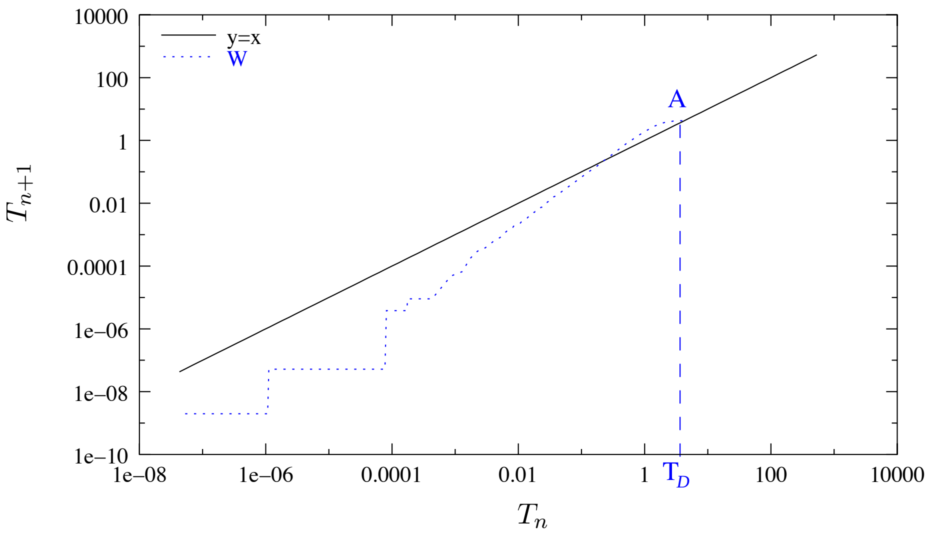

Figure 4: The graph of $I_{\underline{W},N}$, when the process $\underline{W}$ is Gaussian white noise. The point **A** corresponds to the right hand intersection between the graphs of $I_{\underline{W},N}$ and $y = x$.

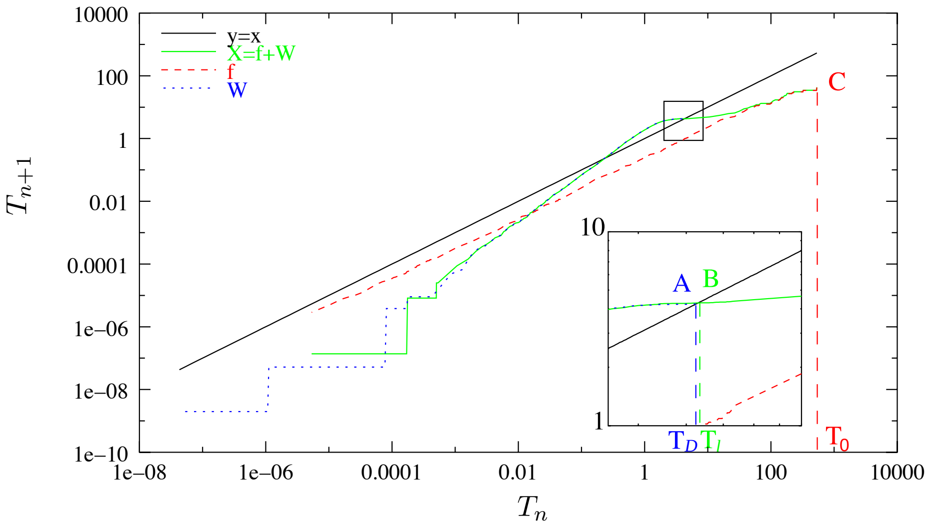

Figure 5: Analysis of the 1D signal with $SNR = 20\ db$: the iteration functions $I_{\underline{W},N}$,$I_{\underline{f},N}$,$I_{\underline{X},N}$ for $\underline{W}$ $\underline{f}$ and $\underline{X}$ respectively. The points **A** and **B** correspond to the intersections between the graphs of $I_{\underline{W},N}$ 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 total signal $\underline{X}$ and its abscissa is T_0.

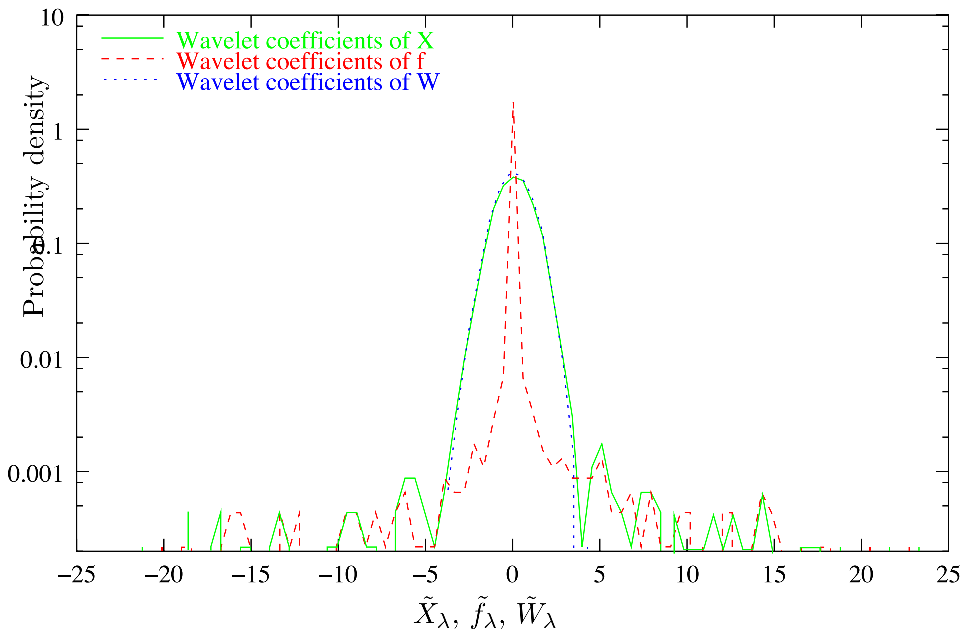

Figure 6: Histograms of the wavelet coefficients $\tilde{X}_\lambda$, $\tilde{f}_\lambda$, and $\tilde{W}_\lambda$ for the 1D signal with $SNR = 20\ db$

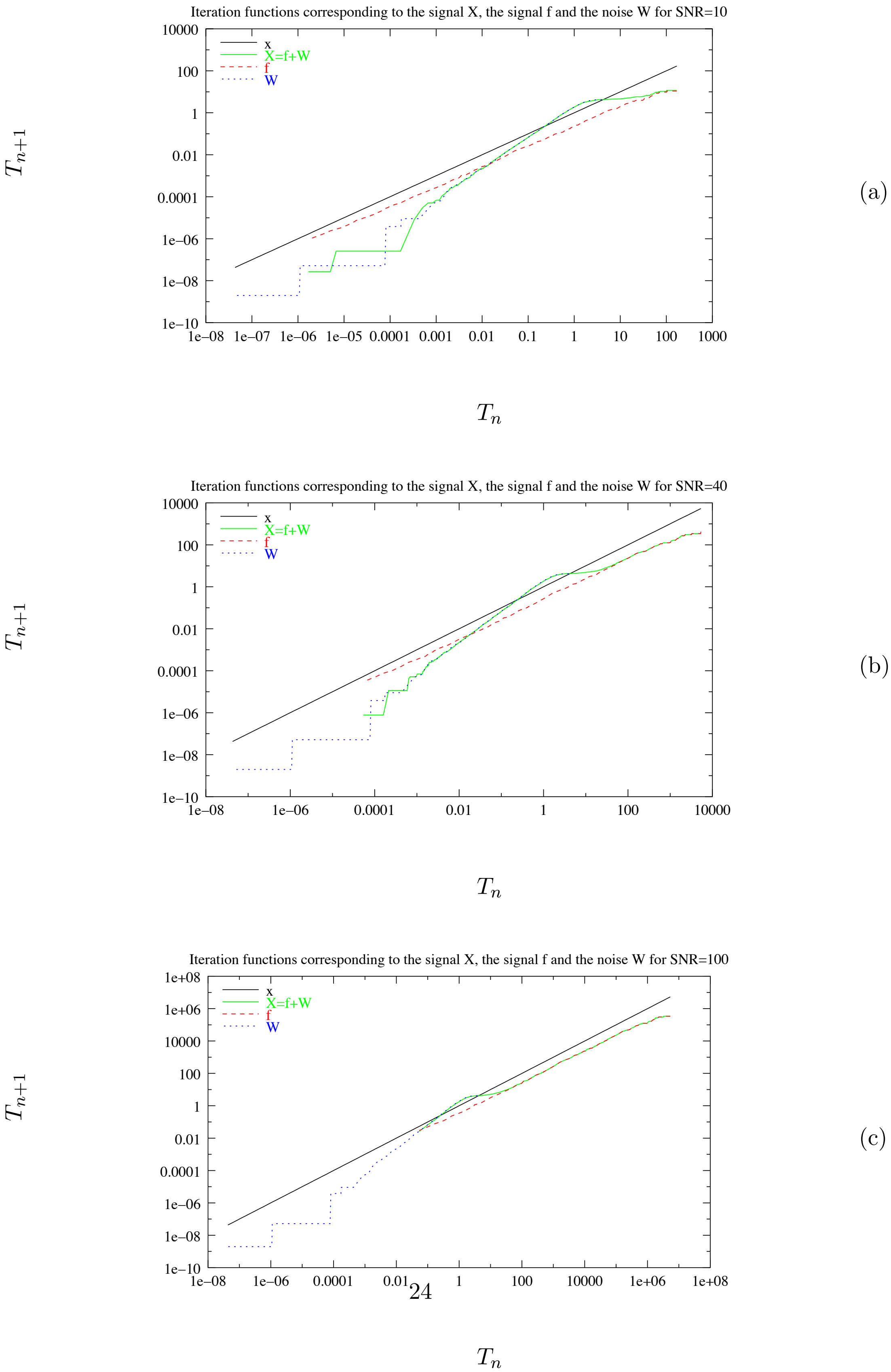

Figure 7: Iteration functions $I_{\underline{X}=\underline{f}+\underline{W},N}$, $I_{\underline{f},N}$ and $I_{\underline{W},N}$ for $\sigma_W = 1$ and σ_f taken successively equal to $10^{1/2}$ (plot (a), $SNR = 10$ db), 10^2 (plot (b) $SNR = 40$ db) and 10^5 (plot (c) $SNR = 100$ db).

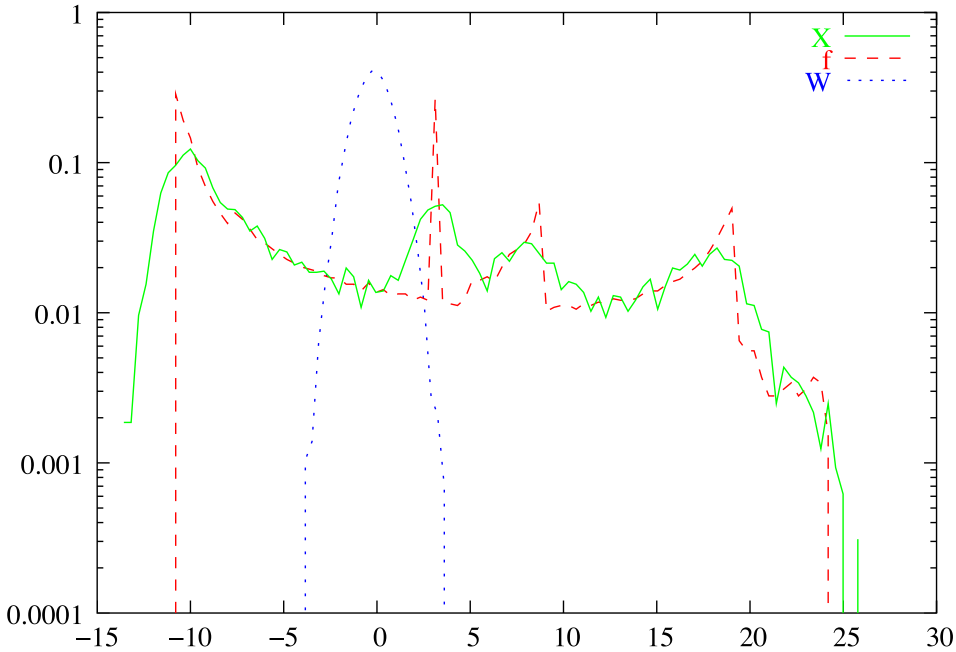

Figure 8: Histograms of signal $\underline{f}$, the noisy signal $\underline{X}$ and the noise $\underline{W}$ for the 1D signal with $SNR = 20\ db$.

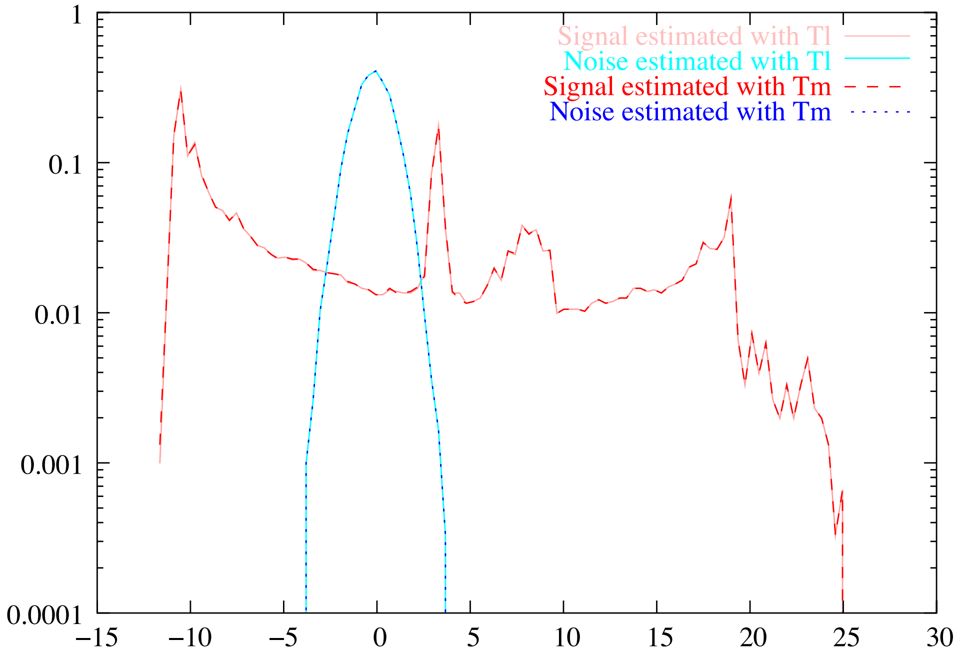

Figure 9: Histograms of the estimated signal and noise using thresholds T_ℓ and T_m for the 1D signal with $SNR = 20\ db$.

4.1.4 Application to a 2D test signal

In this section, we present the results obtained by applying the recursive algorithm to a 2D discrete signal $\underline{X}$ sampled on $N = 512^2$ grid points. The signal $\underline{X}$ is obtained by the superposition of a 2D signal $\underline{f}$ composed of randomly located axisymmetric cusps, and a 2D realization $\underline{W}$ of a Gaussian white noise.

Surface plots of the signals $\underline{X}$, $\underline{f}$ and $\underline{W}$ are shown in Fig. 10 (top and center). We chose $\sigma_f = 3.163$ and $\sigma_W = 1$ resulting in $SNR = 10$.

We compute the corresponding iteration functions $I_{\underline{X},N}$, $I_{\underline{f},N}$ and $I_{\underline{W},N}$. They are displayed on Fig. 11.

We first remark that conversely to the 1D case, the curve $I_{\underline{f},N}$ corresponding to the signal alone intersects the line $y = x$ for a small value of the threshold between $T = 10^{-4}$ and $T = 10^{-3}$. We interpret this as a consequence of a less sparse wavelet representation of the signal $\underline{f}$. These more numerous weak wavelet coefficients of $\underline{f}$ may be due to the fact that in such a 2D signal, axisymmetric singularities are difficult to capture for the 2D orthogonal wavelets of a tensor product multi-resolution analysis. For these reasons, more wavelet coefficients of little energy are required in order to represent $\underline{f}$ correctly.

However, the most striking feature of Fig. 11 is that similarly to the 1D case, the wavelet coefficients of the noise are responsible for deflecting the graph of $I_{\underline{X},N}$ from $I_{\underline{f},N}$ to $I_{\underline{W},N}$ when the threshold value T decreases. We observe that it intersects the line $y = x$ very close to the intersection between $I_{\underline{W},N}$ and $y = x$. This shows that when applied to $\underline{X}$, the recursive algorithm converges to a limit threshold value T_ℓ close to T_D.

When applying the algorithm, we find a limit $T_\ell = 5.13 \simeq T_D = (2\ln(512^2))^{1/2} = 4.99$. The resulting estimations $F_{T_\ell}(\underline{X})$ and $F^c_{T_\ell}(\underline{X})$ for $\underline{f}$ and $\underline{W}$ are shown on Fig. 10 (bottom). As expected, the estimation $F_{T_\ell}(\underline{X})$ shows an efficiently denoised version of the signal $\underline{X}$. The corresponding numerical results are summarized in table 2. As in the 1D case, the results using T_ℓ, T_m and T_D are very close. This shows the validity of the recursive algorithm for estimating the suitable threshold value. This is confirmed when looking at the histograms of Fig. 13 and Fig. 14. The results of the estimations using T_ℓ and T_m are almost perfectly superposed. Fig. 12 shows the PDF of the wavelet coefficients of $\underline{X}$, $\underline{f}$, and $\underline{W}$. One can see that the threshold value $T_\ell = 5.13$ corresponds to the value for which the PDF of $\tilde{X}_\lambda$ moves from the PDF of $\tilde{f}_\lambda$ to the one of $\tilde{W}_\lambda$.

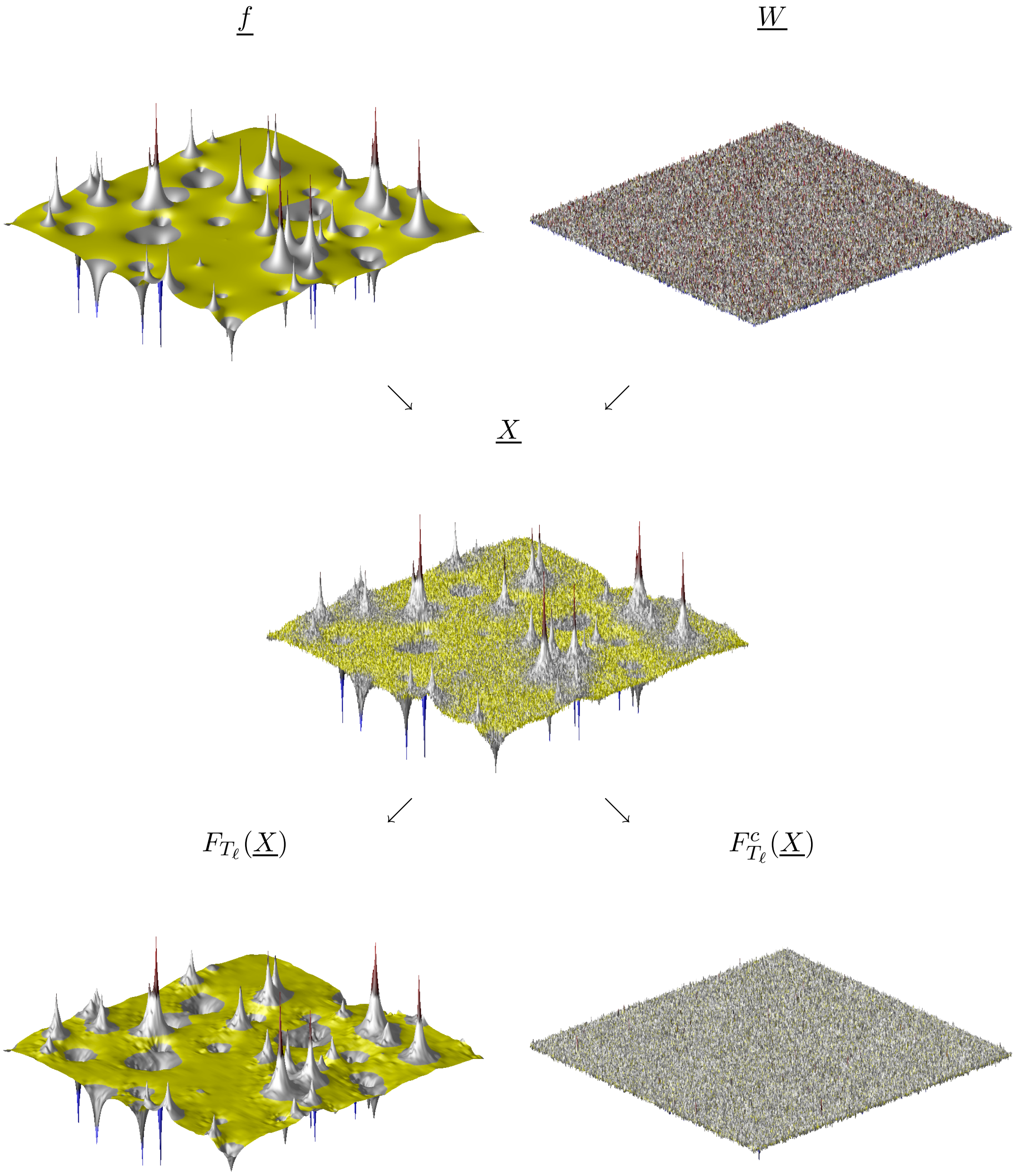

Figure 10: Construction of a 2D noisy signal $\underline{X} = \underline{f} + \underline{W}$ ($SNR = 10\,db$), and its nonlinear wavelet thresholding using the threshold T_ℓ.

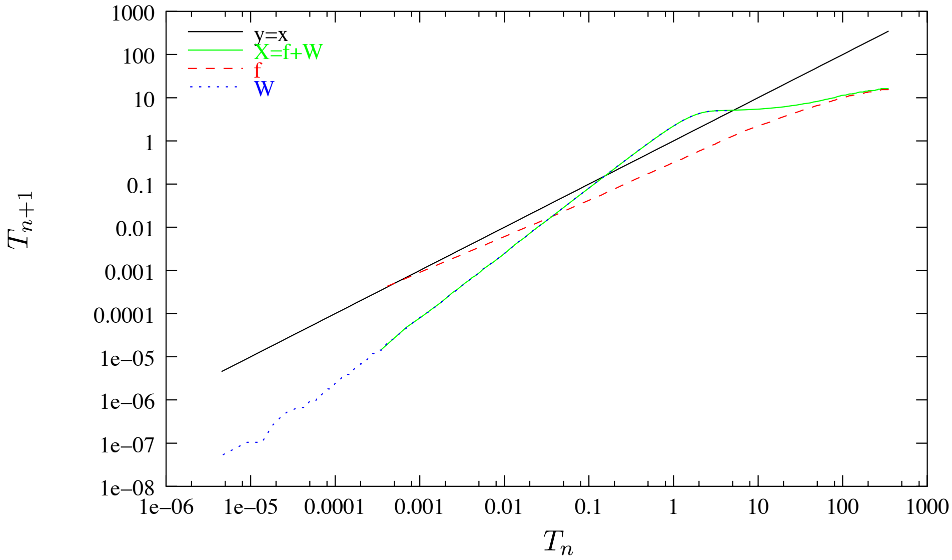

Figure 11: Iteration functions $I_{\underline{X},N}$, $I_{\underline{f},N}$ and $I_{\underline{W},N}$corresponding to the 2D signals $\underline{X}$, $\underline{f}$ and to the noise $\underline{W}$.

SNR	10
σ_f	$10^{\frac{1}{2}}$
n_ℓ	4
T_ℓ	5.13
T_m	5.01
T_D	4.99
$\mathcal{E}(T_\ell)$	$7.57\,10^{-3}$
$\mathcal{E}(T_m)$	$7.34\,10^{-3}$
$\mathcal{E}(T_D)$	$7.28\,10^{-3}$
$SNR'(T_\ell)$	21.20
$SNR'(T_m)$	21.34
$SNR'(T_D)$	21.37
Flat. of $F^c_{T_\ell}(\underline{X})$	3.32
Flat. of $F^c_{T_m}(\underline{X})$	3.30
Flat. of $F^c_{T_D}(\underline{X})$	3.30

Table 2: Thresholds T_ℓ, T_m and T_D, the relative errors $\mathcal{E}(T_\ell)$, $\mathcal{E}(T_m)$ and $\mathcal{E}(T_D)$ and the corresponding enhanced signal to noise ratios SNR' obtained for $SNR = 10\,db$, with $\sigma_W = 1$.

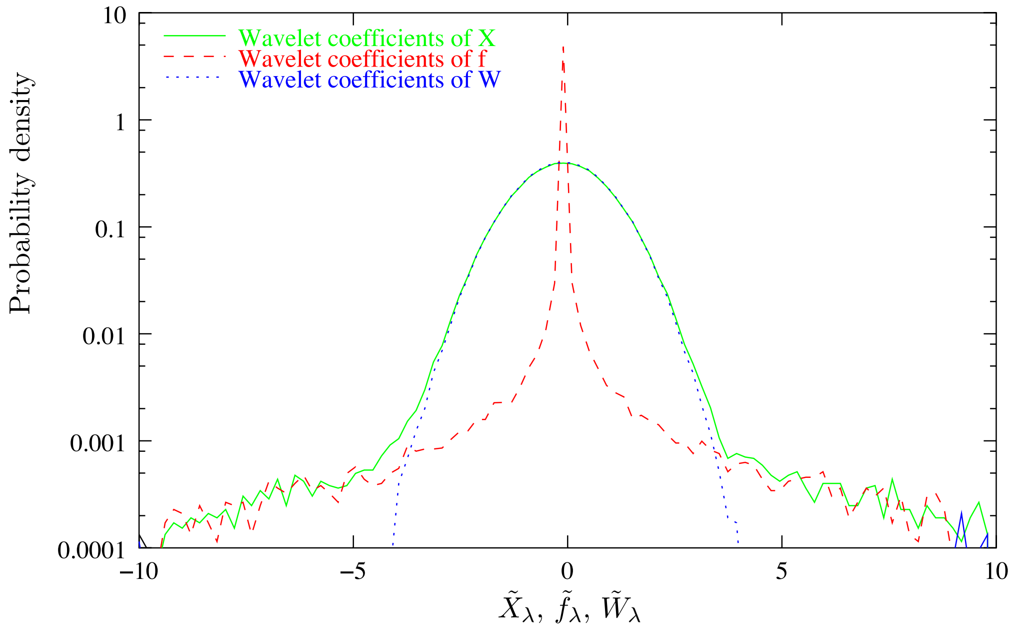

Figure 12: Histograms of wavelet coefficients $\tilde{X}_\lambda$, $\tilde{f}_\lambda$, $\tilde{W}_\lambda$ for the 2D case with$SNR = 10\ db$.

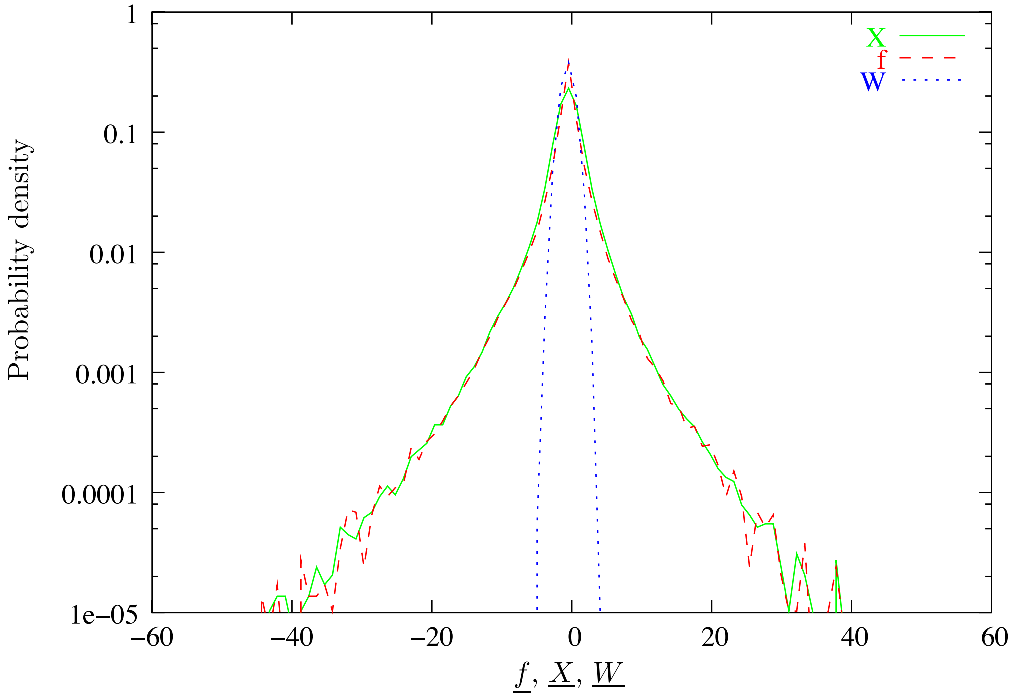

Figure 13: Histograms of the 2D signal $\underline{f}$, noisy signal $\underline{X}$ and noise $\underline{W}$, with $SNR = 10\ db$.

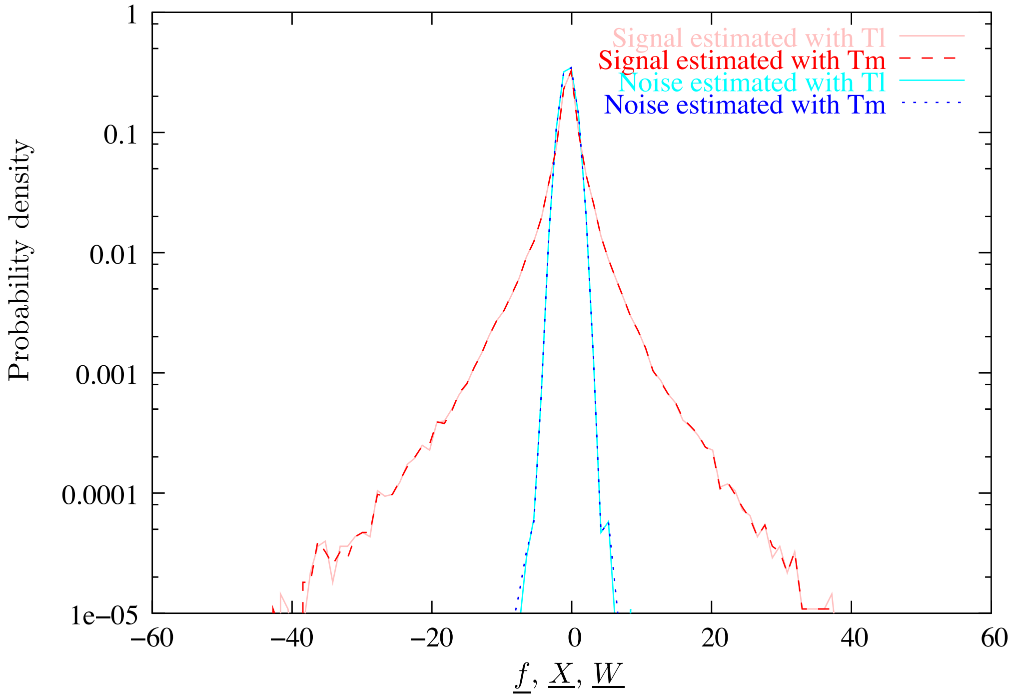

Figure 14: Histograms of estimated signals $F_{T_\ell}(\underline{X})$ and $F_{T_m}(\underline{X})$, and the corresponding noises $F^c_{T_\ell}(\underline{X})$ and $F^c_{T_m}(\underline{X})$ ($SNR = 10\ db$).

Method	recursive	median
Threshold	$T_\ell = 0.32$	$T_m = 0.19$
Retained coefficients	0.24%	1.41%
Retained variance	35%	98%
Estimated signal to noise ratio $\mathcal{SNR}$	-2.88	18.41
Flatness of the estimated noise	3.66	3.16
n_ℓ	2	

Table 3: Numerical results for the experimental signal

4.2 Application to a Bose-Einstein condensate

4.2.1 Description of the signal

The signal presented in this section is an absorption image measuring the density of atoms in a Bose-Einstein condensate [8]. It was obtained using a CCD camera for measuring the optical density of lithium atoms confined in a magnetic trap and cooled by evaporation using a microwave field. Its surface plot is shown in Fig. 15 (top). The number of sample points is $N = 128^2$.

We apply the recursive wavelet thresholding algorithm to remove the strong noise observed in the signal and we compare the results obtained using both thresholds T_ℓ and T_m. As the signal and noise components are unknown, we check the quality of the estimation by looking at the estimated fields in physical space, at their histograms computed both in physical and wavelet space and their Fourier power spectra.

4.2.2 Numerical results

Table 3 presents the numerical results obtained. We define an estimated SNR

$$\mathcal{SNR} = 10 \log_{10} \frac{\|F_T(\underline{X})\|^2}{\|F_T^c(\underline{X})\|^2}$$

corresponding to the estimated signal to the estimated noise ratio. Therefore, it does not quantifies the performance of the denoising. The quality of the filtering is given by the agreement between the estimated SNR and the real SNR that has to be estimated by some other means. Note that $\mathcal{SNR}$ depends on the filtering method used.

We observe a large difference between T_ℓ and T_m as well as between the proportion of coefficients and variance retained by either the recursive or the median based method. The $\mathcal{SNR}$ obtained is very different according to the threshold used. Looking at the noisy signal in physical space (see Fig. 15 and Fig. 15) make us conjecture that $\mathcal{SNR} = 18.41$ does not correspond to the real SNR. The flatness of the estimated noise for both methods is close to 3, which is the flatness of a Gaussian white noise. The noise estimated by the median based method is thus closer to the Gaussianity.

4.2.3 Results in physical space

One observes in Fig. 15(middle) that the recursive algorithm extracts the signal from the noise very well. On the contrary, Fig. 15 (bottom) shows that the median based threshold T_m is too small for removing all parts of the noise. This residual noise is therefore still present in the estimated signal $F_{T_m}(\underline{X})$ which explains the large percentage of retained energy that results in the value of $\mathcal{SNR} = 18.41$.

4.2.4 Histograms of the wavelet coefficients

Figures 16(top) and 16(bottom) show the histogram of the wavelet coefficients of the signal superposed on the histograms obtained for the estimated signal and the estimated noise for both methods. One can observe the effect of the thresholding on these histograms which are simply separated by the value corresponding to the thresholds T_ℓ and T_m. The vertical line at the value zero corresponds to the coefficients set to zero after thresholding. One observes that the threshold T_ℓ corresponds to the point where the histogram of the noisy signal changes from a Gaussian-like shape to the more irregular shape of large non-Gaussian tails. On the contrary, the median based threshold T_m corresponds to an intermediate value of the Gaussian part of the histogram. A comparison with the fitted Gaussian curve shows, however, that the fit is not perfect. Taking the median of the modulus of the weak wavelet coefficients responsible for this part of the histogram could therefore yield an incorrect estimation of the variance of the noise.

4.2.5 Histograms computed in physical space

The histograms of the noisy signal $\underline{X}$ and the estimated signal and noise are presented for both methods in Fig. 17(top) and Fig. 17(bottom). One observes that the tails of the histogram of the estimated signal are shorter when using the threshold T_ℓ than when using the threshold T_m. This is consistent with the fact that no spurious noisy oscillation are observed on the denoised field obtained with T_ℓ whereas the signal estimated using T_m shows this type of noisy structures which results in a wider histogram. As a counterpart, the noise looks more Gaussian when extracted using the median

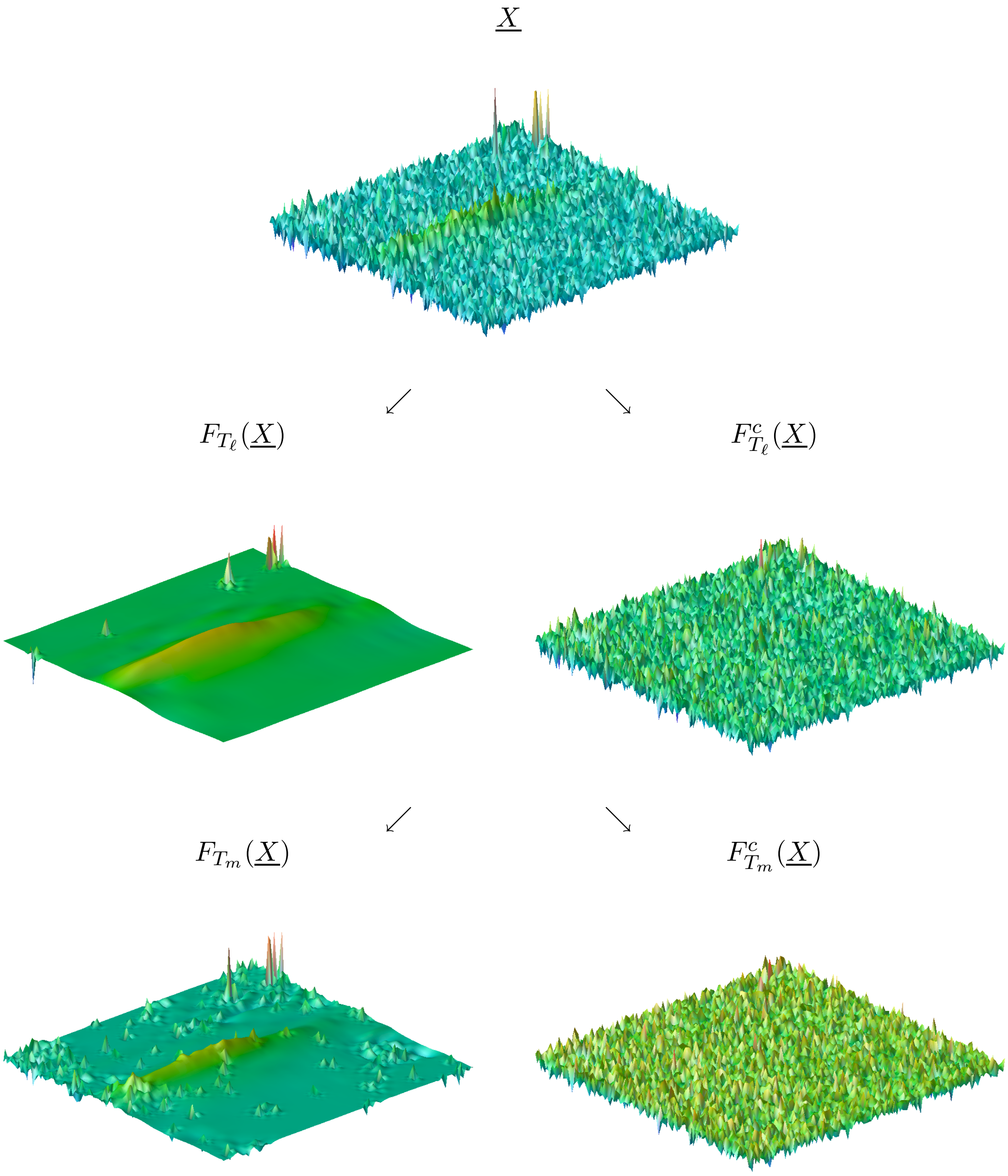

Figure 15: Wavelet filtering of an experimental 2D noisy signal (top) using the threshold T_ℓ found with the recursive algorithm (middle) and the median based threshold T_m (bottom).

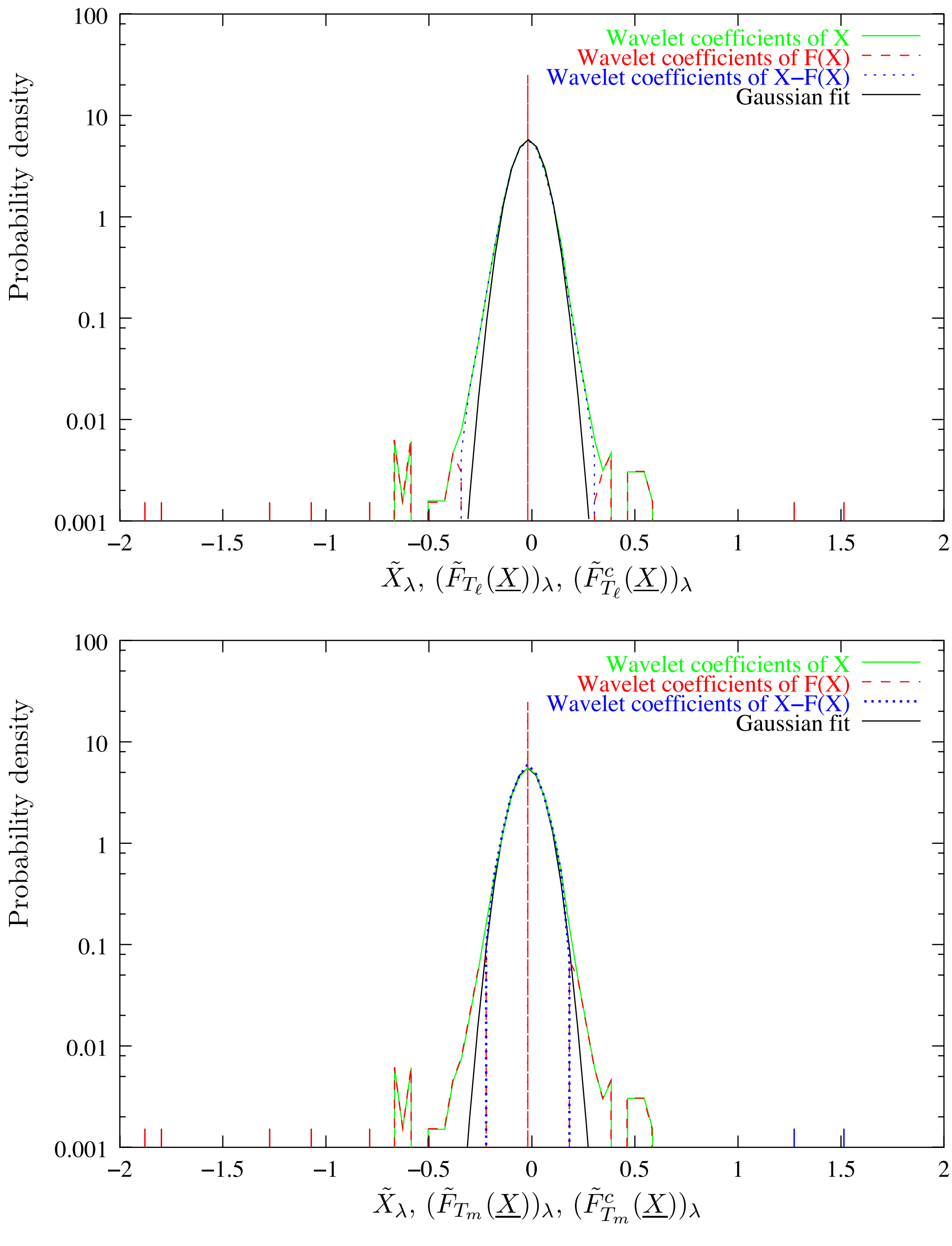

Figure 16: PDFs of the wavelet coefficients $\tilde{X}_\lambda$, $(\tilde{F}_{T_\ell}(\underline{X}))_\lambda$, and $(\tilde{F}^c_{T_\ell}(\underline{X}))_\lambda$ obtained using the recursive algorithm (top) and PDFs of the wavelet coefficients $\tilde{X}_\lambda$, $(\tilde{F}_{T_m}(\underline{X}))_\lambda$ and $(\tilde{F}^c_{T_m}(\underline{X}))_\lambda$ obtained using the median based threshold T_m (bottom)

based threshold. This is related to the value of the flatness of the noise which is closer to 3 for the median method. We also observe that for both methods, the histogram of the estimated signal is not well superposed to the one of the noisy signal. This may be a consequence of the low signal to noise ratio.

4.2.6 Fourier spectra

The isotropic Fourier power spectra of the noisy signal and its two estimated components are shown for both methods in Fig. 18 and Fig. 19. One can observe a very strong signature of the noise as a linear part of the spectrum of $\underline{X}$ at small scales. This linear part exhibits a slope k^{+1}, typical of white noise. We observe that when using the threshold T_m the spectrum of the estimated signal also present such a linear part, revealing the presence of remaining noise. On the contrary, the spectrum of the estimated signal obtained using the recursive algorithm is free from this linear component. Hence, we conclude that the noise has been correctly removed.

4.2.7 Conclusion for the experimental case

The recursive algorithm is more efficient than the median based method for the denoising of this experimental signal. We conjecture that this is a consequence of the not perfectly Gaussian distribution of the noise that leads to a wrong estimation of the threshold computed from the median of the modulus of the wavelet coefficients of the noisy signal. On the contrary, the recursive algorithm, by converging to the fixed point of the iteration function actually converges to the threshold corresponding to the change of behavior in the histogram of the wavelet coefficients of the noisy signal. This convergence is more robust with respect to a not exactly Gaussian distribution of the noise. This robustness explains why the median threshold gives a slightly better result with academic signals for which the Gaussianity of the noise is guaranteed, but fails to remove the noise completely from a real experimental signal.

5 Conclusions

We presented a new recursive algorithm to determine automatically the threshold value of the wavelet coefficients to denoise a signal corrupted by a Gaussian white noise. This efficient and simple algorithm is based on a recursion in wavelet coefficient space to estimate the variance of the noise. A mathematical justification of the scheme has been given by applying fixed point type arguments to guarantee the convergence of the process. We proved that the algorithm is stable and converges with a finite number of iterations bounded from above by the number of samples, but in practice we need very few iterations. Numerical examples in one and two dimensions

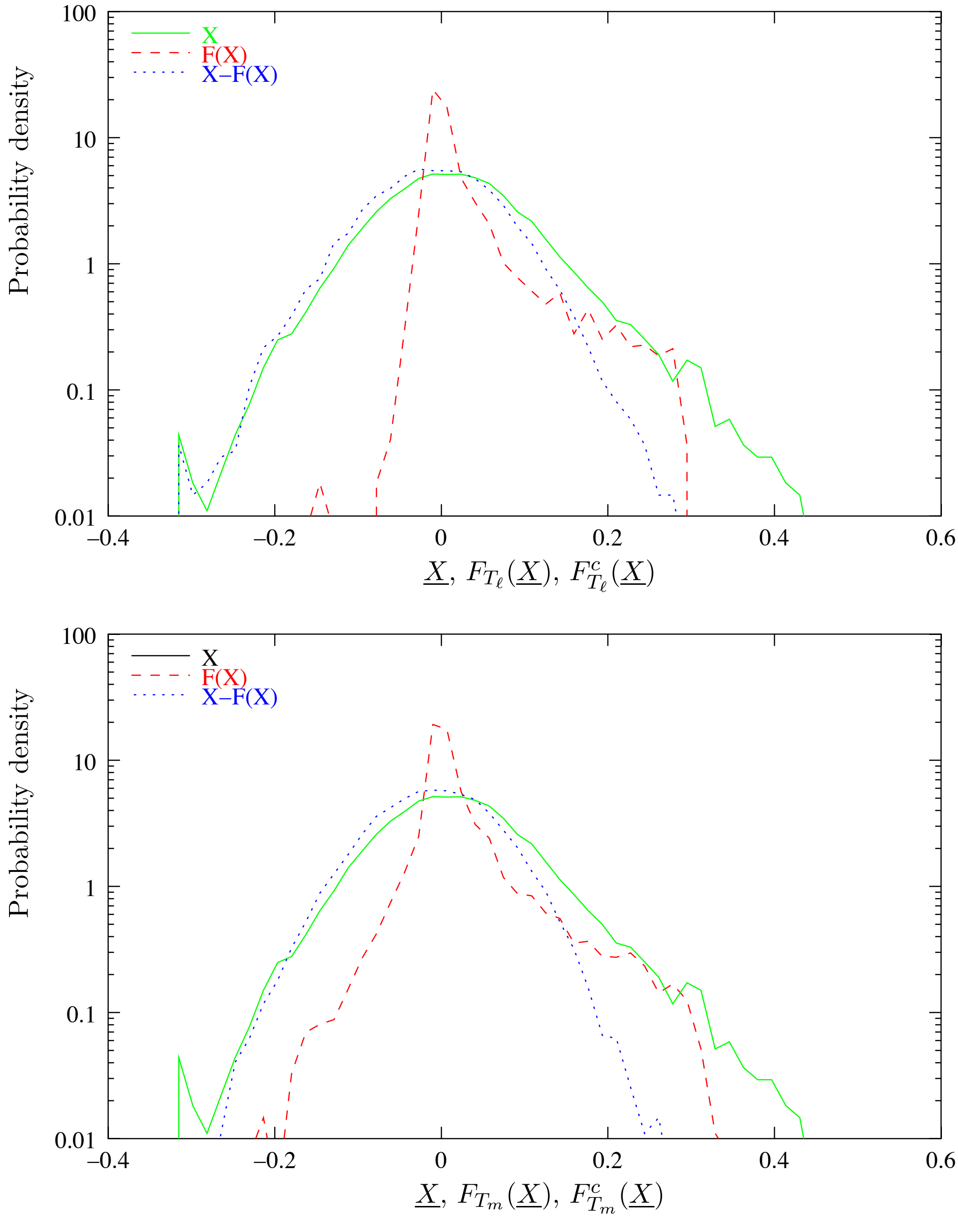

Figure 17: PDFs of $\underline{X}$, $F_{T_\ell}(\underline{X})$ and $F^c_{T_\ell}(\underline{X})$ in physical space obtained using the recursive algorithm (top) and PDFs of $\underline{X}$, $F_{T_m}(\underline{X})$ and $F^c_{T_m}(\underline{X})$ obtained using the median based threshold T_m (bottom)

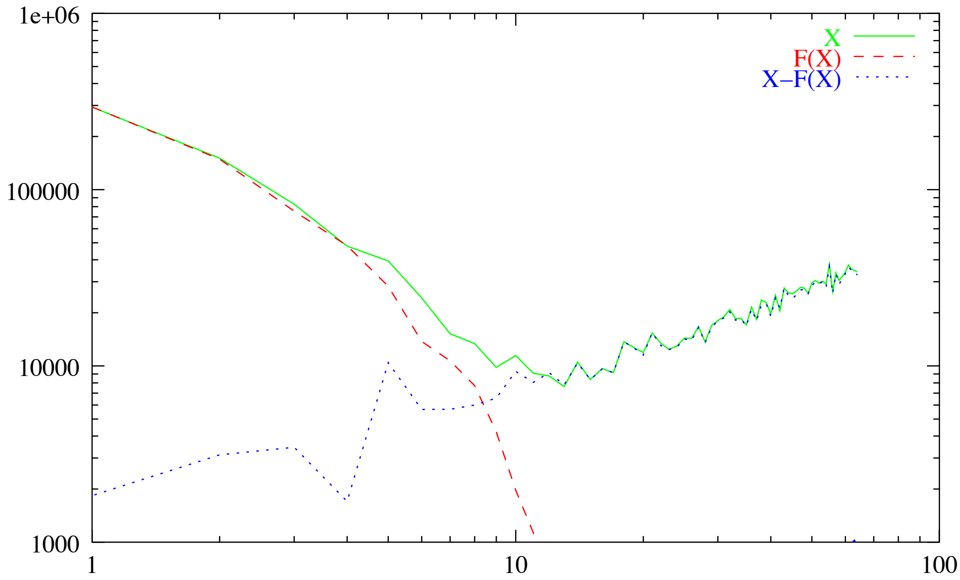

Figure 18: Energy spectra of $\underline{X}$, $F_{T_\ell}(\underline{X})$, $F^c_{T_\ell}(\underline{X})$ obtained using the recursive algorithm.

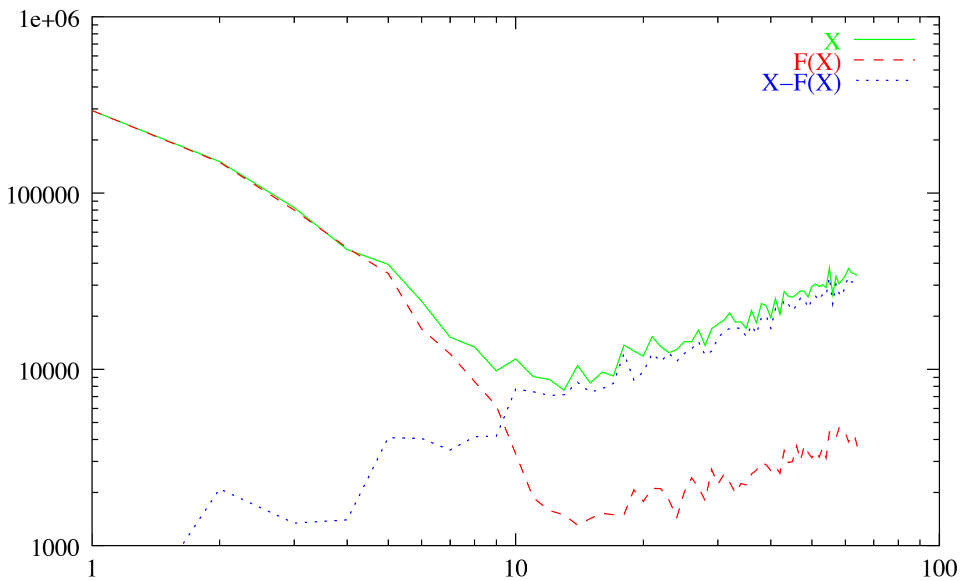

Figure 19: Energy spectra of $\underline{X}$, $F_{T_m}(\underline{X})$, $F^c_{T_m}(\underline{X})$ obtained using the median based threshold T_m.

illustrated the properties and validity of the scheme for different signal to noise ratios. We observed that the number of iterations decreases with the signal to noise ratio. We have some evidence that it actually depends on the sparsity of the wavelet representation of the signal, *i.e.* the more sparse the wavelet representation of the signal, the faster the convergence. For the academic examples studied here we found that the threshold obtained with the algorithm and that the error $\mathcal{E}(T)$ of the corresponding estimation are very close to the ones obtained using the universal threshold proposed by Donoho & Johnstone [2] and the threshold estimated from the median of the wavelet coefficients of the noisy signal at small scales. Moreover, the recursive algorithm gave better results than the median based wavelet thresholding estimator when applied to an experimental signal. This suggests that the algorithm is more robust with respect to a departure from the academic situation of a perfectly Gaussian white noise.

Future work is concerned with the generalization of the algorithm to non Gaussian distributions of the noise and to the case of colored, *i.e.* correlated, noise. We are also extending the approach to vector-valued signals, *e.g.* velocity or vorticity fields, in order to apply it to coherent vortex extraction in turbulent flows.

6 Acknowledgment

We thank F. Schereck, L. Khaykovich, K.L. Corwin, G. Ferrari, T. Bourdel, J. Cubizolles and C. Salomon from the Laboratoire Kastler Brossel of the École Normale Supérieure of Paris, for providing us with the absorption density field of the section 4.2. We acknowledge financial support from the contract between CEA-Euratom and ENS-Paris n^o V.3258.001.

References

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

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

[3] M. Farge, K. Schneider, N. Kevlahan. *Non–Gaussianity and coherent vortex simulation for two–dimensional turbulence using an adaptive orthogonal wavelet basis.* Phys. Fluids, **11**(8), 2187–2201, 1999.

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

[5] S. Mallat. *A wavelet tour of signal processing.* Academic Press, 1998.

[6] 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.

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

[8] F. Schereck, L. Khaykovich, K.L. Corwin, G. Ferrari, T. Bourdel, J. Cubizolles and C. Salomon. Quasipure Bose-Einstein Condensate Immersed in a Fermi Sea *Phys. Rev. Letters*, **87**(8), 2001

[9] Wavelab software written by David Donoho, Mark Reynold Duncan, Xiaoming Huo and Ofer Levi
http://www-stat.stanford.edu/∼wavelab/.