Optimal wavelet shrinkage of a noisy dynamical system with non-linear noise impact

Matthieu Garcin, Dominique Guegan

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Documents de Travail du Centre d'Economie de la Sorbonne

CES Working Papers

Optimal wavelet shrinkage of a noisy dynamical system with non-linear noise impact

Matthieu GARCIN, Dominique GUEGAN

2015.85

cnrs dépasser les frontières

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/
ISSN : 1955-611X

Optimal wavelet shrinkage of a noisy dynamical system with non-linear noise impact*

Matthieu Garcin[†] Dominique Guégan[‡]

February 11, 2015

Abstract

By filtering wavelet coefficients, it is possible to construct a good estimate of a pure signal from noisy data. Especially, for a simple linear noise influence, Donoho and Johnstone (1994) have already defined an optimal filter design in the sense of a good reconstruction of the pure signal. We set here a different framework where the influence of the noise is non-linear. In particular, we propose an optimal method to filter the wavelet coefficients of a discrete dynamical system disrupted by a weak noise, in order to construct good estimates of the pure signal, including Bayes' estimate, minimax estimate, oracular estimate or thresholding estimate. We present the example of a simple chaotic dynamical system as well as an adaptation of our technique in order to show empirically the robustness of the thresholding method in presence of leptokurtic noise. Moreover, we test both the hard and the soft thresholding and also another kind of smoother thresholding which seems to have almost the same reconstruction power as the hard thresholding.

Keywords: wavelets, dynamical systems, chaos, Gaussian noise, Cauchy noise, thresholding, nonequispaced design, non-linear noise impact.

1 Introduction

Donoho and Johnstone (1994) have developed a theory of signal denoising using wavelets [14]. Their optimal filtering method has been used for many applications. However, for some signals, the noise has a non-linear influence and therefore, the classical theory of wavelet-based denoising has to be adapted. This is the aim of the present paper in the particular framework of dynamical systems.

Dynamical systems are used to depict non-linearity by a deterministic way [19, 1]. They differ from other kinds of non-linear models relying on a stochastic description, like heteroskedastic processes [21, 6, 10], jump processes [51] or also long-memory processes [29, 38, 53, 30, 4, 5]. Dynamical systems are particularly relevant in some applications, like in video processing [24], in natural sciences [37] as well as in finance [22, 23, 31].

We consider a dynamical system, defined in discrete time, x_t. Two successive states of that dynamical system are linked by an evolution function z [31]:

$$\forall t \in \{1, ..., T\},\ x_{t+1} = z(x_t). \tag{1}$$

*Submitted paper

[†]Corresponding author, Université Paris 1 Panthéon-Sorbonne, MSE - CES, 106 boulevard de l'hôpital, 75013 Paris, France, and Natixis Asset Management, firstname.lastname [at] polytechnique.edu

[‡]Université Paris 1 Panthéon-Sorbonne, MSE - CES, 106 boulevard de l'hôpital, 75013 Paris, France, firstname.lastname [at] univ-paris1.fr

However, some measurement noise may perturb the observation of that dynamical system [25, 31]. In that case, we do not observe directly the state of the system x_t, but rather a noisy observation u_t, which consists in an alteration of the state x_t by an additive random variable ε_t:

$$\forall t \in \{1, ..., T\},\ u_t = x_t + \varepsilon_t,$$

where $\varepsilon_1, ..., \varepsilon_T$ are independent identically distributed random variables. Therefore, the observed evolution function is not z but the function z^ε that links two successive noisy observations:

$$\forall t \in \{1, ..., T\},\ u_{t+1} = z^\varepsilon(u_t) = z(u_t - \varepsilon_t) + \varepsilon_{t+1}. \tag{2}$$

If we observe N states of the noisy system, we can sort all the observations and therefore we get a discretization of the state space of the dynamical system: $u_{1:N} \leq ... \leq u_{N:N}$, respectively noted for simplicity $u_1 \leq ... \leq u_N$. Hence, we have N discrete observations, $z^\varepsilon(u_1), ..., z^\varepsilon(u_N)$, of the noisy dynamical system:

$$\forall n \in \{1, ..., N\},\ z^\varepsilon(u_n) = z(u_n - \varepsilon_n^\star) + \varepsilon_n, \tag{3}$$

where $\varepsilon_1, ..., \varepsilon_N$, $\varepsilon_1^\star, ..., \varepsilon_N^\star$ are $2N$ independent identically distributed random variables [28][1]. That noisy evolution function, z^ε, is a non-linear function of the noise.

To sum up the problem, we have sparse observations of a noisy evolution function, whereas we are mostly interested in the knowledge of the pure dynamical system. The aim of the present paper is then to present a method to denoise such a noisy signal (the function z^ε) and therefore to estimate the true or pure evolution function z introduced in equation (1).

Trajectories of dynamical systems are often very erratic, like, for instance, in Figure 1, where we represent a logistic chaos. That erratic nature of the pure trajectory makes the denoising of many noisy trajectories very challenging, even though a few methods have already been tested to denoise noisy trajectories of dynamical systems using linear wavelet filtering [46] or other smoothing techniques [26]. Instead, evolution functions are often smoother than time trajectories and their denoising is therefore more feasible. Figure 1 attests the smoothness of the same logistic chaos if we consider its evolution function in the phase space rather than its time trajectory. Therefore, we do not intend to denoise directly the trajectory of a dynamical system in the time domain but in the phase space. Both problems are linked and if we estimate accurately z then we get an estimate of the pure time trajectory.

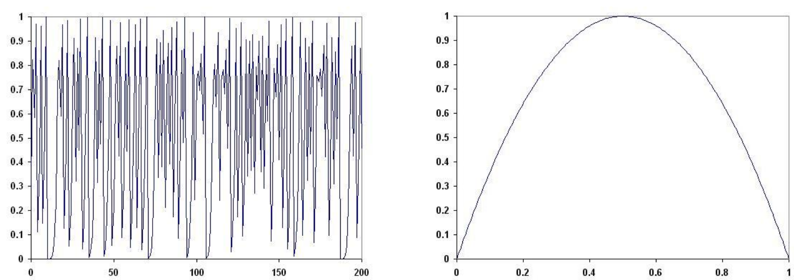

Figure 1: Time trajectory (on the left) and evolution function (on the right) of the logistic map of parameter 4: $z : x \mapsto 4x(1-x)$.

We can use several methods to denoise z^ε: local methods and singular value decomposition [41, 47, 48], maximum-likelihood-based techniques [25, 36], methods based on correlation observation [54], kernel-based non-parametric estimates [7] or methods using radial basis functions [9, 34]. For a review, we refer to [1, 42].

We are mostly interested in the wavelet shrinkage, because this technique of analysis of a signal into localised elements, which is very popular for spatially inhomogeneous signals

[1] We note $\varepsilon_1, ..., \varepsilon_N$, $\varepsilon_1^\star, ..., \varepsilon_N^\star$ the noise in equation (3). These variables are not identically related to the $\varepsilon_1, ..., \varepsilon_T$ appearing in equation (2). As well as u_{n+1} is not $z^\varepsilon(u_n)$, $z(u_n - \varepsilon_n^\star)$ is not disrupted in equation (3) by a variable ε_{n+1} but by another variable noted ε_n.

in which the noise influence is linear, allows a good accuracy and a parsimonious representation [43]. Some empirical papers have already studied that method applied to dynamical systems [26, 32, 33]. The big challenge – that we investigate in this paper – in such a method is the non-linear influence of the noise on the signal. Indeed, the literature mainly deals with the denoising based on wavelets for signals with linear noise influence, and in the present article we adapt those classical methods to the specific case of signals with non-linear noise influence.

The irregular observation grid is another specificity of our framework. Indeed, as we are interested in the evolution function, the step size between two consecutive observations in the phase space is non-constant, whereas the time trajectory is discretized by a regular observation grid. We can then make some remarks about that specificity:

- ▷ The observation grid is not only irregular, it is also stochastic, and then all the following developments are conditional to a set of observations.
- ▷ Irregular grids make the use of classical empirical wavelets time-consuming because fast empirical wavelet transform algorithms are designed for regular observation grids. However, *second-generation* wavelets allow fast empirical wavelet transform using what is known as *lifting* [11, 50, 39, 40]: the main difference with classical wavelets is that wavelets are not built anymore by dilatations and translations of a unique mother wavelet. Second-generation wavelets are particularly pertinent in multiresolution analysis and in the definition of the nested subspaces representing the different scales. Nevertheless in our framework we prefer to use first-generation wavelets because multiresolution is not our goal. Antoniadis and Fan (2001) proposed another method in which the empirical wavelet coefficients defined on a regular dyadic grid are approximated by a solution of an optimization problem involving observations on an irregular grid [3]. Another approach consists in composing first z^ε with $H : n \mapsto u_n$ [8]: then we can easily compute the wavelet coefficients of $z^\varepsilon \circ H$ on an equispaced design and thereafter adapt the inverse wavelet transform using H^{-1}. This last method works well in dimension one but it is not relevant on a multidimensional grid, where there is no order statistics and therefore no known function H. In the present paper, using a simple approximation of the integral, we define the empirical wavelet coefficient $\langle z^\varepsilon, \psi_{j,k} V\rangle$ of resolution level j and translation parameter k by[2]:
$$\langle z^\varepsilon, \psi_{j,k} V\rangle = \sum_{n=1}^{N} z^\varepsilon(u_n)\psi_{j,k}(u_n)V_n, \tag{4}$$
where $\psi_{j,k}$ is obtained by dilatation and translation of a mother wavelet Ψ:
$$\psi_{j,k} : t \in \mathbb{R} \mapsto 2^{j/2}\Psi\left(2^j t - k\right),$$
and V_n is the Voronoi cell size or step size corresponding to the observation u_n [28][3].
- ▷ There is a difference between an empirical wavelet coefficient and the corresponding theoretical wavelet coefficient. The empirical one is introduced in equation (4), whereas the theoretical one is defined for any wavelet function ψ_l by:
$$\hat{z}_l^\varepsilon = \int_{\mathbb{R}} z^\varepsilon(x)\psi_l(x)dx.$$
The empirical wavelet transform does not systematically share with the theoretical wavelet transform some properties such as orthogonality or normality[4]. Therefore, the transition from theoretical results to empirical applications may lead to some adaptations. Donoho (1992) proposed a solution, the hybrid transform, which removes the difference

[2] In the rest of the paper, we use a single index l instead of the resolution level and the translation parameter. Contrarily to both j and k that index l is not interpretable.

[3] If the evolution function is not multidimensional, we can define the steps size in an easier way than using Voronoi cells size. For example, we can define it as $(u_n - u_{n-1})$. We also note that the maximal step size can be explicitly bounded if the dynamical system is ergodic when the number of observations grows [27].

[4] Neither are they exactly equal. But, under certain assumptions, such as ergodicity and Hölder conditions, their difference can be bounded [28].

between both theoretical and empirical wavelet coefficients, based on the choice of the mother wavelet [12]. Such hybrid transform works if the observation grid is regular.

As a consequence, our theoretical results are developed for an appropriate mother wavelet conditioned by our grid[5], whereas we simply use Daubechies wavelets in our examples and we assume that, thanks to the big amount of data, any empirical wavelet coefficient calculated on our oversampling grid is similar to the corresponding theoretical wavelet coefficient. Therefore, we will indifferently write $\langle z, \psi_l V\rangle$, $\hat{z}$ or even $\langle z(u_.), \psi_l(u_.)V_.\rangle$. Moreover, the mother wavelet is supposed to generate an orthonormal family.

The objective of the present article is to determine the optimal filter in the denoising techniques using wavelet coefficients in the case of a dynamical system signal[6] perturbed by a measurement noise in order to reconstruct its true evolution function. The method consists in adapting classical results of signal processing[7] to dynamical systems, in which the noise influence is non-linear. In the example we provide, the emphasis is put on estimators based on the thresholding of the wavelet coefficients, namely the hard thresholding, the soft thresholding and a smoother thresholding which is a compromise between the characteristics of both hard and soft thresholding. Empirically, for small samples, the hard thresholding seems to behave more accurately than the soft one, since in our example[8] it eliminates up to 69.7% of the Gaussian noise, versus 65.9% for the soft thresholding and 68.3% for the smoother one. This is consistent with what is observed in the case of a linear influence of the noise in [44]: the exact error differs from the asymptotic one for a soft thresholding, which is only efficient for large samples, whereas hard thresholding seems powerful for all sample sizes. Besides, thresholding of wavelet coefficients does not work well with strongly non-Gaussian noise in both a linear and a non-linear noise influence environment. Alternative methods using wavelets have been proposed for such frameworks, like the pyramid transform in [18]. In this last method, an interpolation of the median of the observations on a time interval is used to determine an estimate of the median of the observations at a finer scale. We can thereafter filter thanks to a hard thresholding function the difference with the true median. In the case of a non-linear influence of the noise, we derive another method which simply consists in a preliminary smoothing of the outliers of the observed data before the use of the thresholding method previously proposed for a Gaussian noise. In our example[9], that method allows to eliminate up to 97.0% of a Cauchy noise.

The article is divided into three parts, describing successively the method of wavelet shrinkage of signals in which the noise influence is linear (Section 2), the theory of wavelet shrinkage of signals in which the noise influence is non-linear (Section 3), more particularly Bayes' estimate, minimax estimate, oracular estimate and thresholding estimate. Some applications, as well as a study of the robustness of the estimators to leptokurtic noise, are provided in Section 4.

2 Standard wavelet shrinkage

In that section, we briefly recall the classical results about the denoising of a signal linearly disrupted by some noise that we intend to adapt to signals in which the noise has a non-linear influence.

For a signal $(s(n))_{1\leq n\leq N}$ defined on a regular discrete grid and linearly disrupted by a white noise $(\varepsilon_n)_{1\leq n\leq N}$, of mean 0 and variance σ^2, the theory of wavelet denoising has already been developed [14, 43]. Assuming a noisy signal $(s^\varepsilon(n))_{1\leq n\leq N}$ defined by the linear relation

$$s^\varepsilon(n) = s(n) + \varepsilon_n \text{ for } 1 \leq n \leq N,$$

[5] But we do not give an explicit expression for such a wavelet.
[6] In the framework of dynamical systems, we now call signal the evolution function, z.
[7] Interesting details about classical signal processing may be found in [43].
[8] Our example consists in a logistic map of parameter 4 disrupted by a Gaussian noise of standard deviation 5%.
[9] This second example consists in a logistic map of parameter 4 disrupted by a Cauchy noise of scale parameter 2%.

we carry out a discrete wavelet analysis using an orthonormal wavelet basis $\{\psi_l\}_{1\leq l\leq N}$. Each wavelet coefficient of the noisy signal is written $\hat{s}^\varepsilon_l$, for $1 \leq l \leq N$[10]. To reduce the noise, we apply a filter operator F to each wavelet coefficient, and we get an estimator, $\tilde{s}$, of the pure signal, s, using a reconstruction based on the filtered wavelet coefficients, $F\hat{s}^\varepsilon_l$:

$$\tilde{s}(n) = \sum_{l=1}^{N} F\hat{s}^\varepsilon_l \psi_l(n).$$

The idea of wavelet shrinkage is to compare the wavelet coefficients $(\hat{s}_l)_{1\leq l\leq N}$ of the pure signal with the wavelet coefficients $(\hat{s}^\varepsilon_l)_{1\leq l\leq N}$ of the noisy signal, instead of comparing both signals, $(s(n))_{1\leq n\leq N}$ and $(s^\varepsilon(n))_{1\leq n\leq N}$. For a smooth signal, $\hat{s}_l = 0$ for many l, whereas the noise slightly pushes many $\hat{s}^\varepsilon_l$ away from zero. The denoising consists in shrinking to zero the $\hat{s}^\varepsilon_l$ induced by the noise. The filter F is chosen in a set of operators noted $\mathcal{F}$[11], to minimize the gap between the pure signal $(s(n))_{1\leq n\leq N}$ and its estimator $(\tilde{s}(n))_{1\leq n\leq N}$. Several techniques exist to make an adequate choice and to estimate the error made using a wavelet denoising, according to the *a priori* knowledge we have of the pure signal:

- ▷ If we know the *a priori* probability function π of the pure signal, we get the following error using **Bayes' estimate**[12]:

$$r_b(\mathcal{F}, \pi) = \inf_{F\in\mathcal{F}} \mathbb{E}^p \left[\mathbb{E}\left[\sum_{l=1}^{N} (\hat{s}_l - F\hat{s}^\varepsilon_l)^2 \middle| s \right]\right],$$

  where $\mathbb{E}^p$ is the expectation corresponding to the *a priori* distribution of s[13] and where the optimal linear operator F is explicitly known [43].

- ▷ If, instead of an accurate probability function, we only know a set Ξ which contains the pure signal, we choose the optimal linear filter F in the set of operators $\mathcal{F}$ and we get a **minimax estimate** [43], whose error is:

$$r_m(\mathcal{F}, \Xi) = \inf_{F\in\mathcal{F}} \sup_{s\in\Xi} \mathbb{E}\left[\sum_{l=1}^{N} (\hat{s}_l - F\hat{s}^\varepsilon_l)^2 \right].$$

  The minimax estimator is simply the Bayes' estimator with the less *a priori* favourable probability function.

- ▷ If we know s, we use an **oracle**[14] to choose the optimal linear filter such that the error is [43]:

$$r_a(s) = \inf_{F\in\mathcal{F}} \mathbb{E}\left[\sum_{l=1}^{N} (\hat{s}_l - F\hat{s}^\varepsilon_l)^2 \right].$$

- ▷ We can also define the non-linearly projected oracle by imposing $f_l = 1$ if $|\hat{s}^\varepsilon_l| \geq \sigma$ and $f_l = 0$ else [43]. The purpose of such a filter consists in selecting only the coefficients reflecting more the signal than the noise and to shrink the others to zero. The corresponding error is then noted $r_p(s)$. The errors $r_a(s)$ and $r_p(s)$ have roughly the same magnitude since they are linked by the equation

$$\frac{1}{2} r_p(s) \leq r_a(s) \leq r_p(s).$$

[10] In this standard framework, the observation points are equispaced and equal to n, for $n \in \{1, ..., N\}$. Therefore each Voronoi cell size V_n is equal to 1 and $\hat{s}^\varepsilon_l$ is simply $\sum_{n=1}^{N} s^\varepsilon(n)\psi_{j,k}(n)$.

[11] That set may be, for instance, the set of all the linear filters or the set of all the soft-threshold filters.

[12] The error which interests us consists in the difference of the original signal and the approximated one. Therefore, it should be written $\inf_{F\in\mathcal{F}} \mathbb{E}^p \left[\mathbb{E}\left[\left\| s - \sum_{l=1}^{N} F\hat{s}^\varepsilon_l \psi_l \right\|^2_{L^2(\mathbb{R})} \middle| s \right]\right]$. But Parseval's identity asserts that such an error is also equal to the quadratic difference of the wavelet coefficients. In the following theoretical developments, we use directly Parseval's identity and we limit our analysis to the error on wavelet coefficients.

[13] In the other sections of this paper, we will simply write $\mathbb{E}$ instead of $\mathbb{E}^p$ and therefore $\mathbb{E}[.]$ instead of $\mathbb{E}^p[\mathbb{E}[.|s]]$.

[14] The word *oracle* means that we apply the best possible technique using wavelet coefficient attenuation according to a specific signal.

Both oracles estimates are analytical models: in practice the pure signal is not known, but oracular estimate allows to define theoretically a minimal bound for other estimates.

▷ We can use a particular non-linear filter, known as **threshold filter** [14, 52, 43]. It may be a hard-threshold filter, defined by:

$$F\hat{s}_l^\varepsilon = \hat{s}_l^\varepsilon \mathbf{1}_{\{|\hat{s}_l^\varepsilon| \geq \lambda\}}, \tag{5}$$

for a threshold $\lambda \geq 0$, or a soft-threshold filter, defined by:

$$F\hat{s}_l^\varepsilon = (\hat{s}_l^\varepsilon - \lambda)\,\mathbf{1}_{\{\hat{s}_l^\varepsilon \geq \lambda\}} + (\hat{s}_l^\varepsilon + \lambda)\,\mathbf{1}_{\{\hat{s}_l^\varepsilon \leq -\lambda\}}. \tag{6}$$

Those filters are easy to implement and no *a priori* knowledge about the signal is necessary. Moreover, using a hard-thresholding or a soft-thresholding estimator is almost like using an oracle if the threshold is $\lambda^\star = \sigma\sqrt{2\log N}$ [14][15], because the corresponding error $r_t(s)$ is such that:

$$r_t(s) \leq (2\log(N) + 1)\left(\sigma^2 + r_p(s)\right).$$

That threshold value $\lambda^\star$ eliminates a lot of noise. Besides, when N grows, that threshold value also grows and may eliminate all the wavelet coefficients, even those reflecting more the signal than the noise. As a consequence, that threshold $\lambda^\star$ has a high theoretical interest promoting the use of threshold filters, but many other thresholds may be used, like $\lambda^\star/\sqrt{N}$ [13, 17], that do not present such a drawback: in particular $\lambda^\star/\sqrt{N}$ is lower than $\lambda^\star$ and it does not shrink to zero as many coefficients as the threshold $\lambda^\star$ does.

3 Main results

In this paragraph, we present some new results about wavelet-based denoising methods. We generalize the classical theorems concerning Bayes' estimate, minimax estimate, linear and non-linear oracular estimate and threshold estimate when the true signal is the evolution function of a dynamical system given in equation (1). In this framework, we are confronted to non-linear noise influence and irregular observation grids.

We introduce some assumptions for all the following theorems:

(A1) *z is a real differentiable function and its derivative is noted z'.*

(A1') *z' is a Lipschitz continuous function, with Lipschitz constant $K' > 0$.*

(A2) *We observe $I_N = \{u_1, ..., u_N\}$, where $u_1 \leq ... \leq u_N$, and z^ε, which follows equation (3), whose a linear approximation noted $z^{\mathcal{L}}$ is:*

$$\forall n \in \{1, ..., N\},\ z^{\mathcal{L}}(u_n) = z(u_n) - z'(u_n)\varepsilon_n^\star + \varepsilon_n. \tag{7}$$

(A3) *V_n is the Voronoi cell size corresponding to the observation u_n and we use the notation $\mathcal{V}_l$ for $\langle V^2, \psi_l^2\rangle$.*

(A4) *$\varepsilon_1, ..., \varepsilon_N, \varepsilon_1^\star, ..., \varepsilon_N^\star$ are independent identically distributed random variables with at least the four first moments finite. Their mean is zero and their variance is σ^2. Moreover, the covariance matrix of the vector $(\langle \varepsilon_., \psi_l(u_.)V_.\rangle)_{1\leq l\leq N}$, conditionally to the observation grid I_N, is diagonal.*

Assumptions *(A1)* and *(A1')* are standard conditions defining the smoothness of the evolution function z. This smoothness is necessary because of the linear approximation made in assumption *(A2)*. The quantity $\mathcal{V}_l$ defined in assumption *(A3)* is a factor scaling the noise volatility defined in assumption *(A4)* to the volatility of a noisy wavelet coefficient. In the standard case of a wavelet analysis of a time series, where $u_n = n$ for each n, we would simply have $\mathcal{V}_l = 1$ for each l because of the normality[16] of the wavelet. On the contrary, in the case

[15] The noise is assumed to be Gaussian.
[16] Normality in the sense that $\int_{\mathbb{R}} \psi_l^2(x)dx = 1$.

of a dynamical system for which there exists an invariant measure, then when N grows $\mathcal{V}_l$ decreases towards 0 at the same rate than $1/N$.

In the following, for any filter F, from the wavelet coefficients $\hat{z}_l^{\mathcal{L}}$ of the linear approximation of the noisy signal $z^{\mathcal{L}}$ introduced in equation (7), we define an estimate $\tilde{z}$ of the pure signal z by the relation:

$$\tilde{z} = \sum_{l=1}^{N} F\hat{z}_l^{\mathcal{L}}\psi_l, \tag{8}$$

and the error of the estimate $\tilde{z}$ will always be defined as the expected value of the quadratic difference of the wavelet coefficients, conditionally to the observation grid:

$$r^{\mathcal{L}} = \mathbb{E}\left[\sum_{l=1}^{N}\left(\hat{z}_l - F\hat{z}_l^{\mathcal{L}}\right)^2 \middle| I_N\right].$$

We also define, for the same filter F, another estimate $\tilde{\tilde{z}}$ of the pure signal z, from the wavelet coefficients $\hat{z}_l^{\varepsilon}$ of the noisy signal z^{ε}:

$$\tilde{\tilde{z}} = \sum_{l=1}^{N} F\hat{z}_l^{\varepsilon}\psi_l, \tag{9}$$

and its corresponding error:

$$r = \mathbb{E}\left[\sum_{l=1}^{N}\left(\hat{z}_l - F\hat{z}_l^{\varepsilon}\right)^2 \middle| I_N\right].$$

3.1 From Bayes' estimate to minimax estimate

We propose in Theorem 1 a Wiener filter which minimizes the difference between the pure evolution function z and its estimate $\tilde{z}$ obtained by applying a filter to the wavelet coefficients of $z^{\mathcal{L}}$[17]: we do not obtain the closest estimate yielding from the dynamical system but from the linear approximation of the dynamical system. Such a method allows us to get the exact value of the Wiener filter in this simpler framework. However, when the noise is small enough[18], both models (the dynamical system signal and its linear approximation) are very close, so that the wavelet filter minimizing the error for the linear approximation of z^{ε} is also a good wavelet filter for z^{ε} itself. More precisely, we can obtain an upper bound of the error of the estimate in the initial problem, when the signal is the dynamical system.

Theorem 1. *Let assume (A1), (A2), (A3), (A4). We suppose that $\{\psi_l\}_{1\leq l<N}$ is a Karhunen-Loève orthonormal basis for the linear approximation of the noisy signal: the covariance matrix of $Z = (\hat{z}_l)_{1\leq l<N}$ conditioned by I_N is diagonal. Moreover, we suppose that the random vector Z has a conditional expected value equal to zero and we denote its finite conditional variance by $\beta_l^2 = \mathbb{E}[Z_l^2|I_N]$. We also note $\Gamma_l = \langle \mathbb{E}\left[z'^2|I_N\right], \psi_l^2 V^2\rangle$.*

(i) Conditionally to I_N, the optimal filter involved in the estimate $\tilde{z}$ of z from $z^{\mathcal{L}}$, as introduced in equation (8), is $F^{\mathcal{L}} = (f_l^{\mathcal{L}})_l$, where $f_l^{\mathcal{L}} = \frac{\beta_l^2}{\beta_l^2+\sigma^2(\mathcal{V}_l+\Gamma_l)}$, and the resulting conditional error between $\tilde{z}$ and z is:

$$r_b^{\mathcal{L}}(\pi) = \sum_{l=1}^{N} \frac{\beta_l^2\sigma^2(\mathcal{V}_l+\Gamma_l)}{\beta_l^2+\sigma^2(\mathcal{V}_l+\Gamma_l)}.$$

[17] We call *optimal* the filter which minimizes the error function, r or $r^{\mathcal{L}}$, between the pure evolution function z and its estimate.

[18] Garcin and Guégan (2014) used a Kolmogorov-Smirnov test to determine when there is no significant difference between the distribution of $\hat{z}^{\mathcal{L}}$ and the distribution of $\hat{z}^{\varepsilon}$, with respect to the magnitude of the noise [28]. For example, the noise is considered small enough when its standard deviation is lower than 10% for a Gaussian noise and if the signal is a logistic chaotic map, and then no major difference appears between both the distributions.

In particular, if z is a Lipschitz continuous function, with Lipschitz constant $K > 0$, then, we have an upper bound of the error:

$$r_b^{\mathcal{L}}(\pi) \leq \sum_{l=1}^{N} \frac{\beta_l^2 \sigma^2 \mathcal{V}_l (1+K^2)}{\beta_l^2 + \sigma^2 \mathcal{V}_l (1+K^2)}.$$

(ii) *Assuming (A1'), if we apply the same conditional filter, $F^{\mathcal{L}}$, to $\hat{z}^{\varepsilon}$ in order to estimate z, leading to the estimator $\tilde{\tilde{z}}$ introduced in equation* (9)*, then the resulting conditional error $r_b(\pi)$ between $\tilde{\tilde{z}}$ and z is bounded by:*

$$\left| r_b(\pi) - r_b^{\mathcal{L}}(\pi) \right| \leq \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l \left(f_l^{\mathcal{L}} K' \sigma^2 \right)^2,$$

where we denote by κ^4 the kurtosis of the noise (ε_n)[19].

In Theorem 1, we determined the error of the approximation made using a filter which is not optimal as it does not allow to strictly minimize the error between z and $\tilde{\tilde{z}}$. Indeed, the filter we used is only optimal in the case of the linear approximation of the noisy evolution function. Therefore, a different filter than $F^{\mathcal{L}}$, close to $F^{\mathcal{L}}$[20], giving a lower error than $r_b(\pi)$, may exist. But the information disclosed by the observations do not allow to determine easily such a filter and we consider that giving the following close upper bound for the optimal error is sufficient:

$$r_b^{\mathcal{L}}(\pi) + \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l \left(f_l^{\mathcal{L}} K' \sigma^2 \right)^2.$$

Indeed, we have a practical interest in the proposition we made, since the value of the quasi-optimal filter is accurately determined. That consideration can also be applied to all the following theorems, where the idea remains the same: first, we study the optimal linear approximation and then we use that filter in the non-linear case, for which it is a quasi-optimal filter, in order to closely bound the error.

In the case where, despite knowing the probability distribution of the pure signal, we only know a set Ξ to which the pure signal belongs, we can work in the minimax estimator framework. Such estimator is simply the Bayes' estimator for the *a priori* probability distribution of signals in Ξ leading to the worst error possible. More precisely, such worst case is reached for the biggest β_l and Γ_l possible.

3.2 Oracle for diagonal attenuation

The oracle builds an estimate knowing the function to estimate and using a specific method. Therefore, the estimate is not the function z itself, as we impose the use of wavelet filtering. Oracular estimator allows to test the accuracy of a method in the most favourable circumstances and thus to give a benchmark for such a method in another framework, less favourable but more realistic, where the pure signal is not known beforehand. In our specific case, the method to which we apply an oracle is a diagonal attenuation: the noise is supposed to pollute the wavelet coefficients of the signal and the denoising method consists in multiplying each noisy wavelet coefficient by a number, the corresponding filter coordinate $f_l^{\mathcal{L}}$, in order to eliminate a big part of the noise without distorting too much the shape of the pure signal.

Theorem 2. *Let assume (A1), (A2), (A3), (A4). We also note $\Gamma_l = \langle z'^2, \psi_l^2 V^2 \rangle$.*

(i) *Conditionally to I_N, the optimal filter involved in the estimate $\tilde{z}$ of z from $z^{\mathcal{L}}$, as introduced in equation* (8)*, is $F^{\mathcal{L}} = (f_l^{\mathcal{L}})_l$, where $f_l^{\mathcal{L}} = \frac{\hat{z}_l^2}{\hat{z}_l^2 + \sigma^2(\mathcal{V}_l + \Gamma_l)}$, and the resulting conditional error between $\tilde{z}$ and z is:*

$$r_a^{\mathcal{L}}(z) = \sum_{l=1}^{N} \frac{\hat{z}_l^2 \sigma^2 (\mathcal{V}_l + \Gamma_l)}{\hat{z}_l^2 + \sigma^2 (\mathcal{V}_l + \Gamma_l)}.$$

[19] We precisely define it here as $\mathbb{E}[\varepsilon_1^4]/\sigma^4$. If the noise is Gaussian, $\kappa^4 = 3$.
[20] We can then assert that $F^{\mathcal{L}}$ is a quasi-optimal filter.

(ii) *Assuming (A1'), if we apply the same conditional filter, $F^{\mathcal{L}}$, to $\hat{z}^{\varepsilon}$ in order to estimate z, leading to the estimator $\tilde{\tilde{z}}$ introduced in equation* (9), *then the resulting conditional error $r_a(z)$ between $\tilde{\tilde{z}}$ and z is bounded by:*

$$\left| r_a(z) - r_a^{\mathcal{L}}(z) \right| \leq \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l \left(f_l^{\mathcal{L}} K' \sigma^2 \right)^2,$$

where we denote by κ^4 the kurtosis of the noise (ε_n).

Like in all the other theorems, we check that if the signal has a noise with a linear influence, then $K' = 0$ and $r_a(z) = r_a^{\mathcal{L}}(z)$ thanks to the assertion *(ii)* of Theorem 2.

3.3 Oracle for diagonal projection

Another oracle can be defined when each noisy wavelet coefficient is projected in a kind of set of *significant* wavelet coefficients. This means that we only consider the wavelet coefficients which sound to reflect anything else than the noise.

More precisely, conditionally to the observation grid, each coordinate of the optimal filter $F^{\mathcal{L}}$, defining the estimate $\tilde{z}$ introduced in equation (8), now belongs to the set $\{0, 1\}$. It is thus defined by:

$$f_l^{\mathcal{L}} = \begin{cases} 1 & \text{if } \hat{z}_l^2 \geq \sigma^2 \left(\mathcal{V}_l + \Gamma_l \right) \\ 0 & \text{if } \hat{z}_l^2 < \sigma^2 \left(\mathcal{V}_l + \Gamma_l \right), \end{cases} \tag{10}$$

where $\Gamma_l = \langle z'^2, \psi_l^2 V^2 \rangle$. As before, we define the estimate $\tilde{\tilde{z}}$ introduced in equation (8) using the same conditional filter $F^{\mathcal{L}}$.

Theorem 3 provides a link between the estimate error obtained by the oracle using diagonal attenuation and the oracle using diagonal projection.

Theorem 3. *Let assume (A1), (A2), (A3), (A4). We note r_p and r_a the conditional errors between $\tilde{\tilde{z}}$ and z respectively in the projection and the attenuation case of oracles. $r_p^{\mathcal{L}}$ and $r_a^{\mathcal{L}}$ are the conditional error between $\tilde{z}$ and z respectively in the projection and the attenuation case of oracles. We also note $\Gamma_l = \langle z'^2, \psi_l^2 V^2 \rangle$. Then:*

(i) *$r_p^{\mathcal{L}}$ and $r_a^{\mathcal{L}}$ are linked by:*

$$\frac{1}{2} r_p^{\mathcal{L}}(z) \leq r_a^{\mathcal{L}}(z) \leq r_p^{\mathcal{L}}(z).$$

(ii) *If we assume (A1'), then:*

$$\left| r_p(z) - r_p^{\mathcal{L}}(z) \right| \leq \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l \left(f_l^{\mathcal{L}} K' \sigma^2 \right)^2,$$

where κ^4 is the kurtosis of the noise and $f_l^{\mathcal{L}}$ is defined in equation (10).

(iii) *r_p and r_a are linked by:*

$$\frac{1}{2} r_p(z) - 2R_a \leq r_a(z) \leq r_p(z) + 3R_a,$$

where $R_a = \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l \left[\frac{\hat{z}_l^2}{\hat{z}_l^2 + \sigma^2 (\mathcal{V}_l + \Gamma_l)} \right]^2 \left(K' \sigma^2 \right)^2$.

In the last inequality of Theorem 3, we note that R_a is simply the upper bound of the difference between r_a and $r_a^{\mathcal{L}}$ that we introduced in Theorem 2:

$$\left| r_a(z) - r_a^{\mathcal{L}}(z) \right| \leq R_a.$$

3.4 Threshold estimator

Threshold estimators are non-linear estimators which aim to shrink to zero the wavelet coefficients supposed to reflect only the noise. Thus, that estimator does not take into account the coefficients below a predetermined threshold, λ. The underlying idea is that many of the wavelet coefficients of a relatively smooth signal are equal to zero: for example, if $k \in \mathbb{N}$ and if the mother wavelet has k vanishing moments, then any theoretical wavelet coefficient of a polynomial of degree k or lower is equal to zero.

The wavelet coefficients that are not below the threshold λ are supposed to reflect partly the pure signal. They remain unchanged if we use a hard-thresholding filter, defined by equation (5). On the contrary, those coefficients may be shrunk towards zero by the quantity λ, in order to have a continuous thresholding function, as defined in equation (6) relative to the soft-thresholding filter. The continuity of the soft-thresholding function is important for example when we have to estimate the error function without any *a priori* knowledge of the pure function to estimate but directly from the observation panel. Stein's unbiased risk estimator (SURE) can indeed directly be applied to soft-thresholding in the classical framework thanks to the continuity of the thresholding function, whereas no unbiased error estimator can be obtained for hard-thresholding [49, 16, 43].

Before to present how to adapt SURE to our framework, we extend in Theorem 4 the classical results about thresholding to the specific case of a dynamical system signal.

Theorem 4. *Let assume (A1), (A2), (A3), (A4). Let the noise $\varepsilon_1, ..., \varepsilon_N, \varepsilon_1^\star, ..., \varepsilon_N^\star$ be Gaussian.*

(i) Let $\Gamma_l = \langle z'^2, \psi_l^2 V^2 \rangle$. Conditionally to I_N, if the threshold $\lambda_l^\star$, for a hard or a soft thresholding filter, is such that:

$$\lambda_l^\star = \sigma\sqrt{(\mathcal{V}_l + \Gamma_l)\, 2\log N},$$

then, the error between $\tilde{z}$ introduced in equation (8) and z is near the error obtained by a projection oracle:

$$r_t^{\mathcal{L}}(z, (\lambda_l^\star)_l) \leq (1 + 2\log N)\left(\sigma^2\left(\bar{\mathcal{V}} + \bar{\Gamma}\right) + r_p^{\mathcal{L}}(z)\right),$$

where $\bar{\Gamma} = N^{-1}\sum_{l=1}^N \Gamma_l$ and $\bar{\mathcal{V}} = N^{-1}\sum_{l=1}^N \mathcal{V}_l$. In particular, if z is a Lipschitz continuous function, with Lipschitz constant $K > 0$, then $\Gamma_l \leq K^2\mathcal{V}_l$ and $\sigma\sqrt{2\mathcal{V}_l \log N} \leq \lambda_l^\star \leq \sigma\sqrt{2\mathcal{V}_l(1+K^2)\log N}$.

(ii) We assume that z is a Lipschitz continuous function, with Lipschitz constant $K > 0$. If the threshold λ°, for a hard or a soft thresholding filter, is such that:

$$\lambda_l^\circ = \sigma\sqrt{2\mathcal{V}_l\,(1+K^2)\log N},$$

then, the error between $\tilde{z}$ and z is near the error for a projection oracle:

$$r_t^{\mathcal{L}}(z, (\lambda_l^\circ)_l) \leq \left(1 + 2\left(1+K^2\right)\log N\right)\left(\sigma^2\left(\bar{\mathcal{V}} + \bar{\Gamma}\right) + r_p^{\mathcal{L}}(z)\right),$$

where $\bar{\Gamma} = N^{-1}\sum_{l=1}^N \Gamma_l$ and $\bar{\mathcal{V}} = N^{-1}\sum_{l=1}^N \mathcal{V}_l$.

(iii) Moreover, assuming (A1'), then, for any threshold λ_l, the error $r_t(z, (\lambda_l)_l)$ between $\tilde{\tilde{z}}$ introduced in equation (9) and z is such that:

$$\left| r_t(z, (\lambda_l)_l) - r_t^{\mathcal{L}}(z, (\lambda_l)_l) \right| \leq \frac{3}{4} N\bar{\mathcal{V}}\left(K'\sigma^2\right)^2.$$

In the assertion *(i)* of Theorem 4, the threshold depends on the value of z' at the observation points. However, these values are not necessarily known, except if we make the assumption of using an oracle. This is a big challenge in our non-linear framework. Indeed, for a classical time series signal in which the noise influence is linear, the benefit of the threshold estimator is that, contrarily to the other estimators presented here (Bayes, minimax, oracle), it is universal, in the sense that it does not depend on the pure signal: such classical result is interesting because it shows that the method of threshold estimator, for a certain threshold $\sigma\sqrt{2\log N}$, is in the same magnitude than the best linear estimator, the oracular one. In fact, practitioners sometimes prefer other thresholds than that classical theoretical universal threshold, $\sigma\sqrt{2\log N}$, in order to denoise the signal, because when N grows the whole signal is eliminated by the filter as the threshold grows without any bound [2, 17, 35]. Therefore, even in the classical framework of a linear influence of the noise, one may use threshold values depending on the signal in order to get the best estimator.

Moreover, building an estimate for σ from the noisy signal is challenging, at least more than in the linear case[21]. In addition, building an estimate for Γ_l, or for z', is also stimulating. One can imagine an iterative algorithm, building from an arbitrary well-chosen threshold an estimate for the pure signal and therefore of its derivative. Then, using that new estimate for Γ_l, one gets a new threshold thanks to Theorem 4 and a new estimate for the pure signal and of its derivative, and so forth. The main challenge of such an algorithm is to get an estimate for z' from z: it is a well-known ill-posed inverse problem [20].

The assertion *(ii)* of Theorem 4 specifies a more universal threshold than the assertion *(i)*. More precisely, we no longer have to know the value of z' at the observation points. Instead, the belonging of z to a specific set of functions is enough. This is much less restrictive: the knowledge of z' was a kind of oracle whereas the belonging of z to a specific set of functions is close to Bayes or minimax frameworks.

Whereas assertions *(i)* and *(ii)* of Theorem 4 relate to the estimate of z from the linear approximation of the noisy evolution function, the assertion *(iii)* allows to link both the estimates of z from $z^{\mathcal{L}}$ and from z^{ε} whatever the threshold. The difference between both error functions is bounded by $(3/4)N\bar{\mathcal{V}}\left(K'\sigma^2\right)^2$. If, instead of the error, we consider the mean error relatively to the size of the sample, then the difference between the errors is simply bounded by $(3/4)\bar{\mathcal{V}}\left(K'\sigma^2\right)^2$.

A non-parametric choice of the threshold is also possible: besides all these theoretical threshold values, the choice of the threshold may be based on the optimisation problem consisting in reducing the error function, that is the expected value of the quadratic difference between the pure signal and the filtered noisy one. The challenge is then to have an idea of such an error function when the pure signal is unknown. For signals with a linear noise impact, several methods are possible in order to estimate the error function from the noisy observations only.

For example, one may split the sample in two subsets and compare the interpolations obtained from one of both denoised subsets with the noisy realization contained in the other subset and reciprocally. That difference depends on the value of the threshold in the denoising filter. One can then minimize the resulting objective function and therefore choose the optimal filter [45].

The same idea leads to *SureShrink*, which is a soft-threshold minimizing an unbiased estimate of the error, in the case of a signal with a linear noise influence [49, 16, 43]. It is based on Stein's unbiased risk estimator (SURE).

For a signal with a non-linear noise influence, we would be able to determine an unbiased error estimate if we could observe the linear approximation $z^{\mathcal{L}}$ of the noisy signal z^{ε}. But we just observe z^{ε} and we are unable to build an unbiased estimator of the error. We can however bound the bias of the estimator we propose in Theorem 5

Theorem 5. *Let assume (A1), (A2), (A3), (A4) and (A1'). Let the noise $\varepsilon_1, ..., \varepsilon_N, \varepsilon_1^\star, ..., \varepsilon_N^\star$ be Gaussian. Let $\lambda \geq 0$ and $\Gamma_l = \langle z'^2, \psi_l^2 V^2\rangle$. Let $\tilde{r}_t(z,\lambda)$ be an estimator of $r_t(z,\lambda)$, for a soft-thresholding filter, defined by:*

$$\tilde{r}_t(z,\lambda) = \sum_{l=1}^{N} \mathcal{S}_l(\hat{z}_l^{\varepsilon}),$$

[21] Typically, for time series signals with a linear impact of the noise, one can calibrate σ using the median M of all the wavelet coefficients at the finest scale. It is indeed shown in [14] that $\mathbb{E}[M] = 0.6745\sigma$. In our framework, we would rather have $\mathbb{E}[\tilde{M}] = 0.6745\sigma$, where $\tilde{M}$ is the median of all the ratios of each wavelet coefficient by the corresponding $\sqrt{\mathcal{V}_l + \Gamma_l}$ at the finest scale.

where

$$\mathcal{S}_l : x \mapsto \begin{cases} \lambda^2 + \sigma^2(\mathcal{V}_l + \Gamma_l) & \textit{if } |x| \geq \lambda \\ x^2 - \sigma^2(\mathcal{V}_l + \Gamma_l) & \textit{else.} \end{cases}$$

Then, the bias of that estimator, conditionally to I_N, is bounded by:

$$|\mathbb{E}\left[\tilde{r}_t(z,\lambda) - r_t(z,\lambda)|I_N\right]| \leq \sum_{l=1}^{N} K'\sigma^2 \left[2\lambda\mathcal{U}_l + 2\sigma\sqrt{3\mathcal{V}_l(\mathcal{V}_l+\Gamma_l)} + \frac{3}{4}\mathcal{V}_l K'\sigma^2\right],$$

where $\mathcal{U}_l = \langle V, |\psi_l|\rangle$.

In the following example, we will compare that estimator of the error with the true error function for a soft-thresholding filter. But before the examples, we give a formula of the expected error for a threshold filter. That formula will be useful to compute the mean of the reconstruction errors for different values of threshold.

3.5 Expected error for the threshold filters

The knowledge of the probability density function of the linear approximation of all the wavelet coefficients allows to write the expected error for the threshold filter applied to the linear approximation of the wavelet coefficients as a simple expression. If we assume that the noise is Gaussian with mean 0 and variance σ^2 then $\hat{z}_l^{\mathcal{L}}$ conditioned by I_N is a Gaussian random variable of mean $\hat{z}_l$ and variance $\sigma^2(\mathcal{V}_l + \Gamma_l)$, where $\Gamma_l = \langle z'^2, \psi_l^2 V^2\rangle$. Let $\delta = \delta_{\hat{z}_l, \sigma^2(\mathcal{V}_l+\Gamma_l)}$ and $\Delta = \Delta_{\hat{z}_l, \sigma^2(\mathcal{V}_l+\Gamma_l)}$ be the corresponding probability density and cumulative distribution function.

Three cases arise, depending on the nature of the filter we use in this paper: a hard-threshold filter, a soft-threshold filter or a smoother filter defined for a particular parameter $\lambda > 0$ by the function

$$\mathcal{F}_\lambda : x \mapsto x\left(1 - \exp\left(\frac{-x^2}{\lambda\sqrt{\mathcal{V}_l + \Gamma_l}}\right)\right). \tag{11}$$

The first two cases are already known and the corresponding expected error appears in [15]:

$$\begin{cases} r^{\mathcal{L}}_{t,hard}(z,\lambda) = \sum_{l=1}^{N} & \sigma^2(\mathcal{V}_l+\Gamma_l) + (\hat{z}_l^2 - \sigma^2(\mathcal{V}_l+\Gamma_l))[\Delta(\lambda) - \Delta(-\lambda)] \\ & +\sigma^2(\mathcal{V}_l+\Gamma_l)(\lambda - \hat{z}_l)\delta(\lambda) + \sigma^2(\mathcal{V}_l+\Gamma_l)(\lambda+\hat{z}_l)\delta(-\lambda) \\ r^{\mathcal{L}}_{t,soft}(z,\lambda) = \sum_{l=1}^{N} & \sigma^2(\mathcal{V}_l+\Gamma_l) + \lambda^2 + (\hat{z}_l^2 - \sigma^2(\mathcal{V}_l+\Gamma_l) - \lambda^2)[\Delta(\lambda) - \Delta(-\lambda)] \\ & -\sigma^2(\mathcal{V}_l+\Gamma_l)(\lambda + \hat{z}_l)\delta(\lambda) - \sigma^2(\mathcal{V}_l+\Gamma_l)(\lambda - \hat{z}_l)\delta(-\lambda). \end{cases}$$

We now derive the last case. For the filter $\mathcal{F}_\lambda$ defined for a given $\lambda > 0$, we have the following expression for the error of reconstruction:

$$\begin{aligned} r^{\mathcal{L}}_{t,\mathcal{F}_\lambda}(z,\lambda) = \sum_{l=1}^{N} \quad & \Big(\sigma^2(\mathcal{V}_l+\Gamma_l) + 2\exp\left(\tfrac{-\hat{z}_l^2(1-\omega)}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)\omega^{3/2}\left[(1-\omega)\hat{z}_l^2 - \sigma^2(\mathcal{V}_l+\Gamma_l)\right] \\ & +\exp\left(\tfrac{-\hat{z}_l^2(1-\omega')}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)(\omega')^{3/2}\left[\hat{z}_l^2 + \sigma^2(\mathcal{V}_l+\Gamma_l)\right]\Big), \end{aligned}$$

where

$$\begin{cases} \omega = \frac{\lambda}{\lambda + 2\sigma^2\sqrt{\mathcal{V}_l+\Gamma_l}} \\ \omega' = \frac{\lambda}{\lambda + 4\sigma^2\sqrt{\mathcal{V}_l+\Gamma_l}}. \end{cases}$$

We give some details about how we get this formula, in Appendix G.

4 Example

We consider that the signal is a logistic map of coefficient α, with $0 \leq \alpha \leq 4$. Such a signal is defined by the function:

$$z_\alpha : x \in [0,1] \mapsto \alpha g(x),$$

where $g(x) = x(1-x)$. If $\alpha = 4$, then the signal is chaotic [31]. Moreover, z'_α is a Lipschitz continuous function, with Lipschitz constant α.

In this section, we want to compare the filters previously introduced, when the signal is disrupted by a measurement noise. On the one hand, we will consider a Gaussian noise of standard deviation $\sigma = 5\%$ and, on the other hand, we will introduce a leptokurtic noise. Moreover, except in the Bayes and minimax framework, we will always assume that $\alpha = 4$. We also note that in this example I_N is a *small* sample of $N = 130$ observations. Indeed, our theoretical results are not asymptotic and work well insofar as the number of observations is sufficient to overcome discretization problems when calculating the empirical wavelet coefficients.

4.1 Gaussian noise

We assume that the noise is a Gaussian noise of mean 0 and of standard deviation $\sigma = 5\%$.

4.1.1 Bayes' estimator

We consider, for the purpose of our example, that the signal is random and that the probability distribution of α is uniform in $[0, 4]$. Then, we can apply Theorem 1, with:

$$\left\{ \begin{array}{lcl} \beta_l^2 & = & (\mathbb{E}[\alpha]\langle g, \psi_l \rangle)^2 = 4\hat{g}_l^2 \\ \Gamma_l & = & \mathbb{E}[\alpha^2]\langle g'^2, \psi_l^2 V^2 \rangle = \frac{16}{3}\langle g'^2, \psi_l^2 V^2 \rangle \end{array} \right.$$

and the error is 0.00042. This corresponds to 95.4% of the noise eliminated.

4.1.2 Minimax estimator

Instead of knowing the probability distribution of the pure signal, we now consider that we only know that the pure signal is a logistic function of coefficient α, with $0 \leq \alpha \leq 4$. The minimax estimator is the one with the *a priori* distribution giving the worst error in the Bayes framework. Such worst distribution gives the biggest β_l^2 and Γ_l possible. Therefore, we have to find the distribution of α, in $[0, 4]$, leading to the maximum β_l^2 and Γ_l, or, more precisely, leading to the maximum $(\mathbb{E}[\alpha])^2$ and $\mathbb{E}[\alpha^2]$. There is only one solution: the discrete probability distribution for which the value of the probability of $\alpha = 4$ is 1. The worst signal is then the chaotic one, and the minimax estimator is defined by the optimal filter given in Theorem 1, where:

$$\left\{ \begin{array}{lcl} \beta_l^2 & = & 16\hat{g}_l^2 \\ \Gamma_l & = & 16\langle g'^2, \psi_l^2 V^2 \rangle. \end{array} \right.$$

and the error is 0.0011. This corresponds to 87.8% of the noise eliminated.

4.1.3 Oracular estimator

We now consider that the pure signal is a logistic of parameter 4, which is the chaotic case of the logistic function. Since the minimax estimator corresponds to that case, we have already detailed the optimal oracular filter by attenuation. In particular, we note that 87.8% of the noise is eliminated using this linear oracle. Besides, the oracle for diagonal projection uses the following filter.

$$f_l^{\mathcal{L}} = \left\{ \begin{array}{ll} 1 & \text{if } \hat{g}_l^2 \geq \sigma^2 \left(\frac{\mathcal{V}_l}{16} + \langle g'^2, \psi_l^2 V^2 \rangle \right) \\ 0 & \text{if } \hat{g}_l^2 < \sigma^2 \left(\frac{\mathcal{V}_l}{16} + \langle g'^2, \psi_l^2 V^2 \rangle \right). \end{array} \right.$$

We can then apply Theorem 3 in order to get the error of such an estimator, which is 0.0048 and corresponds to the elimination of 48.0% of the noise.

4.1.4 Threshold estimator

We filter the wavelet coefficients thanks to non-linear filters, in particular threshold filters. First, hard thresholding filter has the drawback to produce an error function which is not very smooth in the value of the threshold as one can see in Figure 2. Therefore, one can be tempted to use soft thresholding which is much smoother. Moreover, the error of soft thresholding can be estimated from the noisy wavelet coefficients thanks to the function $\mathcal{S}$ of Theorem 5. We see in Figure 2 that this estimate is not perfect but its shape is similar to the curve of the error calculated with the *a priori* knowledge of the pure evolution function. In particular, their minimum almost coincides. However, the performance for soft thresholding is not better than for hard thresholding, as one can see in Figure 2 and in Table 1. One of the reasons for that is that the continuity of the soft thresholding filter is obtained at the expense of the modification of the value of coefficients bigger than the threshold whereas those wavelet coefficients are considered to be relevant. This disadvantage is particularly marked for small datasets as noticed in [44]. One can then use other non-linear filters. For example, the filter introduced in equation (11) cumulates the advantages of being continuous and not modifying a lot *big* wavelet coefficients, even though it has some drawbacks such as the absence of an estimator for the reconstruction error.

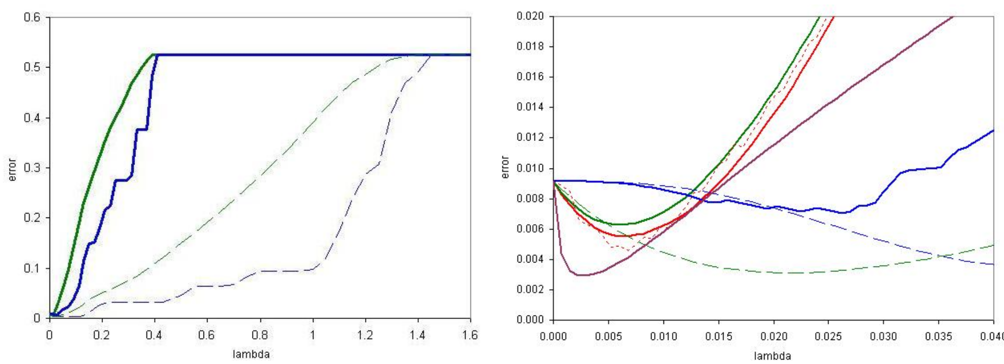

Figure 2: Mean of the reconstruction error for the evolution function of a logistic map of parameter 4 and with a Gaussian noise of standard deviation $\sigma = 5\%$, for different values of a threshold λ and for $N = 130$. The figure on the right is a focus on small values of λ. In green is the error for the soft thresholding, in blue for the hard thresholding. The solid line stands for a thresholding with the same λ for all wavelet coefficients, whereas the dotted line stands for a level-dependant threshold. In that level-dependent case the threshold applied to each wavelet coefficient is simply $\lambda\sqrt{\mathcal{V}_l + \Gamma_l}/\max_l \sqrt{\mathcal{V}_l + \Gamma_l}$. In red is the estimate of the error thanks to the function $\mathcal{S}$ of Theorem 5: the solid line stands for the mean obtained for 100 simulations whereas the dotted line corresponds to only one simulation. In purple is the error function calculated for the smooth non-linear filter $\mathcal{F}_\lambda$ defined in equation (11). In the hard, soft and smooth thresholding cases, the mean error is calculated from the approximation of the probability density of the wavelet coefficients.

Let us now focus on Table 1, in which we have reported the errors obtained for different threshold values. In particular, a naive interpretation of the literature about the denoising of classical signals may lead to use the *universal* threshold $\sigma\sqrt{2\ln N}$ without any adjustment. This choice is worse than not filtering the noisy signal. Indeed, as one can see in Figure 2, whatever the thresholding function the error first decreases when the threshold grows and then the error grows rapidly when the threshold is above a certain value. The convergence of that error function towards a constant value corresponds to the situation where all the wavelet coefficients are shrunk to zero. Therefore, the threshold $\sigma\sqrt{2\ln N}$ is much bigger than the admissible values which are closer to zero.

Next, if one amends the classical *universal* threshold to take into account the particular irregular grid, then the threshold becomes $\sigma\sqrt{2\mathcal{V}_l \ln N}$ and the accuracy of the estimate is improved, even though only less than 5% of the noise is eliminated for the hard threshold, whereas up to almost 40% of the noise is eliminated for the soft threshold. In addition to that grid adjustment, if the non-linearity of the noise influence appears in the threshold, which is now $\sigma\sqrt{2(\mathcal{V}_l + \Gamma_l)\ln N}$, then the hard threshold leads to almost 70% of the noise eliminated, whereas the soft threshold eliminates only 12.9% of the noise. The good performance of the hard thresholding filter for such a threshold in comparison to previously tested thresholds underlines the importance of using both adjustments when filtering the wavelet coefficients of a noisy dynamical system.

Threshold	Hard threshold error	Noise eliminated	Soft threshold error	Noise eliminated
$\sigma\sqrt{2\ln N}$	0.1484	-	0.2699	-
$\sigma\sqrt{2\mathcal{V}_l \ln N}$	0.0088	4.1%	0.0055	39.8%
$\sigma\sqrt{2(\mathcal{V}_l+\Gamma_l)\ln N}$	0.0028	69.6%	0.0080	12.9%
best constant threshold	0.0070	22.9%	0.0063	31.1%
best threshold in $\sqrt{\mathcal{V}_l+\Gamma_l}$	0.0028	69.7%	0.0031	65.9%

Table 1: Mean of the reconstruction errors obtained for different thresholds for the evolution function of a logistic map of parameter 4 and with a Gaussian noise of standard deviation $\sigma = 5\%$ and for $N = 130$.

In a theoretical point of view or if one uses a good estimator of the error, what sounds possible for a soft threshold, on can choose a threshold minimizing the error function. If one uses the same threshold for all the wavelet coefficients, then the proportion of the noise eliminated is 22.9% for a hard threshold and 31.1% for a soft threshold. The performance is greater if we use level-dependent thresholds. Several types of level-dependence are possible but we have chosen a dependence in $\sqrt{\mathcal{V}_l+\Gamma_l}$ which is consistent with the dependence of the thresholds $\lambda_l^\star$ in Theorem 4. In our example, using such a threshold eliminates almost 70% of the noise for a hard threshold: this is almost the same performance than using $\sigma\sqrt{2(\mathcal{V}_l+\Gamma_l)\ln N}$ as a threshold value. For the soft threshold, the best threshold dependent in $\sqrt{\mathcal{V}_l+\Gamma_l}$ leads to the elimination of 65.9% of the noise. This is quite a good result comparing to all the other soft thresholds we have tested. That result makes the use of the error estimate of Theorem 5 relevant, since it allows a good estimation of this best soft threshold. On the contrary, we did not present any error estimate for the hard thresholding and we are thus unable to reach the optimal hard threshold without *a priori* knowledge of the pure signal, whereas in this example the hard threshold $\sigma\sqrt{2(\mathcal{V}_l+\Gamma_l)\ln N}$ seems very efficient.

In comparison to the best hard and soft thresholds, the smooth non-linear filter we have tested, $\mathcal{F}_\lambda$, leads to results in the same range than with a hard threshold and slightly better than with a soft threshold: at best the error is 0.0029, which corresponds to 68.3% of the noise eliminated.

4.2 Robustness of the threshold filters to leptokurtic noise

Theorems 4 and 5 relate to a Gaussian noise. One may wonder if the method of thresholding is robust to other kinds of noise, particularly leptokurtic noise. In absence of theoretical response to that matter, we will focus on an empirical analysis of the robustness.

We first consider the impulse response if the observed evolution function if sharply disrupted at one point. None of the aforementioned methods allow to eliminate such an outlier. Indeed, the wavelet coefficients at the finest scale are strongly impacted by such a noise: therefore, their value is far from zero and cannot be eliminated by any filter. In particular, thresholding methods are not robust to extreme noise as we can see in the impulse response graphs in Figure 3. This drawback is also noticed by [18] in the traditional framework of a linear noise influence.

Let us now consider another kind of extreme noise which would be different from a Gaussian noise with an impulse. We specifically study the Cauchy noise, which has a high probability to draw variables in the tail of its distribution. The probability density of the empirical wavelet coefficients of a dynamical system disrupted by a Cauchy noise has been studied in [28]. For a Cauchy distribution of small scale parameter, we then know a good approximation of that probability density of wavelet coefficients and we could therefore calculate expected quadratic errors as well as for a Gaussian noise. However, such expected error is not defined for the Cauchy noise, especially because of its leptokurtic feature. Therefore, we will limit our analysis to an example with only one simulation and not an average of several simulations. The error will therefore be the exact quadratic error for the wavelet coefficients of that simulation. In

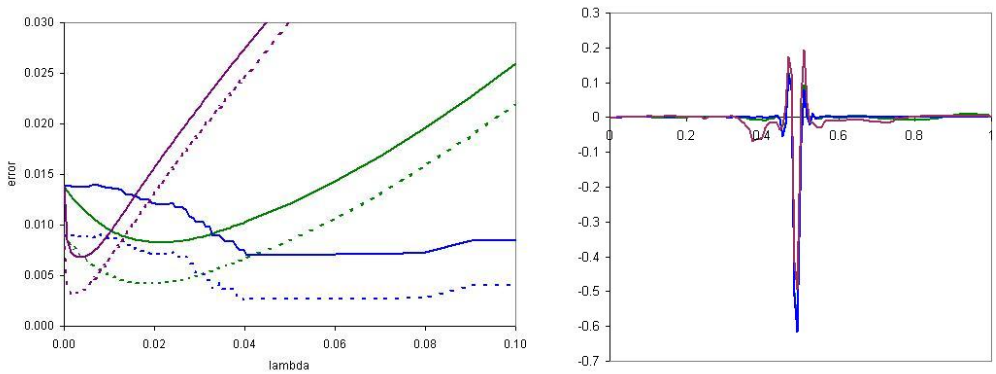

Figure 3: We consider the evolution function of a logistic map of parameter 4 with a Gaussian noise of standard deviation $\sigma = 5\%$ for $N = 130$. We have disrupted that signal with an impulse point: we have subtracted 1 to the evolution function at the abscissa 0.5. On the **left** figure, we see the empirical reconstruction error depending on the threshold parameter λ for such a simulation, with (solid line) and without (dotted line) the impulse for a level-dependent hard thresholding (blue), a level-dependent soft thresholding (green) and a smooth thresholding $\mathcal{F}_\lambda$ (purple). The impulse seems to translate the error curves upwards; it keeps quite unchanged the optimal λ for each thresholding kind. On the **right** figure, we see the resulting difference between the reconstructed evolution function with and without the negative impulse for an arbitrary level-dependent threshold, $\sigma\sqrt{2(\mathcal{V}_l + \Gamma_l)\ln N}$. The impulse noise is spread over a large range of abscissas. Such impulse response highlights the lack of robustness of the method to leptokurtic noise.

Figure 4, we present two kinds of curves. In the first kind, we present the error when we apply a thresholding filter to the raw data disrupted by a Cauchy noise. In the second kind of curves, we first smooth the data in a proper way in order to get rid of the extreme noise. In order to smooth the data, we could truncate the series of wavelet coefficients. But we propose a better method in which the wavelet coefficients at the finest scale are not removed. More precisely, for each observation point of the noisy evolution function, we calculate a local median on a small number of close observations.[22] If that median is far from the observation, we consider that the observation is an outlier and we replace it by the median. Since the median is robust, such a method allows to get rid of most of the observations disrupted by an extreme noise without changing the other observations.

As we can see in Figure 4, the estimate $\mathcal{S}$ of the error for soft-thresholding fails. If we do not smooth the data, the benefit of thresholding is narrow. However, if we smooth the observations as we previously detailed, the thresholding is efficient for a large range of thresholds. This offsets the fact that we are not able to estimate accurately the optimal threshold using $\mathcal{S}$. The errors reported in Table 2 emphasize the effectiveness of our method. We also note that only smoothing the noisy observations of the evolution function eliminates 91.4% of the noise. Of course, the weight of extreme noise, which is eliminated while smoothing the observations, is preponderant in the reconstruction error. Besides, the filtering of the wavelet coefficients of the smoothed data improves a lot the denoising: up to 65.5% of the noise remaining after the smoothing of the dataset is eliminated next when we use a hard thresholding.

5 Conclusion

We have presented several methods to filter the wavelet coefficients of the evolution function of a noisy dynamical system. Bayes's estimate, as well as minimax estimate and oracular estimates are studied in Theorems 1, 2 and 3. Non-linear filters are also introduced for the particular signals we consider. Threshold estimate is indeed presented in Theorem 4. Theorem 5 gives a way to optimise the threshold used for such estimates when we only have an estimate of the error. That estimate of the error has a small bias that we are able to bound.

[22] Increasing the number of considered neighbours worsens the accuracy.

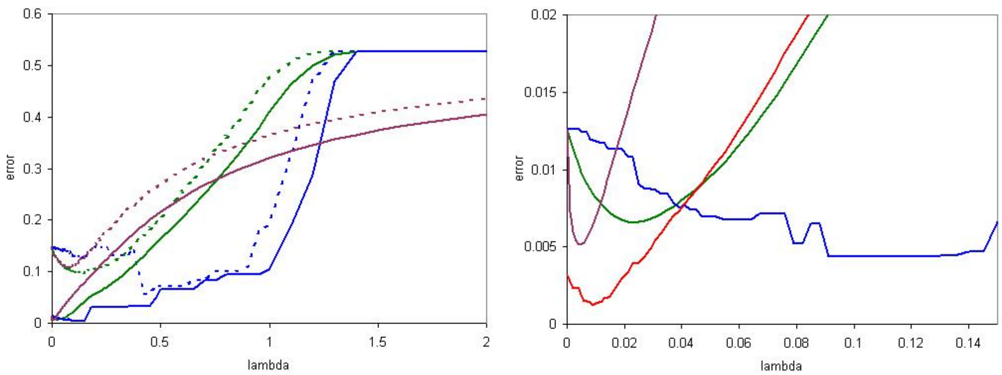

Figure 4: Reconstruction error for one simulation of the evolution function of a logistic map of parameter 4 and with a Cauchy noise of scale parameter $\gamma = 2\%$, for different parameters of a level-dependent threshold λ and for $N = 130$. In that level-dependent case the threshold applied to each wavelet coefficient is still $\lambda\sqrt{\mathcal{V}_l + \Gamma_l}/\max\limits_l \sqrt{\mathcal{V}_l + \Gamma_l}$. The figure on the right is a focus on small values of λ. In green is the error for the soft thresholding, in blue for the hard thresholding, in purple for the smooth thresholding $\mathcal{F}_\lambda$. The dotted line stands for the thresholding of the wavelet coefficients calculated on the raw noisy observations of the evolution function, whereas the solid line stands for the thresholding of the wavelet coefficients calculated on the smoothed noisy observations of the evolution function. In red is the estimate of the error thanks to the function $\mathcal{S}$ of Theorem 5 if one makes the error of considering the Cauchy noise as a Gaussian noise: the scale parameter of the Gaussian noise is robustly approximated by $\hat{\sigma}$, which is equal to 0.6745 times the median of the wavelet coefficients at the finest scale, whereas such a median is equal to γ. Therefore, $\hat{\sigma} = 2.965\%$ and the red line is not a particularly relevant estimate of the green one.

Thresholding type	Raw observations	Smoothed observations
Best soft threshold	32.7%	95.5%
Best hard threshold	61.3%	97.0%
Best smooth threshold $\mathcal{F}_\lambda$	25.8%	96.5%

Table 2: Proportion of the noise eliminated for different thresholds, with a dependence in $\sqrt{\mathcal{V}_l + \Gamma_l}$, for the evolution function of a logistic map of parameter 4 and with a Cauchy noise of scale parameter $\gamma = 2\%$ and for $N = 130$.

To sum up, the classical results on wavelet filtering of signals with linear noise influence are transformed in different ways:

- ▷ The noisy wavelet coefficients standard deviation, which appears when calibrating the different filters, is different from the one faced when denoising a standard time series with linear noise influence:
  - it is first scaled by a factor $\mathcal{V}_l$ to the phase space;
  - it is then enhanced by a factor $(\mathcal{V}_l + \Gamma_l)/\mathcal{V}_l$ taking into account the non-linearity of the noise influence.
- ▷ Biases appear as we are only able to work with linear approximations of the noisy evolution function. However, those biases are all explicitly bounded, permitting the use of these theoretical results in practice, what is a very interesting challenge.

The example provided in Section 4 shows the pertinence of our method to denoise a noisy dynamical system based on a filtering of the wavelet coefficients of its evolution function. We also improve the method by a preliminary smoothing of outliers in order to eliminate leptokurtic noise.

References

[1] Abarbanel, H., Brown, R., Sidorowich, J. and Tsimring, L. (1993), *The analysis of observed chaotic data in dynamical systems*, Reviews of modern physics, 65, 4: 1331-1392

[2] Abramovich, F., Sapatinas, T. and Silverman, B. (1998), *Wavelet thresholding via a Bayesian approach*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60, 4: 725-749

[3] Antoniadis, A. and Fan, J. (2001), *Regularization of wavelet approximations*, Journal of the American Statistical Association, 96: 939-967

[4] Bardet, J.-M. (1999), *La mémoire longue en économie : discussion et commentaires*, Journal de la société française de statistique, 140, 2: 49-54

[5] Bardet, J.-M., Lang, G., Moulines, E. and Soulier, P. (2000), *Wavelet estimator of long-range dependent processes*, Statistical inference for stochastic processes, 3, 1-2: 85-99

[6] Bollerslev, T. (1986), *Generalized autoregressive conditional heteroskedasticity*, Journal of econometrics, 31, 3: 307-327

[7] Bosq, D. and Guégan, D. (1995), *Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system*, Statistics and probability letters, 25, 3: 201-212

[8] Cai, T. and Brown, L. (1998), *Wavelet shrinkage for nonequispaced samples*, Annals of statistics, 26, 5: 1783-1799

[9] Casdagli, M. (1989), *Nonlinear prediction of chaotic time series*, Physica D: Nonlinear Phenomena, 35, 3: 335-356

[10] Chorro, C., Guégan, D. and Ielpo, F. (2012), *Option pricing for GARCH-type models with generalized hyperbolic innovations*, Quantitative Finance, 12, 7: 1079-1094

[11] Daubechies, I., Guskov, I., Schröder, P. and Sweldens, W. (1999), *Wavelets on irregular point sets*, Philosophical transactions of the Royal Society of London, series A: mathematical, physical and engineering sciences, 357, 1760: 2397-2413

[12] Donoho, D. (1992), *Interpolating wavelet transforms*, Preprint, Department of Statistics, Stanford University, 2, 3

[13] Donoho, D. (1995), *Denoising by soft thresholding*, IEEE transactions on information theory, 41, 3: 613-627

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

[15] Donoho, D. and Johnstone, I. (1994), *Minimax risk over l_p-balls for l_p-error*, Probability theory and related fields, 99, 2: 277-303

[16] Donoho, D. and Johnstone, I. (1995), *Adapting to unknown smoothness via wavelet shrinkage*, Journal of the American statistical association, 90, 432: 1200-1244

[17] Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995), *Wavelet shrinkage: Asymptopia?*, Journal of the Royal Statistical Society, Series B, 57, 2: 301-369

[18] Donoho, D. and Yu, T.P. (2000), *Nonlinear pyramid transforms based on median-interpolation*, SIAM Journal on mathematical analysis, 31, 5: 1030-1061

[19] Eckmann, J.-P. and Ruelle, D. (1985), *Ergodic theory of chaos and strange attractors*, Reviews of modern physics, 57, 3: 617-656

[20] Engl, H., Hanke, M. and Neubauer, A. (1996), *Regularization of inverse problems*, Kluwer, Dordrecht, Netherlands

[21] Engle, R. (1982), *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, Econometrica, 50, 4: 987-1008

[22] Faggini, M. (2014), *Chaotic time series analysis in economics: balance and perspectives*, Chaos, 24, 4: 042101

[23] Faggini, M. and Parziale, A. (2012), *The Failure of Economic Theory. Lessons from Chaos Theory*, Modern economy, 3, 1: 1-10

[24] Farmer, M. (2013), *Robust pre-attentive attention direction using chaos theory for video surveillance*, Applied Mathematics, 4: 43-55

[25] Farmer, J. and Sidorowich, J. (1991), *Optimal shadowing and noise reduction*, Physica D: Nonlinear Phenomena, 47, 3: 373-392

[26] Gao, J., Sultan, H., Hu, J. and Tung, W.-W. (2010), *Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison*, IEEE Signal Processing Letters, 17, 3: 237-240

[27] Garcin, M. and Guégan, D. (2012), *Extreme values of random or chaotic discretization steps and connected networks*, Applied Mathematical Sciences, 119, 6: 5901-5926

[28] Garcin, M. and Guégan, D. (2014), *Probability density of the empirical wavelet coefficients of a noisy chaos*, Physica D: Nonlinear Phenomena, 276: 28-47

[29] Granger, C. and Joyeux, R. (1980), *An introduction to long-memory time series models and fractional differencing*, Journal of time series analysis, 1, 1: 15-29

[30] Guégan, D. (2000), *A new model: the k-factor GIGARCH process*, Journal of Signal Processing, 4, 3: 265-271

[31] Guégan, D. (2003), *Les Chaos en finance : approche statistique*, Economica, Paris, France

[32] Guégan, D. (2008), *Effect of noise filtering on predictions: on the routes of chaos*, Brussels Economic Review, 53, 2: 255-272

[33] Guégan, D. and Hoummiya, K. (2005), *Denoising with wavelets method in chaotic time series: application in climatology, energy and finance* in *Noise and fluctuations in econophysics and finance*, ed. by D. Abbott, J.P. Bouchaud, X. Gabaix, J.L. McCaulay, Proc. SPIE 5848: 174-185

[34] Guégan, D. and Mercier, L. (2005), *Prediction in chaotic time series: methods and comparisons with an application to financial intra-day data*, The European Journal of Finance, 11, 2: 137-150

[35] Hall, P. and Patil, P. (1996), *Effect of threshold rules on performance of wavelet-based curve estimators*, Statistica Sinica, 6, 2: 331-345

[36] Hammel, S. (1990), *A noise reduction method for chaotic systems*, Physics letters A, 148, 8: 421-428

[37] Hao, B.-L. (Ed.) (1988), *Directions in chaos*, volume 2, World Scientific Publishing, Singapore

[38] Hosking, J. (1981), *Fractional differencing*, Biometrika, 68, 1: 165-176

[39] Jansen, M. (2013), *Multiscale local polynomial smoothing in a lifted pyramid for non-equispaced data*, IEEE transactions on signal processing, 61, 3: 545-555

[40] Jansen, M., Nason, G. and Silverman, B. (2009), *Multiscale methods for data on graphs and irregular multidimensional situations*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71, 1: 97-125

[41] Kostelich, E.J. and Yorke, J.A. (1988), *Noise reduction in dynamical systems*, Physical Review A, 38, 3: 1649-1652

[42] Kostelich, E.J. and Yorke, J.A. (1990), *Noise reduction: finding the simplest dynamical system consistent with the data*, Physica D: Nonlinear Phenomena, 41, 2: 183-196

[43] Mallat, S. (2000), *Une exploration des signaux en ondelettes*, Ellipses, Éditions de l'École Polytechnique, Paris, France

[44] Marron, J.S., Adak, S., Johnstone, I., Neumann, M.H. and Patil, P. (1998), *Exact risk analysis of wavelet regression*, Journal of computational and graphical statistics, 7, 3: 278-309

[45] Nason, G.P. (1996), *Wavelet shrinkage using cross-validation*, Journal of the royal statistical society, series B, 58: 463-479

[46] Roy, M., Kumar, V., Kulkarni, B., Sanderson, J., Rhodes, M. and Stappen, M. (1999), *Simple denoising algorithm using wavelet transform*, AIChE Journal, 45, 11: 2461-2466

[47] Sauer, T. (1992), *A noise reduction method for signal from non linear systems*, Physica D: Nonlinear Phenomena, 58, 1: 193-201

[48] Shin, K., Hammond, J. and White, P. (1999), *Iterative SVD method for noise reduction of low-dimensional chaotic time series*, Mechanical systems and signal processing, 13, 1: 115-124

[49] Stein, C. (1981), *Estimation of the mean of a multivariate normal distribution*, The Annals of Statistics, 9, 6: 1135-1151

[50] Sweldens, W. (1998), *The lifting scheme: A construction of second generation wavelets*, SIAM Journal on Mathematical Analysis, 29, 2: 511-546

[51] Tong, H. and Lim, K.S. (1980), *Threshold autoregression, limit cycles and cyclical data*, Journal of the Royal Statistical Society, Series B (Methodological), 42, 3: 245-292

[52] Vidakovic, B. (1998), *Nonlinear wavelet shrinkage with Bayes rules and Bayes factors*, Journal of the American Statistical Association, 93, 441: 173-179

[53] Woodward, W., Cheng, Q. and Gray, H. (1998), *A k-factor GARMA long-memory model*, Journal of time series analysis, 19, 4: 485-504

[54] Xu, Y., Weaver, J., Healy, D. and Lu, J. (1994), *Wavelet transform domain filters: a spatially selective noise filtration technique*, IEEE transactions on image processing, 3, 6: 747-758

A Lemmas

We introduce three lemmas that we will use several times in the proofs of the theorems of the present article.

Lemma 1. *Let $X_1, ..., X_N$ be independent random variables with zero mean and let g be a function in $L^2(\mathbb{R})$. Then*

$$\mathbb{E}\left[\langle X, g\rangle^2\right] = \langle \mathbb{E}[X^2], g^2\rangle.$$

Proof. We can write:

$$\begin{array}{rcl} \mathbb{E}\left[\langle X, g\rangle^2\right] & = & \mathbb{E}\left[\sum_{m=1}^N \sum_{n=1}^N X_n X_m g(n) g(m)\right] \\ & = & \sum_{m=1}^N \sum_{n=1}^N \mathbb{E}\left[X_n X_m\right] g(n) g(m) \\ & = & \sum_{n=1}^N \mathbb{E}\left[X_n^2\right] g(n)^2, \end{array}$$

because of the independence of the random variables. Finally, that last line is equal to $\langle \mathbb{E}[X^2], g^2\rangle$. □

Lemma 2. *Let $X_1, ..., X_N$ be independent identically distributed random variables with zero mean, variance σ^2 and kurtosis κ^4 and let g be a positive function in $L^2(\mathbb{R})$ such that $\langle g, g\rangle = \mathcal{G}$. Then*

$$\mathbb{E}\left[\langle X^2, g\rangle^2\right] \leq \sigma^4 \kappa^4 \mathcal{G}.$$

Proof. We can write:

$$\begin{array}{rcl} \mathbb{E}\left[\langle X^2, g\rangle^2\right] & = & \mathbb{E}\left[\sum_{m=1}^N \sum_{n=1}^N X_n^2 X_m^2 g(n) g(m)\right] \\ & = & \sum_{m=1}^N \sum_{n=1}^N \mathbb{E}\left[X_n^2 X_m^2\right] g(n) g(m) \\ & = & \sum_{n=1}^N \mathbb{E}\left[X_n^4\right] g(n)^2 + \sum_{m=1}^N \sum_{n=1}^N \mathbb{E}\left[X_n^2\right] \mathbb{E}\left[X_m^2\right] g(n) g(m) - \sum_{n=1}^N \mathbb{E}\left[X_n^2\right]^2 g(n)^2, \end{array}$$

because of the independence of the random variables. Finally, we get:

$$\begin{array}{rcl} \mathbb{E}\left[\langle X^2, g\rangle^2\right] & = & \langle \mathbb{E}\left[X^4\right] - \mathbb{E}\left[X^2\right]^2, g^2\rangle + \langle \mathbb{E}\left[X^2\right], g\rangle^2 \\ & = & \langle \sigma^4\left(\kappa^4 - 1\right), g^2\rangle + \langle \sigma^2, g\rangle^2 \\ & = & \sigma^4\left(\kappa^4 - 1\right)\mathcal{G} + \langle \sigma^2, g\rangle^2 \\ & \leq & \sigma^4\left((\kappa^4 - 1)\mathcal{G} + \langle g, g\rangle\right) = \sigma^4 \kappa^4 \mathcal{G}, \end{array}$$

where we used the Cauchy-Schwarz inequality in the last line. □

Lemma 3. *Let assume (A1), (A2), (A3), (A4) and (A1'). For a linear filter $F = (f_l)_l$ we define the estimator $\tilde{z}$ of z from $z^{\mathcal{L}}$, introduced in equation* (8)*, and the estimator $\tilde{\tilde{z}}$ of z from z^{ε}, introduced in equation* (9)*. The error between these estimators and the true function z is respectively, conditionally to I_N:*

$$\left\{ \begin{array}{rcl} r^{\mathcal{L}}(z) & = & \mathbb{E}\left[\sum_{l=1}^N \left(f_l \hat{z}_l^{\mathcal{L}} - \hat{z}_l\right)^2 \middle| I_N\right] \\ r(z) & = & \mathbb{E}\left[\sum_{l=1}^N \left(f_l \hat{z}_l^{\varepsilon} - \hat{z}_l\right)^2 \middle| I_N\right]. \end{array}\right.$$

Then:

$$\left|r(z) - r^{\mathcal{L}}(z)\right| \leq \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l \left(f_l K' \sigma^2\right)^2.$$

Proof. Thanks to the triangle inequality, we have:

$$\begin{array}{rcl} \left|r(z) - r^{\mathcal{L}}(z)\right| & = & \left|\mathbb{E}\left[\sum_{l=1}^{N} \left(f_l \hat{z}_l^{\varepsilon} - \hat{z}_l\right)^2 \middle| I_N\right] - \mathbb{E}\left[\sum_{l=1}^{N} \left(f_l \hat{z}_l^{\mathcal{L}} - \hat{z}_l\right)^2 \middle| I_N\right]\right| \\ & \leq & \mathbb{E}\left[\sum_{l=1}^{N} \left(f_l \hat{z}_l^{\varepsilon} - f_l \hat{z}_l^{\mathcal{L}}\right)^2 \middle| I_N\right], \end{array}$$

Then, according to the Taylor expansion made in [27], since the filter is linear, we get:

$$\left|r(z) - r^{\mathcal{L}}(z)\right| \leq \mathbb{E}\left[\sum_{l=1}^{N} f_l^2 \frac{1}{4} \max\left(|z''|^2\right) \langle |\varepsilon_\cdot^\star|^2, |\psi_l(u_\cdot)| V_\cdot \rangle^2 \middle| I_N\right].$$

Finally, because z' is a Lipschitz continuous function of constant K', we get:

$$\begin{array}{rcl} \left|r(z) - r^{\mathcal{L}}(z)\right| & \leq & \sum_{l=1}^{N} f_l^2 \frac{(K')^2}{4} \mathbb{E}\left[\langle |\varepsilon_\cdot^\star|^2, |\psi_l(u_\cdot)| V_\cdot \rangle^2 \middle| I_N\right] \\ & \leq & \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l \left(f_l K' \sigma^2\right)^2, \end{array}$$

using Lemma 2. □

B Proof of Theorem 1

Proof. (i) This part of the theorem is a direct consequence of Wiener's Theorem [43][23], for an independent but not identically distributed noise, $(\varepsilon_n - z'(u_n)\varepsilon_n^\star)_n$, whose conditional variance in the wavelet basis is:

$$\begin{array}{rcl} \mathbb{E}\left[\langle \varepsilon_\cdot - z'(u_\cdot)\varepsilon_\cdot^\star, \psi_l(u_\cdot) V_\cdot \rangle^2 \middle| I_N\right] & = & \mathbb{E}\left[\langle \varepsilon_\cdot, \psi_l(u_\cdot) V_\cdot \rangle^2 \middle| I_N\right] + \mathbb{E}\left[\langle z'(u_\cdot)\varepsilon_\cdot^\star, \psi_l(u_\cdot) V_\cdot \rangle^2 \middle| I_N\right] \\ & = & \langle \mathbb{E}\left[\varepsilon_\cdot^2\right], \psi_l(u_\cdot)^2 V_\cdot^2 \rangle + \langle \mathbb{E}\left[\left(z'(u_\cdot)\varepsilon_\cdot^\star\right)^2 \middle| I_N\right], \psi_l(u_\cdot)^2 V_\cdot^2 \rangle \\ & = & \langle \sigma^2, \psi_l^2 V^2 \rangle + \sigma^2 \langle \mathbb{E}\left[z'^2 \middle| I_N\right], \psi_l^2 V^2 \rangle \\ & = & \sigma^2 \left(\mathcal{V}_l + \Gamma_l\right), \end{array}$$

where we used the independence of $\varepsilon_\cdot$ and $\varepsilon_\cdot^\star$ in the first line and Lemma 1 in the second line. In particular, as soon as z is a Lipschitz continuous function, with Lipschitz constant $K > 0$, then, for all x in the support of z, $|z'(x)| \leq K$ and $\Gamma_l \leq K^2 \mathcal{V}_l$. Therefore, conditionally to I_N:

$$r_b^{\mathcal{L}}(\pi) = \sum_{l=1}^{N} \frac{\beta_l^2 \sigma^2}{\frac{\beta_l^2}{\mathcal{V}_l + \Gamma_l} + \sigma^2} \leq \sum_{l=1}^{N} \frac{\beta_l^2 \sigma^2 \mathcal{V}_l (1 + K^2)}{\beta_l^2 + \sigma^2 \mathcal{V}_l (1 + K^2)}.$$

(ii) We can use directly Lemma 3, because z' is a Lipschitz continuous function of constant K':

$$\left|r_b(\pi) - r_b^{\mathcal{L}}(\pi)\right| \leq \frac{\kappa^4}{4} \sum_{l=1}^{N} \mathcal{V}_l (f_l^{\mathcal{L}} K' \sigma^2)^2.$$

□

[23] Theorem 10.2 page 434.

C Proof of Theorem 2

Proof. (i) In the oracular case, the wavelet coefficients $(\hat{z}_l)_l$ of the signal are deterministic. Therefore, for any filter $F = (f_l)_l$, the error made, conditionally to I_N, is:

$$\begin{array}{rcl} r\left((f_l)_l\right) & = & \mathbb{E}\left[\sum_{l=1}^{N}\left(\hat{z}_l - f_l\hat{z}_l^{\mathcal{L}}\right)^2 \Big| I_N\right] \\ & = & \sum_{l=1}^{N}\hat{z}_l^2 + f_l^2\mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}}\right)^2 \Big| I_N\right] - 2f_l\hat{z}_l\mathbb{E}\left[\hat{z}_l^{\mathcal{L}}\big| I_N\right] \\ & = & \sum_{l=1}^{N}\hat{z}_l^2\left(1-f_l\right)^2 + f_l^2\sigma^2\left(\mathcal{V}_l+\Gamma_l\right). \end{array}$$

That error is minimal when $f_l = f_l^{\mathcal{L}}$, where $f_l^{\mathcal{L}} = \frac{\hat{z}_l^2}{\hat{z}_l^2+\sigma^2(\mathcal{V}_l+\Gamma_l)}$. As a consequence, the minimal error $r_a^{\mathcal{L}}$ is, conditionally to I_N:

$$\begin{array}{rcl} r_a^{\mathcal{L}}(z) & = & \inf_{(f_l)_l} r\left((f_l)_l\right) \\ & = & r\left((f_l^{\mathcal{L}})_l\right) \\ & = & \sum_{l=1}^{N}\hat{z}_l^2\left(1-f_l^{\mathcal{L}}\right)^2 + \left(f_l^{\mathcal{L}}\right)^2\sigma^2\left(\mathcal{V}_l+\Gamma_l\right) \\ & = & \sum_{l=1}^{N}\frac{\hat{z}_l^2\sigma^2(\mathcal{V}_l+\Gamma_l)}{\hat{z}_l^2+\sigma^2(\mathcal{V}_l+\Gamma_l)}. \end{array}$$

(ii) We can use directly Lemma 3, because z' is a Lipschitz continuous function of constant K':

$$\left| r_a(z) - r_a^{\mathcal{L}}(z)\right| \leq \frac{\kappa^4}{4}\sum_{l=1}^{N}\mathcal{V}_l(f_l^{\mathcal{L}}K'\sigma^2)^2.$$

□

D Proof of Theorem 3

Proof. (i) The wavelet coefficient of the linearised noisy evolution function is

$$\hat{z}_l^{\mathcal{L}} = \hat{z}_l + \langle \varepsilon_. - z'(u_.)\varepsilon_.^\star, \psi_l(u_.)V_.\rangle.$$

Therefore, the error between the filtered linearised coefficients and the true ones, conditionally to I_N, is defined by:

$$\begin{array}{rcl} r_p^{\mathcal{L}}(z) & = & \mathbb{E}\left[\sum_{l=1}^{N}\left(\hat{z}_l - f_l^{\mathcal{L}}\hat{z}_l^{\mathcal{L}}\right)^2 \Big| I_N\right] \\ & = & \sum_{l=1}^{N}\hat{z}_l^2 + \left(f_l^{\mathcal{L}}\right)^2\mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}}\right)^2 \Big| I_N\right] - 2f_l^{\mathcal{L}}\hat{z}_l\mathbb{E}\left[\hat{z}_l^{\mathcal{L}}\big| I_N\right] \\ & = & \sum_{l=1}^{N}\hat{z}_l^2\left(1-f_l^{\mathcal{L}}\right)^2 + \left(f_l^{\mathcal{L}}\right)^2\sigma^2\left(\mathcal{V}_l+\Gamma_l\right), \end{array} \tag{12}$$

because, according to Lemma 1:

$$\begin{array}{rcl} \mathbb{E}\left[\langle \varepsilon_. - z'(u_.)\varepsilon_.^\star, \psi_l(u_.)V_.\rangle^2 \big| I_N\right] & = & \mathbb{E}\left[\langle(\varepsilon_. - z'(u_.)\varepsilon_.^\star)^2, \psi_l(u_.)^2V_.^2\rangle \big| I_N\right] \\ & = & \langle \mathbb{E}\left[(\varepsilon_. - z'(u_.)\varepsilon_.^\star)^2 \big| I_N\right], \psi_l(u_.)^2V_.^2\rangle \\ & = & \sigma^2\left(\mathcal{V}_l+\Gamma_l\right), \end{array}$$

where $\Gamma_l = \langle z'^2, \psi_l^2V^2\rangle$. Then, from equation (12), we conclude that, conditionally to I_N:

$$r_p^{\mathcal{L}}(z) = \sum_{l=1}^{N}\min\left(\hat{z}_l^2, \sigma^2\left(\mathcal{V}_l+\Gamma_l\right)\right). \tag{13}$$

In another side, for all $x, y > 0$:

$$\frac{1}{2}\min(x,y) \leq \frac{xy}{x+y} \leq \min(x,y),$$

then, we get from equation (13):

$$\frac{1}{2}r_p^{\mathcal{L}}(z) \leq r_a^{\mathcal{L}}(z) \leq r_p^{\mathcal{L}}(z).$$

(ii) For the estimator $\tilde{\tilde{z}}$, the error is $r_p(z)$ and is such that:

$$\left|r_p(z) - r_p^{\mathcal{L}}(z)\right| = \left|\mathbb{E}\left[\sum_{l=1}^{N}\left(f_l^{\mathcal{L}}\hat{z}_l^{\varepsilon} - \hat{z}_l\right)^2 \middle| I_N\right] - \mathbb{E}\left[\sum_{l=1}^{N}\left(f_l^{\mathcal{L}}\hat{z}_l^{\mathcal{L}} - \hat{z}_l\right)^2 \middle| I_N\right]\right|.$$

We can use directly Lemma 3, because z' is a Lipschitz continuous function of constant K':

$$\left|r_p(z) - r_p^{\mathcal{L}}(z)\right| \leq \frac{\kappa^4}{4}\sum_{l=1}^{N}\mathcal{V}_l(f_l^{\mathcal{L}}K'\sigma^2)^2.$$

(iii) We note $R_p = \frac{\kappa^4}{4}\sum_{l=1}^{N}\mathcal{V}_l\left(f_l^{\mathcal{L}}K'\sigma^2\right)^2$ and $R_a = \frac{\kappa^4}{4}\sum_{l=1}^{N}\mathcal{V}_l\left[\frac{\hat{z}_l^2}{\hat{z}_l^2+\sigma^2(\mathcal{V}_l+\Gamma_l)}\right]^2\left(K'\sigma^2\right)^2$. On the one hand if $f_l^{\mathcal{L}} = 1$ then, by definition of $f_l^{\mathcal{L}}$:

$$\frac{\hat{z}_l^2}{\hat{z}_l^2 + \sigma^2(\mathcal{V}_l + \Gamma_l)} \geq \frac{\hat{z}_l^2}{2\hat{z}_l^2} = \frac{1}{2} = \frac{1}{2}f_l^{\mathcal{L}}. \tag{14}$$

On the other hand if $f_l^{\mathcal{L}} = 0$ we have:

$$\frac{\hat{z}_l^2}{\hat{z}_l^2 + \sigma^2(\mathcal{V}_l + \Gamma_l)} \geq 0 = \frac{1}{2}f_l^{\mathcal{L}}. \tag{15}$$

Therefore, thanks to equations (14) and (15), we get:

$$\frac{1}{2}R_p \leq R_a. \tag{16}$$

Using the item (ii) of Theorem 2 and the items (i) and (ii) of Theorem 3 jointly with the triangle inequality, we can write:

$$\begin{array}{rcl} r_p(z) & \leq & r_p^{\mathcal{L}}(z) + R_p \\ & \leq & 2r_a^{\mathcal{L}}(z) + R_p \\ & \leq & 2r_a(z) + 2R_a + R_p \end{array} \tag{17}$$

and

$$\begin{array}{rcl} r_a(z) & \leq & r_a^{\mathcal{L}}(z) + R_a \\ & \leq & r_p^{\mathcal{L}}(z) + R_a \\ & \leq & r_p(z) + R_p + R_a. \end{array} \tag{18}$$

Finally, putting equations (16), (17) and (18) together, we obtain the double inequality:

$$\frac{1}{2}r_p(z) - 2R_a \leq r_a(z) \leq r_p(z) + 3R_a.$$

□

E Proof of Theorem 4

Proof. (i) This part of the theorem is a direct consequence of a Theorem of Donoho and Johnstone [14], as cited in [43][24], for an independent but not identically distributed noise, $(\varepsilon_n - z'(u_n)\varepsilon_n^\star)_n$, whose conditional variance in the wavelet basis is:

$$\mathbb{E}\left[\langle \varepsilon_. - z'(u_.)\varepsilon_.^\star, \psi_l(u_.)V_.\rangle^2 \middle| I_N\right] = \sigma^2\left(\mathcal{V}_l + \Gamma_l\right),$$

according to Lemma 1. Therefore, conditionally to I_N:

$$r_t^{\mathcal{L}}(z, (\lambda_l^\star)_l) \leq (2\log N + 1)\left(\sigma^2\left(\bar{\mathcal{V}} + \bar{\Gamma}\right) + r_p^{\mathcal{L}}(z)\right).$$

[24] Theorem 10.4 page 452.

(ii) Let $\lambda > 0$. We define $R(\lambda)$ by the equation:

$$R(\lambda) = 2\int_0^{+\infty} x^2 \phi(x+\lambda)dx,$$

where ϕ is the Gaussian probability density function. We can see the integrated function as a product of three functions and we can integrate it by parts because $-\phi(x)$ is the primitive integral of $x\phi(x)$:

$$\begin{array}{rcl} R(\lambda) & = & 2\exp\left(-\frac{\lambda^2}{2}\right)\int_0^{+\infty} x \times x\phi(x) \times \exp(-\lambda x)dx \\ & = & 2\exp\left(-\frac{\lambda^2}{2}\right)\left\{\int_0^{+\infty}\phi(x)\exp(-\lambda x)dx - \lambda\int_0^{+\infty} x\phi(x)\exp(-\lambda x)dx\right\} \\ & \leq & 2\exp\left(-\frac{\lambda^2}{2}\right)\left\{\int_0^{+\infty}\phi(x)dx - \lambda\int_0^{+\infty} x\phi(x)(1-\lambda x)dx\right\} \\ & \leq & 2\exp\left(-\frac{\lambda^2}{2}\right)\left\{\frac{1}{2} - \frac{\lambda}{2} + \frac{\lambda^2}{2}\right\} \\ & \leq & \exp\left(-\frac{\lambda^2}{2}\right)\left(1+\lambda^2\right). \end{array} \tag{19}$$

From [43][25], we know that, for any threshold $\lambda_l > 0$, conditionally to I_N:

$$r_t^{\mathcal{L}}(z,\lambda_l) \leq \sum_{l=1}^{N}\sigma^2(\mathcal{V}_l+\Gamma_l)R\left(\frac{\lambda_l}{\sigma\sqrt{\mathcal{V}_l+\Gamma_l}}\right) + \sum_{l=1}^{N}\left(1+\frac{\lambda_l^2}{\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)r_p^{\mathcal{L}}(z)_l, \tag{20}$$

where $r_p^{\mathcal{L}}(z)_l = \min((\hat{z}_l^{\mathcal{L}})^2, \sigma^2(\mathcal{V}_l+\Gamma_l))$. Then, conditionally to I_N, we get, using equation (19) in equation (20):

$$r_t^{\mathcal{L}}(z,\lambda_l) \leq \sum_{l=1}^{N}\sigma^2(\mathcal{V}_l+\Gamma_l)\exp\left(-\frac{\lambda_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)\left(1+\frac{\lambda_l^2}{\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)+\sum_{l=1}^{N}\left(1+\frac{\lambda_l^2}{\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)r_p^{\mathcal{L}}(z)_l. \tag{21}$$

Let $\lambda_l = \lambda_l^\circ = \sigma\sqrt{2\mathcal{V}_l\left(1+K^2\right)\log N}$. Hence, equation (21) gives, conditionally to I_N:

$$\begin{array}{rcl} r_t^{\mathcal{L}}(z,\lambda_l^\circ) & \leq & \sum_{l=1}^{N}\sigma^2(\mathcal{V}_l+\Gamma_l)\exp\left(-\frac{\mathcal{V}_l\left(1+K^2\right)\log N}{\mathcal{V}_l+\Gamma_l}\right)\left(1+\frac{\mathcal{V}_l\left(1+K^2\right)2\log N}{\mathcal{V}_l+\Gamma_l}\right)+\sum_{l=1}^{N}\left(1+\frac{\mathcal{V}_l\left(1+K^2\right)2\log N}{\mathcal{V}_l+\Gamma_l}\right)r_p^{\mathcal{L}}(z)_l \\ & \leq & \sum_{l=1}^{N}\sigma^2(\mathcal{V}_l+\Gamma_l)\frac{1}{N}\left(1+2\left(1+K^2\right)\log N\right)+\sum_{l=1}^{N}\left(1+2\left(1+K^2\right)\log N\right)r_p^{\mathcal{L}}(z)_l \\ & \leq & \left(1+2\left(1+K^2\right)\log N\right)\left(\sigma^2\left(\mathcal{V}+\bar{\Gamma}\right)+r_p^{\mathcal{L}}(z)\right). \end{array}$$

(iii) We have to adapt Lemma 3 to a non-linear filter. If f_l is a soft-threshold filter then:

$$\left(f_l\hat{z}_l^{\varepsilon} - f_l\hat{z}_l^{\mathcal{L}}\right)^2 \leq \left(\hat{z}_l^{\varepsilon} - \hat{z}_l^{\mathcal{L}}\right)^2.$$

Thus, using the triangle inequality and a Taylor expansion we get, conditionally to I_N:

$$\left|r(z) - r^{\mathcal{L}}(z)\right| \leq \mathbb{E}\left[\sum_{l=1}^{N}\frac{1}{4}\max\left(|z''|^2\right)\langle|\varepsilon_\cdot^\star|^2, |\psi_l(u_\cdot)|V_\cdot\rangle^2\middle| I_N\right].$$

Finally, because z' is a Lipschitz continuous function of constant K' and $\kappa^4 = 3$ for a Gaussian noise we get:

$$\begin{array}{rcl} \left|r(z) - r^{\mathcal{L}}(z)\right| & \leq & \frac{\kappa^4}{4}\sum_{l=1}^{N}\mathcal{V}_l\left(K'\sigma^2\right)^2 \\ & \leq & \frac{3}{4}N\bar{\mathcal{V}}\left(K'\sigma^2\right)^2, \end{array}$$

using Lemma 2.

□

[25] Page 449.

F Proof of Theorem 5

The following proof is inspired by the bias calculation of Stein's unbiased risk estimator for a signal with a linear noise impact [49, 16, 43]. We contribute to adapt it to non-linear noise impact.

Proof. Let g be the function defined by:

$$g : x \mapsto -x\mathbf{1}_{\{|x|<\lambda\}} + \lambda\left(\mathbf{1}_{\{x\leq -\lambda\}} - \mathbf{1}_{\{x\geq\lambda\}}\right).$$

Thus, the filtered wavelet coefficient obtained by soft-thresholding is $h(\hat{z}_l^\varepsilon) = \hat{z}_l^\varepsilon + g(\hat{z}_l^\varepsilon)$. Then, we get the following expression for the bias of the estimate of the error function conditionally to the grid:

$$\begin{aligned}
\mathbb{E}\left[\tilde{r}_t(z,\lambda) - r_t(z,\lambda)\middle| I_N\right] &= \mathbb{E}\left[\sum_{l=1}^N \mathcal{S}_l(\hat{z}_l^\varepsilon) - \sum_{l=1}^N \left\{h(\hat{z}_l^\varepsilon) - \hat{z}_l\right\}^2 \middle| I_N\right] \\
&= \sum_{l=1}^N \mathbb{E}\left[\mathcal{S}_l(\hat{z}_l^\varepsilon) - \left\{g(\hat{z}_l^\varepsilon) + (\hat{z}_l^{\mathcal{L}} - \hat{z}_l) + (\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})\right\}^2 \middle| I_N\right] \\
&= \sum_{l=1}^N \left\{\mathbb{E}\left[\mathcal{S}_l(\hat{z}_l^\varepsilon)\middle| I_N\right] - \mathbb{E}\left[g(\hat{z}_l^\varepsilon)^2\middle| I_N\right] - \mathbb{E}\left[(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)^2\middle| I_N\right] - 2\mathbb{E}\left[g(\hat{z}_l^\varepsilon)(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)\middle| I_N\right]\right. \\
&\quad \left. -\mathbb{E}\left[(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})^2\middle| I_N\right] - 2\mathbb{E}\left[g(\hat{z}_l^\varepsilon)(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})\middle| I_N\right] - 2\mathbb{E}\left[(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})\middle| I_N\right]\right\}.
\end{aligned} \tag{22}$$

We need to develop the expression of $\mathbb{E}\left[\mathcal{S}_l(\hat{z}_l^\varepsilon)\middle| I_N\right]$. It requires the intermediate calculation of $\mathbb{E}\left[g(\hat{z}_l^{\mathcal{L}})(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)\middle| I_N\right]$. According to [28] since the noise is Gaussian the wavelet coefficient $\hat{z}_l^{\mathcal{L}}$ is a Gaussian random variable of mean $\hat{z}_l$ and of variance $\sigma^2(\mathcal{V}_l+\Gamma_l)$. Therefore, by integrating by part we get:

$$\begin{aligned}
\mathbb{E}\left[g(\hat{z}_l^{\mathcal{L}})(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)\right] &= \frac{1}{\sigma\sqrt{2\pi(\mathcal{V}_l+\Gamma_l)}}\int_{\mathbb{R}} g(x)(x-\hat{z}_l)\exp\left(-\frac{(x-\hat{z}_l)^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)dx \\
&= \frac{\sigma^2(\mathcal{V}_l+\Gamma_l)}{\sigma\sqrt{2\pi(\mathcal{V}_l+\Gamma_l)}}\int_{\mathbb{R}} g'(x)\exp\left(-\frac{(x-\hat{z}_l)^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)dx \\
&= \sigma^2(\mathcal{V}_l+\Gamma_l)\mathbb{E}\left[g'(\hat{z}_l^{\mathcal{L}})\right],
\end{aligned} \tag{23}$$

where g' is the derivative of g. Then, conditionally to I_N and using equation (23), we get:

$$\mathbb{E}\left[g\left(\hat{z}_l^{\mathcal{L}}\right)^2 + 2g\left(\hat{z}_l^{\mathcal{L}}\right)\left(\hat{z}_l^{\mathcal{L}} - \hat{z}_l\right) + \left(\hat{z}_l^{\mathcal{L}} - \hat{z}_l\right)^2\middle| I_N\right] = \mathbb{E}\left[\mathcal{S}\left(\hat{z}_l^{\mathcal{L}}\right)\middle| I_N\right], \tag{24}$$

because, for $x\in\mathbb{R}$, $g'(x) = \mathbf{1}_{\{|x|<\lambda\}}$ and $g(x)^2 = x^2\mathbf{1}_{\{|x|<\lambda\}} + \lambda^2\mathbf{1}_{\{|x|\geq\lambda\}}$. Therefore, using jointly equations (22) and (24) we get, conditionally to the I_N:

$$\begin{aligned}
\left|\mathbb{E}\left[\tilde{r}_t(z,\lambda) - r_t(z,\lambda)\middle| I_N\right]\right| &= \left|\sum_{l=1}^N\left\{\mathbb{E}\left[g(\hat{z}_l^{\mathcal{L}})^2 - g(\hat{z}_l^\varepsilon)^2\middle| I_N\right] + 2\mathbb{E}\left[\left(g(\hat{z}_l^{\mathcal{L}}) - g(\hat{z}_l^\varepsilon)\right)(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)\middle| I_N\right]\right.\right. \\
&\quad \left.\left. -\mathbb{E}\left[(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})^2\middle| I_N\right] - 2\mathbb{E}\left[g(\hat{z}_l^\varepsilon)(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})\middle| I_N\right] - 2\mathbb{E}\left[(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})\middle| I_N\right]\right\}\right| \\
&\leq \sum_{l=1}^N\left\{\mathbb{E}\left[\left|g(\hat{z}_l^{\mathcal{L}})^2 - g(\hat{z}_l^\varepsilon)^2\right|\middle| I_N\right] + 2\mathbb{E}\left[\left|\left(g(\hat{z}_l^{\mathcal{L}}) - g(\hat{z}_l^\varepsilon)\right)(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)\right|\middle| I_N\right]\right. \\
&\quad \left. +\mathbb{E}\left[(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})^2\middle| I_N\right] + 2\mathbb{E}\left[\left|g(\hat{z}_l^\varepsilon)(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})\right|\middle| I_N\right] + 2\mathbb{E}\left[\left|(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)(\hat{z}_l^\varepsilon - \hat{z}_l^{\mathcal{L}})\right|\middle| I_N\right]\right\}.
\end{aligned} \tag{25}$$

thanks to the triangle inequality. Moreover, for all $x, y\in\mathbb{R}$ we have $|g(x) - g(y)| \leq |x-y|$, $|g(x)| \leq \lambda$ and, using a Taylor development:

$$|g(x)^2 - g(y)^2| \leq 2|x-y|\max_z(|g(z)g'(z)|) = 2\lambda|x-y|.$$

Therefore, conditionally to I_N, we get from equation (25):

$$\begin{aligned}
&\left|\mathbb{E}\left[\tilde{r}_t(z,\lambda) - r_t(z,\lambda)\middle| I_N\right]\right| \\
&\leq \sum_{l=1}^N\left\{4\lambda\mathbb{E}\left[\left|\hat{z}_l^{\mathcal{L}} - \hat{z}_l^\varepsilon\right|\middle| I_N\right] + 4\mathbb{E}\left[\left|\left(\hat{z}_l^{\mathcal{L}} - \hat{z}_l^\varepsilon\right)\left(\hat{z}_l^{\mathcal{L}} - \hat{z}_l\right)\right|\middle| I_N\right] + \mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}} - \hat{z}_l^\varepsilon\right)^2\middle| I_N\right]\right\} \\
&\leq \sum_{l=1}^N\left\{4\lambda\mathbb{E}\left[\left|\hat{z}_l^{\mathcal{L}} - \hat{z}_l^\varepsilon\right|\middle| I_N\right] + 4\sqrt{\mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}} - \hat{z}_l^\varepsilon\right)^2\middle| I_N\right]\mathbb{E}\left[(\hat{z}_l^{\mathcal{L}} - \hat{z}_l)^2\middle| I_N\right]} + \mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}} - \hat{z}_l^\varepsilon\right)^2\middle| I_N\right]\right\},
\end{aligned} \tag{26}$$

where we used the Cauchy-Schwarz inequality. According to [28] the wavelet coefficient $\hat{z}_l^{\mathcal{L}}$ is a Gaussian random variable of mean $\hat{z}_l$ and of variance $\sigma^2(\mathcal{V}_l+\Gamma_l)$ and, according both to [28] and Lemma 2, $\mathbb{E}\left[\left|\hat{z}_l^{\mathcal{L}}-\hat{z}_l^{\varepsilon}\right|\middle| I_N\right]\leq \mathcal{U}_l K'\sigma^2/2$ and $\mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}}-\hat{z}_l^{\varepsilon}\right)^2\middle| I_N\right]\leq 3\mathcal{V}_l\left(K'\sigma^2\right)^2/4$. Thus, equation (26) leads to:

$$|\mathbb{E}\left[\tilde{r}_t(z,\lambda)-r_t(z,\lambda)\right| I_N]| \quad \leq \quad \sum_{l=1}^N \left\{2\lambda\mathcal{U}_l K'\sigma^2+2K'\sigma^3\sqrt{3\mathcal{V}_l(\mathcal{V}_l+\Gamma_l)}+\tfrac{3}{4}\mathcal{V}_l\left(K'\sigma^2\right)^2\right\},$$

which concludes Theorem 5. □

G Proof for the expected error for the smoother threshold filter

For the filter $\mathcal{F}_\lambda$ defined for a given $\lambda>0$, we have the following expression for the error of reconstruction:

$$\begin{aligned} r_t^{\mathcal{L}}(z,\lambda) &= \mathbb{E}\left[\sum_{l=1}^N\left(\hat{z}_l-\mathcal{F}_\lambda(\hat{z}_l^{\mathcal{L}})\right)^2\middle| I_N\right]\\ &= \sum_{l=1}^N\left(\hat{z}_l^2-2\hat{z}_l\mathbb{E}\left[\mathcal{F}_\lambda(\hat{z}_l^{\mathcal{L}})\middle| I_N\right]+\mathbb{E}\left[\left(\mathcal{F}_\lambda(\hat{z}_l^{\mathcal{L}})\right)^2\middle| I_N\right]\right)\\ &= \sum_{l=1}^N\left(\sigma^2(\mathcal{V}_l+\Gamma_l)-2\hat{z}_l\mathbb{E}\left[\hat{z}_l^{\mathcal{L}}\ \mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\middle| I_N\right]+\mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}}\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\right)^2\middle| I_N\right]+2\mathbb{E}\left[(\hat{z}_l^{\mathcal{L}})^2\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\middle| I_N\right]\right), \end{aligned} \tag{27}$$

where

$$\mathcal{G}_\lambda : x\mapsto \frac{\mathcal{F}_\lambda(x)}{x}-1=-\exp\left(\frac{-x^2}{\lambda\sqrt{\mathcal{V}_l+\Gamma_l}}\right).$$

We then look for a simple expression for $\mathbb{E}\left[\hat{z}_l^{\mathcal{L}}\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\middle| I_N\right]$, $\mathbb{E}\left[(\hat{z}_l^{\mathcal{L}})^2\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\middle| I_N\right]$ and $\mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}}\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\right)^2\middle| I_N\right]$. We note that

$$\begin{cases} \mathcal{G}_\lambda(x)\delta(x)=-\exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega\right]\right)\exp\left(\frac{-(x-\hat{z}_l\omega)^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)\omega}\right)\frac{1}{\sigma\sqrt{2\pi(\mathcal{V}_l+\Gamma_l)}}\\ \mathcal{G}_\lambda(x)^2\delta(x)=\exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega'\right]\right)\exp\left(\frac{-(x-\hat{z}_l\omega')^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)\omega'}\right)\frac{1}{\sigma\sqrt{2\pi(\mathcal{V}_l+\Gamma_l)}}, \end{cases}$$

where

$$\begin{cases} \omega=\frac{\lambda}{\lambda+2\sigma^2\sqrt{\mathcal{V}_l+\Gamma_l}}\\ \omega'=\frac{\lambda}{\lambda+4\sigma^2\sqrt{\mathcal{V}_l+\Gamma_l}}. \end{cases}$$

Therefore

$$\begin{aligned} \mathbb{E}\left[\hat{z}_l^{\mathcal{L}}\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\middle| I_N\right] &= -\exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega\right]\right)\times\frac{1}{\sigma\sqrt{2\pi(\mathcal{V}_l+\Gamma_l)}}\int_{\mathbb{R}}x\exp\left(\frac{-(x-\hat{z}_l\omega)^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)\omega}\right)dx\\ &= -\exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega\right]\right)\hat{z}_l\omega^{3/2} \end{aligned} \tag{28}$$

and

$$\begin{aligned} \mathbb{E}\left[(\hat{z}_l^{\mathcal{L}})^2\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\middle| I_N\right] &= -\exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega\right]\right)\times\frac{1}{\sigma\sqrt{2\pi(\mathcal{V}_l+\Gamma_l)}}\int_{\mathbb{R}}x^2\exp\left(\frac{-(x-\hat{z}_l\omega)^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)\omega}\right)dx\\ &= -\exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega\right]\right)\left[(\hat{z}_l\omega)^2+\sigma^2(\mathcal{V}_l+\Gamma_l)\omega\right]\sqrt{\omega}. \end{aligned} \tag{29}$$

The third expectation we are looking for is equal to :

$$\begin{aligned} \mathbb{E}\left[\left(\hat{z}_l^{\mathcal{L}}\mathcal{G}_\lambda(\hat{z}_l^{\mathcal{L}})\right)^2\middle| I_N\right] &= \exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega'\right]\right)\times\frac{1}{\sigma\sqrt{2\pi(\mathcal{V}_l+\Gamma_l)}}\int_{\mathbb{R}}x^2\exp\left(\frac{-(x-\hat{z}_l\omega')^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)\omega'}\right)dx\\ &= \exp\left(\frac{-\hat{z}_l^2}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\left[1-\omega'\right]\right)\left[(\hat{z}_l\omega')^2+\sigma^2(\mathcal{V}_l+\Gamma_l)\omega'\right]\sqrt{\omega'}. \end{aligned} \tag{30}$$

Then, equations (27) to (30) lead to :

$$\begin{aligned} r_t^{\mathcal{L}}(z,\lambda)=\sum_{l=1}^N\quad &\left(\sigma^2(\mathcal{V}_l+\Gamma_l)+2\exp\left(\frac{-\hat{z}_l^2(1-\omega)}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)\omega^{3/2}\left[(1-\omega)\hat{z}_l^2-\sigma^2(\mathcal{V}_l+\Gamma_l)\right]\right.\\ &\left.+\exp\left(\frac{-\hat{z}_l^2\left(1-\omega'\right)}{2\sigma^2(\mathcal{V}_l+\Gamma_l)}\right)(\omega')^{3/2}\left[\hat{z}_l^2+\sigma^2(\mathcal{V}_l+\Gamma_l)\right]\right). \end{aligned}$$