Interpretation and improvement of an iterative wavelet–based denoising method

R. Ranta[1], C. Heinrich[2], Valérie Louis-Dorr[1], and D. Wolf[1]

[1] Centre de Recherche en Automatique de Nancy – CRAN UMR CNRS 7039,

2 av. de la Forêt de Haye, 54516 Vandoeuvre, France

[2] Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection – LSIIT UMR CNRS-ULP 7005,

Pôle API Parc d'innovation, bd. Sébastien Brant, 67400 Illkirch-Graffenstaden, France

***Abstract*— The goal of this communication is to shed new light on a wavelet–based denoising method developed by Hadjileontiadis *et al.* [1], [2], which is derived from an iterative denoising algorithm by Coifman and Wickerhauser [3], [4]. The underlying algorithm is revisited and interpreted as a fixed point algorithm. This allows to derive a new version of the algorithm largely increasing computational efficiency.**

I. Introduction

The general framework of this communication is wavelet–based denoising of discrete–time signals. We deal with a method proposed by Coifman and Wickerhauser [3], [4], which has been further developed and applied by Hadjileontiadis *et al.* [1], [2] to lung and bowel–sound denoising, segmentation and analysis (see also [5])[1]. The denoised signal is estimated with the help of an iterative scheme, yielding successive refinements of this signal. This may be seen as "peeling off successive layers" of the signal, in Coifman and Wickerhauser's own terms: at a given iteration, the noise residual is decomposed on a wavelet basis and the largest coefficients of each scale are used to reconstruct a signal which is added to the current estimate of the denoised signal. This denoising procedure may be interwoven with a best basis search at each iteration [4].

The goal of this paper is to show that, under certain conditions, these successive refinements may be interpreted as fixed–point algorithm searches for determining independent thresholds for each scale. If no best basis procedure is considered, we show moreover that successive wavelet–decompositions, –reconstructions (WDRs) are useless, thus allowing an important reduction of the computational burden.

This communication is organized as follows. In the second section, we detail the method originally proposed by Coifman and Wickerhauser, and used by Hadjileontiadis *et al.* in a biomedical framework. In the third section, we detail the proposed fixed–point interpretation. Conclusion is given, with particular focus on the CPU time stakes.

[1] See for example [6] and the references therein for a broader point of view on wavelet–based denoising methods.

II. ITERATIVE WAVELET–BASED DENOISING METHODS

We consider the model $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{z}$ is the given discrete–time signal to be denoised, $\boldsymbol{x}$ is the denoised unknown version of $\boldsymbol{z}$ and $\boldsymbol{n}$ the noise. Coifman and Wickerhauser proposed an iterative denoising scheme as $\boldsymbol{z} = \boldsymbol{x}_k + \boldsymbol{n}_k$, where k is the iteration step. The current noise estimation $\boldsymbol{n}_k$, which is initialized for $k = 0$ as $\boldsymbol{n}_0 = \boldsymbol{z}$, is decomposed on an orthogonal wavelet basis as:

$$\boldsymbol{n}_k = \sum_{p,j} w_{n,k}^{j,p} \; \boldsymbol{\psi}^{j,p} + \sum_{p} w_{n,k}^{M,p} \; \boldsymbol{\phi}^{M,p}.$$

We use the following notations: j is the scale, p the position (translation index), $\boldsymbol{\psi}$ the wavelet, $\boldsymbol{\phi}$ the scaling function and M the decomposition depth [7]. As we consider finite duration signals and compactly supported wavelets, the dimension of the transformed vector is finite. Let $\boldsymbol{\Omega}_{n,k}^{j}$ be the vector containing the noise coefficients at scale j, $w_{n,k}^{j,p}$ and $\boldsymbol{\Omega}_{n,k}$ the complete noise coefficients vector. By thresholding $\boldsymbol{\Omega}_{n,k}$, one obtains the current "peeled off layer" $\boldsymbol{\Omega}_{\Delta x,k+1}$. The noise coefficient vector $\boldsymbol{\Omega}_{n,k+1}$, derived from $\boldsymbol{\Omega}_{\Delta x,k+1} + \boldsymbol{\Omega}_{n,k+1} = \boldsymbol{\Omega}_{n,k}$, is used to reconstruct $\boldsymbol{n}_{k+1}$. The iterations end when the stop criterion:

$$STC_{k+1} = \|\boldsymbol{n}_k\|^2 - \|\boldsymbol{n}_{k+1}\|^2 < \varepsilon, \tag{1}$$

is validated, for a user chosen ε.

The threshold selection rule in the algorithm of Coifman and Wickerhauser [3], [4] is different from the one in the algorithm of Hadjileontiadis *et al.* [1], [2]. The latter, which we develop in this communication, writes:

1) compute σ_k^j as $\left(\sigma_k^j\right)^2 = \frac{1}{N} \left\|\boldsymbol{\Omega}_{n,k}^j\right\|^2$;
2) compute the threshold T_{k+1}^j as $T_{k+1}^j = f(\sigma_k^j)$. The particular case considered here is: $T_{k+1}^j = F_a^j \; \sigma_k^j$, where F_a^j is as user–defined constant[2], which can depend on the scale j;
3) compute $\boldsymbol{\Omega}_{n,k+1}^j$ and $\boldsymbol{\Omega}_{\Delta x,k+1}^j$ by hard–thresholding[3] $\boldsymbol{\Omega}_{n,k}^j$ using the threshold T_{k+1}^j;
4) compute $\boldsymbol{n}_{k+1}$ and $\Delta\boldsymbol{x}_{k+1}$ as wavelet reconstructions of $\boldsymbol{\Omega}_{n,k+1}$ and $\boldsymbol{\Omega}_{\Delta x,k+1}$ respectively, and set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \Delta\boldsymbol{x}_{k+1}$;
5) loop to the top if the stop criterion (1) is not reached.

The general idea of the algorithm is to consider, at each iteration step, the empirical distribution of the wavelet coefficients of the current noise estimation for each scale. The large coefficients, whose values exceed the current threshold defined as F_a^j times the current empirical standard deviation, are supposed to belong to the denoised signal and are excluded from the noise estimate. This procedure thus iteratively calculates final different thresholds for each scale. This enables an adaptation of the threshold to the coefficient distribution of a given scale, hence to a colored noise.

III. A FIXED–POINT INTERPRETATION OF THE ALGORITHM

The goal of this section is to show that the aforementioned iterative determination of the final thresholds may be interpreted as finding fixed points of functions whose expressions will be given. Moreover, we show that successive WDRs are useless, thus saving much precious computational time, providing that no best basis search is considered.

[2]Hadjileontiadis *et al.* considered an unique $F_a = 3$, on a medical expertise of denoised signals basis.

[3]See [8] for the distinction between hard– and soft–thresholding

A. Preliminaries

Let us notice that the orthogonality of the wavelet transform allows us to rewrite the stop criterion (1) as:

$$STC_{k+1} = \|\boldsymbol{n}_k\|^2 - \|\boldsymbol{n}_{k+1}\|^2 = \|\boldsymbol{\Omega}_{n,k}\|^2 - \|\boldsymbol{\Omega}_{n,k+1}\|^2 . \quad (2)$$

Moreover, for $\varepsilon = 0$, iterations are stopped when $\boldsymbol{\Omega}_{n,k} = \boldsymbol{\Omega}_{n,k+1}$, i.e., when no wavelet coefficient is thresholded and so $\boldsymbol{\Omega}_{\Delta x,k+1} = \mathbf{0}$.

Furthermore, because of the linearity property of the wavelet decomposition, we have:

$$\boldsymbol{\Omega}_z = \boldsymbol{\Omega}_{n,k} + \sum_{i=0}^{k} \boldsymbol{\Omega}_{\Delta x,i} = \boldsymbol{\Omega}_{n,k} + \boldsymbol{\Omega}_{x,k}, \ \forall\ k. \quad (3)$$

B. Consequences

The consequences are straightforward:

- because of the orthogonality of the wavelet transform we can use the modified stop criterion (2) in the 5^{th} step of the previously described algorithm. This makes the reconstruction of the estimated noise $\boldsymbol{n}_k = \mathrm{WR}(\boldsymbol{\Omega}_{n,k})$ useless;
- because of the linearity of the wavelet transform, we don't have to compute any $\Delta\boldsymbol{x}_k = \mathrm{WR}(\boldsymbol{\Omega}_{\Delta x,k})$ (the goal of this computation is to yield $\boldsymbol{x}$ as $\boldsymbol{x} = \sum_{k=0}^{K} \Delta x_k = \sum_{k=0}^{K} \mathrm{WR}(\boldsymbol{\Omega}_{\Delta x,k})$. Instead, the final denoised–signal estimate is computed as $\boldsymbol{x} = \mathrm{WR}(\boldsymbol{\Omega}_x) = \mathrm{WR}(\sum_{k=0}^{K} \boldsymbol{\Omega}_{\Delta x,k})$, where K is the number of iterations till convergence.

The 4^{th} step of Hadjileontiadis *et al.*'s algorithm can then be completely eliminated.

C. Thresholding

The thresholding will then be implemented as follows: for all j, $T_k^j = F_a^j \sigma_{k-1}^j$ thresholds the coefficients of $\boldsymbol{\Omega}_{n,k-1}^j$ in order to obtain $\boldsymbol{\Omega}_{n,k}^j$ and $\boldsymbol{\Omega}_{\Delta x,k}^j$. If we have at least one index q such as $\boldsymbol{\Omega}_{\Delta x,k}^j(q) \neq 0$, we can write $|\boldsymbol{\Omega}_{\Delta x,k}^j(q)| \geq T_k^j$. Moreover, $T_k^j > |\boldsymbol{\Omega}_{n,k}^j(q)|, \forall\ q$. We also have

$$|\boldsymbol{\Omega}_{\Delta x,k}^j(q)| \geq T_k^j > |\boldsymbol{\Omega}_{\Delta x,k+1}^j(q)| \geq T_{k+1}^j > \ldots , \quad (4)$$

since $\boldsymbol{\Omega}_{\Delta x,k+1}^j$ is obtained by thresholding $\boldsymbol{\Omega}_{n,k}^j$. As a consequence, every non vanishing coefficient of $\boldsymbol{\Omega}_{x,k}^j = \sum_{i=0}^{k} \boldsymbol{\Omega}_{\Delta x,i}^j$ is greater or equal, in modulus, than T_k^j, which is greater than the modulus of any coefficient of $\boldsymbol{\Omega}_{n,k}^j$. The thresholds sequence is initialized as $T_0^j = \infty$, yielding $\boldsymbol{n}_0 = \boldsymbol{z}$, which leads to a decreasing sequence T_k^j (see also eq. (4)), with $T_0^j > T_1^j > \ldots > T_{K-1}^j = T_K^j$, where T_K^j is the final threshold (reached when $\boldsymbol{\Omega}_{\Delta x,K}^j(q) = 0$ for all q).

Thus, T_k^j splits the coefficients $\boldsymbol{\Omega}_z^j = \boldsymbol{\Omega}_{n,k}^j + \boldsymbol{\Omega}_{x,k}^j$ in two disjoint vectors, which means that thresholding $\boldsymbol{\Omega}_{n,k}^j$ by T_{k+1}^j amounts to thresholding directly $\boldsymbol{\Omega}_z^j$. Hence, successive thresholdings of $\boldsymbol{\Omega}_z^j$ are useless and the final thresholds T_K^j can be computed in a previous step of the algorithm[4].

D. Fixed point algorithm

Provided that we choose $\varepsilon = 0$ (see eq. (1)), the algorithm writes as: for each scale j and T_k^j being given:

1) compute σ_k^j as $\left(\sigma_k^j\right)^2 = \dfrac{1}{N} \sum_p \left(w_z^{j,p}\ \mathbb{I}_{(|w_z^{j,p}|<T_k^j)}\right)^2$, where

$$\mathbb{I}_{(|x|<T)} = \begin{cases} 1, & \text{if } |x| < T, \\ 0, & \text{if } |x| \geq T; \end{cases}$$

2) compute T_{k+1}^j as $T_{k+1}^j = F_a^j\ \sigma_k^j$,

and loop till convergence. This boils down to the following iterative scheme:

$$T_{k+1}^j = F_a^j \sqrt{\frac{1}{N} \sum_p \left(w_z^{j,p}\ \mathbb{I}_{(|w_z^{j,p}|<T_k^j)}\right)^2}.$$

[4]In this paper, the same basis is considered during successive iterations. This is the case, as in [1], [2], if no best basis selection procedure is included in every iteration.

Considering function f defined as

$$f(x) = F_a^j \sqrt{\frac{1}{N}\sum_p \left(w_z^{j,p}\ \mathbb{I}_{(|w_z^{j,p}|<x)}\right)^2}, \qquad (5)$$

the final threshold is computed by the fixed-point descent algorithm $T_{k+1}^j = f(T_k^j)$ (see figure 1(b)).

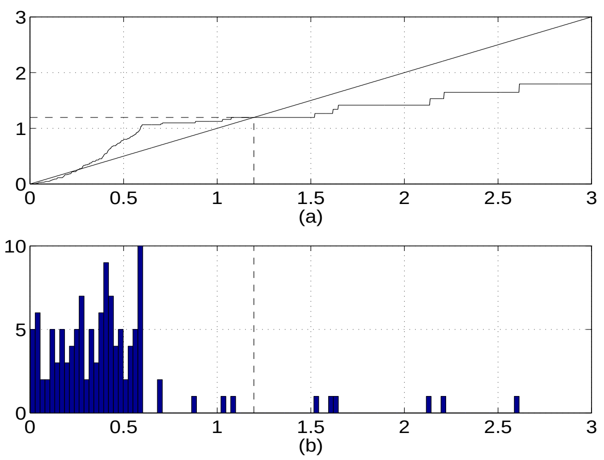

Fig. 1. For a given empirical distribution of $w_z^{j,p}$ and a given j: (a) functions $y = f(x)$ (for $F_a^j = 3$) and $y = x$, showing the fixed point interpretation and the final value $x_0 = T_K^j$ of the algorithm; (b) histogram of the absolute values of $w_z^{j,p}$. The dashed lines represent the threshold, which separates the noise (small values) from the informative signal.

We have to verify that the function defined in (5) has at least one fixed point. Indeed, it is straightforward that f (see figure 1(a)) is a piecewise constant monotonic (increasing) positive function, defined for all $x \in \mathbb{R}^+$ and taking values in a finite set of real numbers, since the dimension of $\mathbf{\Omega}_z^j$ is finite (for compactly supported wavelets). The values of the function f depend on the multiplicative user constant F_a^j, so certains conditions must be fulfilled by this constant in order to assure the existence of a fixed point:

- given a so that there exists $w_z^{j,p}$ such that $0 < |w_z^{j,p}| < a$, one can find a finite constant F_a^j which assures $f(a) > a$;
- for the given F_a^j, one can find $b > a$ large enough to have $f(b) < b$.

As a consequence of these observations, there exists at least a fixed-point x_f for f: $f(x_f) = x_f, x_f \in \mathbb{R}^+$. One can prove that the iteration $T_{k+1}^j = f(T_k^j)$ yields successive decreasing values of the thresholds (see figure 1(a)) and converge towards the first encountered fixed point, i.e., the greatest (see Appendix). This fixed-point satisfies the stop criterion (2) for $\varepsilon = 0$, so it is the final researched threshold T_K^j.

To conclude this discussion, let us mention that the number of the fixed-points of the function f may vary, depending on the value of the constant F_a^j and the shape of the histogram of the $w_z^{j,p}$s. Thus, the fixed point interpretation enlightens the role of the constant in the iterative thresholding: the user must choose an F_a^j big enough to ensure the convergence of the algorithm to a strictly positive value. On the other hand, the thresholds values increase with the value of F_a^j, so an excessive value of F_a^j will lead to an early stop of the algorithm (i.e., no or little signal is extracted from noise). In fact, in practical implementations, the value of F_a^j must be bounded: if $F_a^j > \dfrac{max(|\mathbf{\Omega}_z^j|)}{\sigma_0^j}$, one can easily see that the convergence point is reached after the first iteration and no thresholding is performed.

We also have to note that, as the function f may have several fixed points, the final value of the threshold depends on the initializing value of the T_0^j sequence: an initialization $T_0^j \neq \infty$ can lead to an increasing sequence of thresholds and/or a convergence point that is not necessarily the largest fixed-point of the function f. In fact, the user must respect the following restrictions in the choice of the parameters:

- the algorithm must be initialized with $\boldsymbol{n}_0 = \boldsymbol{z}$ ($T_0^j = \infty$);

- the multiplicative constant F_a^j must be big enough to assure the convergence to a non-vanishing point, but not superior to $\frac{max(|\mathbf{\Omega}_z^j|)}{\sigma_0^j}$, in order to separate informative signal from noise.

IV. Conclusion

In this communication, an iterative wavelet–based denoising method by Hadjileontiadis *et al.* [1], [2] is interpreted as a fixed–point algorithm and intermediate WDR steps are suppressed, which is possible providing that no best basis selection is done.

Besides shedding light on the initial algorithm, this enables to save much computational time. The amount of saved time depends of course on the number of iterations of the initial algorithm: the larger the number of iterations, the larger the gain of CPU time. Implementation on bowel sound denoising and segmentation has shown a factor 4 reduction of CPU time [5].

V. Appendix

The goal of this appendix is to show that the fixed point iterations don't miss the largest fixed point. In the following, we drop the scale index j for the sake of simplicity. Let us note $\mathbf{\Gamma}$ the finite set of values taken by the function f (5), and $\{T_k\}$ the set of the iteratively computed thresholds ($k \neq 0$). Obviously, as the sequence of thresholds is computed as $f(x)$, $T_k \in \mathbf{\Gamma}, \forall\ k \neq 0$.

Let us suppose that the largest fixed point of the function f is missed by the iterative algorithm. That means that there exists $\gamma = f(\gamma)$ such as $\gamma \in \mathbf{\Gamma}$ and $\gamma \neq T_k, \forall\ k$. This can be further rewritten as: there exists k such as $T_{k+1} < \gamma < T_k$. But $T_{k+1} = f(T_k)$, so the first inequality writes $f(T_k) < \gamma$. On the other hand, f is monotonic increasing, so the second inequality implies $f(\gamma) \leq f(T_k)$, which implies (because γ is a fixed-point) $\gamma \leq f(T_k)$, which contradicts the first inequality. So the hypothesis made at the beginning of the paragraph is false: the largest fixed point is not missed by the iterative algorithm.

References

[1] L. Hadjileontiadis and S. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 12, pp. 1269–1281, 1997.

[2] L. Hadjileontiadis, L. Liatsos, C. Mavrogiannis, T. Rokkas, and S. Panas, "Enhancement of bowel sounds by wavelet-based filtering," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 876–886, 2000.

[3] R. Coifman and M. Wickerhauser, "Adapted waveform de-noising for medical signals and images," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 5, pp. 578–586, 1995.

[4] R. Coifman and M. Wickerhauser, *Experiments with adapted wavelet de–noising for medical signals and images*, ch. 12, pp. 323–346. Time frequency and wavelets in biomedical signal processing, IEEE Press Series in Biomedical Engineering, M. Akay ed., 1998.

[5] R. Ranta, C. Heinrich, V. Louis-Dorr, D. Wolf, and F. Guillemin, "Wavelet–based bowel sounds denoising, segmentation and characterization," in 23^{rd} *Annual International Conference of the IEEE EMBC*, Istanbul, Turkey, 25–28 October 2001.

[6] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 909–919, 1999.

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

[8] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.