Wavelet de-noising for blind source separation in noisy mixtures.

Bertrand Rivet[1], Vincent Vigneron[1], Anisoara Paraschiv-Ionescu[2] and Christian Jutten[1]

[1] Institut National Polytechnique de Grenoble.
Laboratoire des Images et des Signaux
Grenoble, France
[2] Swiss Federal Institute of Technology
Lausanne, Switzerland

Abstract. Blind source separation, which supposes that the sources are independent, is a well known domain in signal processing. However, in a noisy environment the estimation of the criterion is harder due to the noise. In strong noisy mixtures, we propose two new principles based on the combination of wavelet de-noising processing and blind source separation. We compare them in the cases of white/correlated Gaussian noise.

1 Introduction

Blind source separation (BSS) is a well known domain in signal processing. Introduced by J. Hérault, C. Jutten and B. Ans [1], its goal is to recover unknown source signals of which only mixtures are observed with only assumptions that the source signals are mutually statistically independent. A lot of BSS models such as instantaneous linear mixtures, convolutive mixtures are presented in recent publications [2–4]. The success of the BSS is its wide range of applications whether it is in telecommunication, speech or medical signal processing. However, the best performances of these methods are obtained for the ideal BSS model and their effectiveness is definitely decreased with observations corrupted by additive noise.

The aim of this paper is to present how to associate wavelet de-noising processing and BSS in order to improve the estimated sources. This paper is organized as follows. Section 2 introduces the BSS problem in noisy mixtures. Section 3 explains the wavelet de-noising principles and proposes two new principles for associating wavelet de-noising and BSS. Section 4 proposes numerical experiments before conclusion and perspectives in section 5.

2 Modelization of the problem

In an instantaneous linear problem of source separation, the unknown source signals and the observed data are related by (Fig. 1):

$$\mathbf{y}(k) = \mathcal{A}\,\mathbf{s}(k) + \mathbf{n}(k) = \mathbf{x}(k) + \mathbf{n}(k) \tag{1}$$

where $\mathcal{A}$ is an unknown full rank $p \times q$ mixing matrix ($p \geq q$), $\mathbf{s}(k)$ is a column vector of q source signals assumed mutually statistically independent, $\mathbf{y}(k)$ a column vector of p mixtures and $\mathbf{n}(k)$ an additive noise. By estimating a $q \times p$ full rank matrix $\mathcal{B}$ one provides estimated sources which are the components (as independent as possible) of the output signal vector $\hat{\mathbf{s}}(k)$ defined as (Fig. 1):

$$\hat{\mathbf{s}}(k) = \mathcal{B}\,\mathbf{y}(k) = \mathcal{B}\,\mathcal{A}\,\mathbf{s}(k) + \mathcal{B}\,\mathbf{n}(k) \tag{2}$$

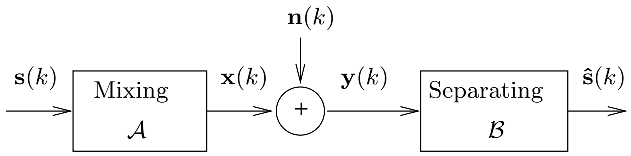

Fig. 1. Blind source separation model in noisy mixtures.

This equation shows that the estimated sources $\hat{\mathbf{s}}$ are affected by the additive noise. Let us illustrate this phenomenon with the figure 2. c) which shows that the estimated separating matrix $\hat{\mathcal{B}}$ is not well estimated ($\hat{\mathcal{B}} \neq \mathcal{B}$) since the ideal sources (see definition 1 below) are different from the original sources. Moreover, even if the separating matrix is well estimated ($\hat{\mathcal{B}} = \mathcal{B}$), the noise affects the estimated sources as shown in d).

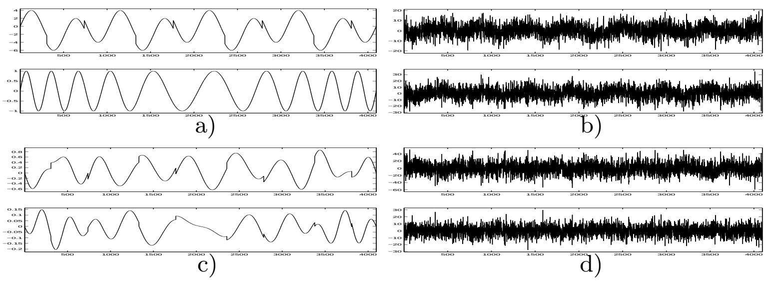

Fig. 2. Illustration of the harmful presence of the noise. a) the original sources $\mathbf{s}$, b) the noisy mixtures $\mathbf{y}$, c) the "ideal" sources (see definition 1) and d) the noisy estimated sources $\hat{\mathbf{s}}$.

Definition 1. *The ideal source signal* $\mathbf{s}_{ideal}(k)$ *is defined as the product of the separating matrix* $\mathcal{B}$ *which is estimated from the noisy mixtures, by the noisy-free mixtures* $\mathbf{x}(k)$*:*

$$\mathbf{s}_{\text{ideal}}(k) = \mathcal{B}_{\text{noisy}}\,\mathbf{x}(k) \tag{3}$$

3 Wavelet de-noising for BSS

As we said, in BSS the estimated separating matrix is affected by the additive noise. Thus, a powerful de-noising processing before separation seems to be a good solution. In this section we first recall the bases of wavelet de-noising (3.1). Then we propose three methods of wavelet de-noising for BSS: the method proposed by Paraschiv-Ionescu *et al.* [5] (3.2) and two new methods (3.3 and 3.4).

3.1 Fundamental of wavelet de-noising

The discret wavelet transform (DWT) is a batch processing, which analyses a finite length time domain signal by breaking up the initial domain in two parts: the detail and approximation information [6]. The approximation domain is successively decomposed into detail and approximation domains.

We use two properties of the discret wavelet transform (DWT):

- the DWT is scattered[3]: a few number of large coefficients dominates the representation,
- the wavelet coefficients are less correlated than the temporal ones.

As a result, we use a nonlinear thresholding function and we treat the coefficients independently to each other. Practicaly, the wavelet de-noising processing consists in applying the DWT to the original noisy signal, chosing the value of the threshold, thresholding the detail coefficients, then inversing the DWT.

Denote $\mathcal{W}(\cdot)$ and $\mathcal{W}^{-1}(\cdot)$ the forward and reverse DWT operators, $d(\cdot)$ the operator which selects the value of the threshold and $\mathcal{T}(\cdot,\lambda)$ the thresholding operator with the threshold λ. Considering the $i-$th noisy observed signal $\mathbf{y}_i$ from (1), the wavelet de-noising processsing is defined as

$$\begin{cases} \mathbf{w}_i = \mathcal{W}(\mathbf{y}_i) = \theta_i + \mathbf{b}_i \\ \lambda = d(\mathbf{w}_i) \\ \hat{\theta}_i = \mathcal{T}(\mathbf{w}_i, \lambda) \\ \hat{\mathbf{x}}_i = \mathcal{W}^{-1}(\hat{\theta}_i) \end{cases} \tag{4}$$

where $\mathbf{x}_i = (\mathcal{A}\,\mathbf{s})_i$ is the $i-$th noisy free mixture. $\theta_i = \mathcal{W}(\mathbf{x}_i)$ and $\mathbf{b}_i = \mathcal{W}(\mathbf{n}_i)$ are respectively the DWT coefficients of the noisy free mixture and the noise. Let denote by $\hat{\mathbf{x}}_i = \mathcal{D}(\mathbf{w}_i)$ the de-noising processing summarizing the four previous stages. Note that the choice of the wavelet function used in the transform is based on one's needs. The choice of a DWT operator can have significant effects on the scheme's performance in terms of noise/signal ratio. Generally, the best choice will resemble "theoretically" the desired feature in the profile, the counterpart being that the analysis fails in direct comparison of the profiles' wavelet transform and identify not-so-similar features. In our analysis, we apply a wavelet transform with the data-adaptive threshold selection rule Sureshrink of ©MATLAB wavelet toolbox to identify sharp gradients.

3.2 P.S. method

The wavelet de-noising Pre-Separating processing (P.S.) [5] consists in introducing a wavelet de-noising processing before the separating algorithm (Fig. 3). Thus, the separating matrix $\mathcal{B}$ is estimated from de-noised mixtures $\hat{\mathbf{x}}(k)$. The estimated sources and the noisy mixtures are related by

$$\hat{\mathbf{s}} = \mathcal{B}_{\mathcal{D}(\mathbf{y})}\,\hat{\mathbf{x}} \qquad \text{with} \quad \hat{\mathbf{x}} = \mathcal{D}(\mathbf{y}) \tag{5}$$

[3] This property is based on the fact that the noise is broad band and is present over all coefficients while deterministic signal is narrow band.

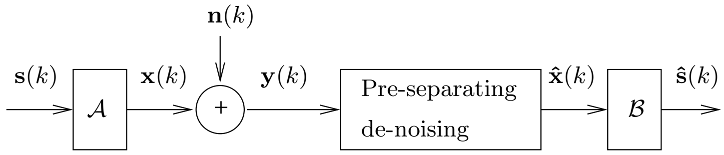

Fig. 3. Principle of the P.S. method.

where the index $\mathcal{D}(\mathbf{y})$ recalls that the separating matrix $\mathcal{B}_{\mathcal{D}(\mathbf{y})}$ is estimated from the de-noised mixtures.

3.3 Serial P.S.P. method

However, the P.S. method is definitely not efficient. Indeed, the frequency bands or the scales occupied by the mixtures correspond to the union of those occupied by the sources since the mixtures are linear combinations of the sources.

Consequently, we propose the following Serial wavelet Pre-Separating and Post-separating de-noising processing (Serial P.S.P). This method (Fig. 4) allows

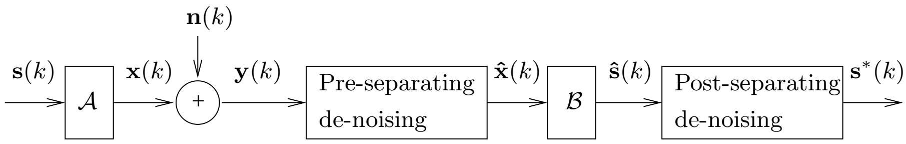

Fig. 4. Principle of the Serial P.S.P. method.

us to adapt the pre-separating de-noising processing to the mixtures and the post-separating de-noising processing to the sources. However, a classical de-noising processing using the variance of the noise $\hat{\sigma}^2$ estimated from the wavelet coefficients at scale 1 cannot succeed. Indeed, the de-noising pre-processing changes the white nature of the noise. In order to overcome this difficulty, we propose the following stages:

1. estimate the variance of the noise $\hat{\sigma}^2_{\mathbf{y}} = (\sigma^2_{y_1}, \cdots, \sigma^2_{y_q})^T$ which corrupts the observed data (*cf* [6] page 447),
2. calculate $\hat{\sigma}^2_{\hat{\mathbf{s}}} = \mathcal{B}^{*2}\, \hat{\sigma}^2_{\mathbf{y}}$ which is an estimation of the variance of the noise[4] present in the estimated sources $\hat{\mathbf{s}}(k)$, since the noise is white and Gaussian,
3. use a de-noising processing on $\hat{\mathbf{s}}(k)$ using $\hat{\sigma}^2_{\mathbf{y}}$ for determining the value of the threshold.

Using Serial P.S.P., we have to choose carefully the pre-denoising scale. If this scale is overestimated, it provides a distortion of the mixtures $\mathbf{x} = \mathcal{A}\mathbf{s}$, which becomes $\tilde{\mathbf{x}} = \tilde{\mathcal{A}}\tilde{\mathbf{s}}$, where both $\tilde{\mathcal{A}} \neq \mathcal{A}$ and $\tilde{\mathbf{s}} \neq \mathbf{s}$. Thus both estimation of $\tilde{\mathcal{A}}$, and restitution of $\tilde{\mathbf{s}}$, even perfect, do not lead to the good solutions.

[4] Let denote $\mathcal{B}^{*2}$ the operator which means $(\mathcal{B}^{*2})_{i,j} = (\mathcal{B}_{i,j})^2$.

3.4 Parallel P.S.P. method

One of the major problems of the previous methods (P.S. or Serial P.S.P.) lies in the pre-separating de-noising processing: it could remove signal and especially the details (*i.e.* differences between the used wavelet and the signal). Even if this can provide a good estimate of the separating matrix, this may be disastrous for estimating the source signals since the details can contain low power sources (ECG fetal sources for instance). To overcome this problem we propose the following principle (Fig. 5): the Parallel Pre-Separating de-noising and Post-separating de-noising processing (Parallel P.S.P.).

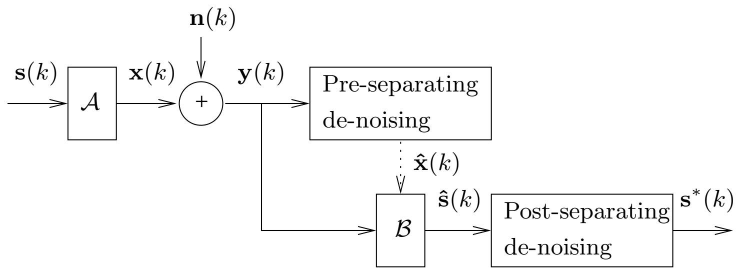

Fig. 5. Principle of Parallel P.S.P. denoising

The algorithm consists in

1. de-noising the noisy observed data $\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{b}(k)$ using an *ad-hoc* principle to obtain estimated mixed signals $\hat{\mathbf{x}}(k) = \mathcal{D}(\mathbf{y})$,
2. using these estimated mixed signals $\hat{\mathbf{x}}(k)$ in order to estimate the separating matrix $\mathcal{B}_{\mathcal{D}(\mathbf{y})}$,
3. estimating noisy source signals defined as $\hat{\mathbf{s}}(k) = \mathcal{B}_{\mathcal{D}(\mathbf{y})}\,\mathbf{y}(k)$,
4. de-noising the noisy estimated source signal thanks to a post-separating de-noising processing $\mathbf{s}^*(k) = \mathcal{D}(\hat{\mathbf{s}})$.

Thus, noisy estimated source signals and observed data are related by

$$\hat{\mathbf{s}}(k) = \mathcal{B}_{\mathcal{D}(\mathbf{y})}\,\mathbf{y}(k) = \mathcal{B}_{\mathcal{D}(\mathbf{y})}\,\mathcal{A}\,\mathbf{s}(k) + \mathcal{B}_{\mathcal{D}(\mathbf{y})}\,\mathbf{b}(k) \tag{6}$$

This principle allows us to distinguish the estimation of the separating matrix $\mathcal{B}$ and the restitution of the denoised sources $\mathbf{s}^*$.

4 Simulated experiments

In the following, we will consider the case of two sources mixed by a 2×2 matrix. We suppose that the mixed signals are corrupted by an additive noise. In order to compare the principles, we need two indexes: the performance index which quantifies the separating accuracy and the decay index which quantifies the remaining signal after de-noising processing.

Definition 2. *The performance index (PI) [7] which quantifies the separation accuracy is defined as*

$$PI = \sum_{i=1}^{q} \left\{ \left(\sum_{j=1}^{q} \frac{|c_{i,j}|^2}{\max_l |c_{i,l}|^2} - 1 \right) + \left(\sum_{j=1}^{q} \frac{|c_{j,i}|^2}{\max_l |c_{l,i}|^2} - 1 \right) \right\} \tag{7}$$

where $c_{i,j}$ *is the* $(i,j)-th$ *element of the global system* $\mathcal{C} = \mathcal{B}\,\mathcal{A}$.

Definition 3. *The remaining signal* $\mathbf{s}_{remaining}$ *after de-noising processing is defined as the inverse wavelet transform of the coefficients of the noisy-free signal* $\mathbf{s}$ *from index where the noisy coefficients* $\mathcal{O}(\mathbf{x})$ *are larger than the value of the threshold.*

Definition 4. *The decay index (DI) which quantifies the removed signal by the de-noising processing is defined as*

$$DI = \frac{\mathcal{P}_{original}}{\mathcal{P}_{remaining}} \tag{8}$$

where $\mathcal{P}_{original}$ *is the power of the noisy-free signal and* $\mathcal{P}_{remaining}$ *the power of the remaining signal after de-noising processing.*

We compare the different principles for two separating algorithms (JADE and EASI) and for different signal to noise ratios (SNR) for the observed mixtures. Each simulation run is repeated 50 times, holding all factors constant except the noise samples. We use the hard shrinkage of the stationary wavelet transform [8] as de-noising processing.

4.1 Case of a white Gaussian noise

Let us begin with a white Gaussian additive noise:

$$\mathbf{y}(k) = \mathcal{A}\;\mathbf{s}(k) + \mathbf{n}(k) \; with \; \mathbf{n}(k) \overset{iid}{\sim} \mathcal{N}(0, \Gamma_{\mathbf{n}}) \; and \; \Gamma_{\mathbf{n}} \; diagonal. \tag{9}$$

The table 1 regroups the DI (dB), the SNR (dB) of the denoised mixtures and the PI (dB) versus the scale used for the wavelet de-noising for a SNR of the mixtures egals to -5dB. We note that even if the SNR of the de-noised mixtures

Scale	0	1	2	3	4	5	6	7	8	9	10
DI_1	0	0,001	0,003	0,005	0,011	0,027	0,360	1,182	1,515	1,788	1,795
DI_2	0	0	0,001	0,002	0,004	0,023	0,763	1,541	2,042	2,225	2,244
SNR_1	-5,0	-2,0	1,0	4,0	6,7	9,5	10,2	10,1	10,0	10,0	10,1
SNR_2	-5,0	-2,0	1,0	3,9	6,8	9,4	10,3	10,1	9,9	9,9	10,0
PI_{JADE}	-5	-16	-17	-19	-20	-21	-23	-22	-21	-20	-21

Table 1. Performance for various scales.

is better on scale 6 than on scale 5, the decay index underlines that the removed signal by the de-noising processing is definitively larger at scale 6 than at scale 5.

The table 2 summarizes the results in order to compare the different methods. The numbers between brackets represent the scales used for the de-noising processing pre-separation and post-separation. The performance index PI (dB) and the SNR (dB) for the two estimated sources are reported versus the SNR (dB) of the observed mixtures.

	SNR	-10	-5	0	5		-10	-5	0	5
without denoising	PI_{JADE}	0	-5	-22	-24	SNR_1	-9,2	-5,0	-0,3	4,7
	PI_{EASI}	1	-2	-15	-24	SNR_2	-9,1	-4,5	0,3	5,3
P.S. (6,0)	PI_{JADE}	-18	-23	-27	-29	SNR_1	5,9	9,9	13,8	17,3
	PI_{EASI}	-14	-19	-22	-24	SNR_2	5,6	10,1	14,8	19,0
Serial P.S.P. (5,6)	PI_{JADE}	-17	-21	-23	-24	SNR_1	7,4	11,9	15,6	17,8
	PI_{EASI}	-13	-18	-21	-23	SNR_2	7,0	11,6	16,6	21,4
Parallel P.S.P. (6,6)	PI_{JADE}	-18	-23	-27	-29	SNR_1	8,1	12,5	16,1	18,1
	PI_{EASI}	-14	-19	-22	-24	SNR_2	7,8	11.3	16,4	21,4

Table 2. Performance for a white Gaussian noise.

The denoised principles provide more accurate estimating of the separating matrix $\hat{\mathcal{B}}$ and improve the quality of the estimated sources. Serial P.S.P. and Parallel P.S.P. methods have similar performance which is better than the performance obtained with P.S. method.

4.2 Case of a colored Gaussian noise

Now, let study the methods with an additive colored Gaussian noise. The simulations were performed with short time dependence noise, modeled by a 2nd-order auto-regressive process AR(2):

$$\mathbf{n}(k) = 1.33\,\mathbf{n}(k-1) - 0.88\,\mathbf{n}(k-2) + \mathbf{w}(k) \tag{10}$$

with $\mathbf{w}(k)$ an iid Gaussian noise. Since the noise $\mathbf{n}$ has scale-dependent wavelet coefficients, we used a scale-dependent threshold.

In this case we only report the PI (dB) and the SNR of the estimated sources versus the SNR (dB) of the mixtures. The table 3 illustrates the interest of a post-separation de-noising. As in the case of the white noise case, the principles Serial P.S.P. and Parallel P.S.P. improve the performances.

5 Conclusion

The noise strongly limits the separation performance, encouraging us to use wavelet de-noising processing. In this paper, we propose two new principles, Serial P.S.P. and Parallel P.S.P., which associate wavelet de-noising and blind source separation. In noisy mixtures, these new principles improve the separation performance and give comparable results, but Parallel P.S.P. method is

	SNR	-10	-5	0	5		-10	-5	0	5
without de-noising	PI_{JADE}	0	-1	-15	-23	SNR_1	-9,3	-4,7	-0,3	4,7
	PI_{EASI}	0	-2	-12	-21	SNR_2	-9	-4,4	0,3	5,3
P.S. (6,0)	PI_{JADE}	-10	-22	-24	-24	SNR_1	7,2	11,6	16,0	19,0
	PI_{EASI}	-14	-19	-23	-24	SNR_2	7,9	12,8	17,4	21,5
Serial P.S.P. (5,6)	PI_{JADE}	-21	-23	-24	-24	SNR_1	11,2	14,9	17,6	19,6
	PI_{EASI}	-18	-20	-23	-24	SNR_2	10,6	14,8	20,0	25,0
Parallel P.S.P. (6,6 or 5)	PI_{JADE}	-22	-24	-24	-25	SNR_1	11,8	15,3	17,5	19,7
	PI_{EASI}	-19	-21	-23	-24	SNR_2	10,3	14,9	20,0	24,9

Table 3. Performance for colored Gaussian noise.

more robust to a bad pre-separation de-noising with white as well as colored Gaussian noise. Moreover its implementation is easier since there is no trade-off for determining the pre-separation scale.

Finaly, we addressed the fetal ECG extraction from sensors located on the mother's skin [9]. Preliminary experiments [10], performed with success from strong noisy signals, confirm the efficacy of these methods.

References

1. J. Hérault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimtrique en apprentissage non supervis. In *Gretsi*, volume 2, pages 1017–1020, Nice, France, May 1985.
2. C. Jutten and A. Taleb. Source separation: from dusk till dawn. In *Independent compoment analysis 2000*, pages 15–26, Helsinki, Finlande, June 2000.
3. S.I. Amari and A. Cichocki. *Adaptive Blind Signal and Image Processing, Learning Algorithms and Applications.* Wiley, 2002.
4. A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* Wiley, 2001.
5. A. Paraschiv-Ionescu, C. Jutten, K. Aminian, and *al.* Source separation in strong noisy mixtures: a study of wavelet de-noising pre-processing. In *ICASSP'2002*, Orlando, Floride, 2002.
6. S. Mallat. *A wavelet tour of signal processing.* Academic Press, second edition, 1999.
7. H.H. Yang, S.I. Amari, and A. Cichocki. Information-theoric approach to blind separation of sources in non-linear mixture. *Signal Processing*, 64(3):291–300, February 1998.
8. R.R. Coifman and D.L. Donoho. Translation-invariant de-noising. In *Wavelets and statistics, Springer lecture notes in Statistics 103*, pages 125–150. New York: Springer-Verlag.
9. L. De Lathauwer, D. Callaerts, B. De Moor, and al. Fetal electrocardiogram extraction by source subspace separation. In *Proc. IEEE Workshop on HOS*, pages 134–138, Girona, Spain, June 12–14 1995.
10. B. Rivet, C. Jutten, and V. Vigneron. Wavelet de-noising for blind source separation in noisy mixtures. Technical report, Technical report for the BLInd Source Separation project (BLISS IST 1999-14190)., 2003.