I&ANN98

Adaptive Blind Elimination of Artifacts in ECG Signals

Allan Kardec Barros[1], **Ali Mansour**[2]and **Noboru Ohnishi**[1,2]

1 - Graduate School of Engineering
Department of Information Engineering, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya 464-01, Japan.
2 - RIKEN BMC Research Center, Nagoya, Japan.
E-mail: allan@ohnishi.nuie.nagoya-u.ac.jp

Abstract— In this work, we deal with the elimination of artifacts (electrodes, muscle, respiration, etc.) from the electrocardiographic (ECG) signal. We use a new tool called independent component analysis (ICA) that blindly separate mixed statisticaly independent signals. ICA can separate the interference, even if they overlap in frequency. In order to estimate the mixing parameters in real-time, we propose a self-adaptive step-size, derived from the study of the averaged behavior of those parameters, and a two-layers neural network. Simulations were carried out to show the performance of the algorithm using a standard ECG database.

Keywords: Independent component analysis, blind separation, adaptive filtering, cardiac artifacts, ECG analysis.

1 Introduction

Many attempts were carried out to eliminate corrupting artifacts from the actual cardiac one when measuring the electrocardiographic (ECG) signal.

Cardiac signals show the well known repeating and almost periodic pattern. This caracteristic of physiological signals was already explored in some works(e.g, [5, 17, 21]) by synchronizing the parameters of the filter with the period of the signal. However, those filters fail to remove the interference when it has the same frequency of the cardiac signal.

Recently, many works were carried out in the field of blind source separation (BSS), using a new tool called *independent component analysis* (ICA). This large number of works may be explained because the ICA algorithms are in general elegant, simple and may deal with signals that second order statistics (SOS) methods[1] in general do not work. This is because SOS algorithms usually search for a solution that *decorrelates* the input signals while ICA looks for an *independent solution*.

ICA is based on the following principle. Assuming that the original (or source) signals have been mixed linearly, and that these mixed signals are available, ICA finds in a *blind* manner a linear combination of the mixed signals which recovers the original source signals, possibly re-scaled. This is carried out by using the principle of entropy maximization of non-linearly transformed signals.

Our main scope here is not necessarily to propose new algorithms. There are many of them already available. Our study goes toward speed of convergence and quality of the output signal.

In this work we propose a self-adaptive step size for ICA algorithms. This study was motivated by the necessity of a faster convergence, since we are mainly thinking in the implementation of this system for real-time. Instead of dealing with the non-linear cost-function of ICA algorithms which would be optimum, we carry out our analysis in a mean squared framework. For this approach, we can solve the problem of bounds to the step-size and derive the optimum one for one step convergence. In this field, there is the work of Douglas and Cichocki[15], with focus on decorrelation networks. Cichocki and his colleagues [11] also proposed a self-adaptive step-size. However, our attempt here is to find a step-size which is directly based on the evolution of the algorithm.

Moreover, we propose a neural network consisting of two layers of ICA algorithms. Some works [6, 16, 9] suggested to carry out whitening before the ICA algorithm in order to orthogonalize the inputs, which yields a faster convergence. The basis of our two-layer network is the same. However, we argue that using a cascade of two ICA algorithms is a stronger principle, because both are searching for independent solutions. Belouchrani *et. al.*[8] also proposed a multi-layer network, but they were not interested in comparing the multi-layer results with the

[1] such as the one proposed in [5, 17, 21].

pre-whitening. This is carried out here, by simulations, for different initial conditions.

2 Independent Component Analysis (ICA)

The principle of ICA may be understood as follows. Consider n source signals $\mathbf{s} = [s_1, s_2, .., s_n]^T$ arriving at m receivers. Each receiver gets a linear combination of the signals, so that we have

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{A}$ is an $m \times n$ matrix, and $\mathbf{n}$ is the noise, which is omitted because it is usually impossible to distinguish noise from the source signals, therefore, we omit it from now on. The purpose of ICA is to find a matrix $\mathbf{B}$, that multiplied by $\mathbf{A}$, will cancel the mixing effect. For simplicity, we assume that matrix $\mathbf{A}$ is a $n \times n$ invertible square matrix. Ideally, $\mathbf{B}\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity.

The system output is then given by

$$\mathbf{z} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{s} = \mathbf{C}\mathbf{s}, \tag{2}$$

where the elements of vector $\mathbf{s}$ must be *mutually independent*. In mathematical terms, it means that the joint probability density of the source signals must be the product of the marginal densities of the individual sources

$$p(\mathbf{s}) = \Pi_{i=1}^{M} p(s_i) \tag{3}$$

Thus, instead of searching for a solution that uncorrelates the signals, ICA looks for the most independent signals. As one can see, this principle is much stronger.

3 Deriving ICA

With ICA, one wants to find a way to estimate the true distribution $p(\mathbf{s}, \theta)$ of a random variable, given the samples $\mathbf{z}_1, \ldots, \mathbf{z}_N$. In other words, ICA is a probability density estimator, or density shaper.

Given the modifiable parameters $\hat{\theta}$, we should find a density estimator $\hat{p}(\mathbf{z}, \hat{\theta})$ of the true density $p(\mathbf{s}, \theta)$. This may be performed by entropy maximization, mutual information minimization, maximum likelihood or Kullback-Liebler (K-L) divergence. We take, for instance, the K-L divergence, given by

$$l\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\} = \int p(\mathbf{s}, \theta) log \frac{p(\mathbf{s}, \theta)}{\hat{p}(\mathbf{z}, \hat{\theta})} d\mathbf{z}. \tag{4}$$

A small value of the K-L divergence $l\{p(\mathbf{s}), \hat{p}(\mathbf{z}, \hat{\theta})\}$ indicates that $\hat{p}(\mathbf{z}, \hat{\theta})$ is close to the true density $p(\mathbf{s}, \theta)$. Hence, we should minimize $l\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\}$, and this can be done by using a gradient method. However, instead of the conventional Euclidean gradient method [6], that reads

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \mu_k \frac{\partial}{\partial \hat{\theta}_k} l\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\}, \tag{5}$$

we rather use the following gradient

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \mu_k \Im \frac{\partial}{\partial \hat{\theta}_k} l\{p(\mathbf{s}, \theta), \hat{p}(\mathbf{z}, \hat{\theta})\}, \tag{6}$$

where $\Im$ is a positive definite matrix.

This is called the relative [9], natural, or Riemannian gradient [1, 2]. This algorithm works better in general because the parameter space of neural networks is Riemannian [3].

To obtain a better estimation, Pearlmutter and Parra [20] derived an algorithm that extracts many parameters related to the signal, and therefore their parameter space $S = \{\hat{\theta}\}$ was built with many variables. However, in most of the works, the gradient method shown above was derived using only the weight matrix $\mathbf{B}$ as the parameter to be estimated. For this case, we have $\Im = \mathbf{B}^T\mathbf{B}$ and the weights of $\mathbf{B}$ are updated by [9]

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \mu_k [\mathbf{I} - N(\mathbf{z}_k)] \mathbf{B}_k, \tag{7}$$

where $N(\cdot)$ is a non-linear function. In this work, we use the following function, as suggested by Bell and Sejnowski[6]

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mu_k (\mathbf{I} - \mathbf{y}_k \mathbf{z}_k^T) \mathbf{B}_k, \tag{8}$$

with $\mathbf{z} = \mathbf{B}\mathbf{x}$ and $\mathbf{y} = \tanh(\mathbf{z})$.

And this is the *secret* of ICA: this non-linearity *tries* to shape the sources distribution. In other words, if one expands this non-linearity in a series of Taylor, higher moments appear. For example, if we use one sigmoidal function, which is used frequently in neural networks, one can see that

$$\tanh(u) = u - \frac{u^3}{3} + \frac{2u^5}{15} + \cdots \tag{9}$$

It should be added, however, that even though these methods are said to *blindly* estimate the sources, some *prior* knowledge is necessary in order to choose this non-linearity. As we have seen, ICA is also known to be a density estimator. In order words, the non-linearity g should be chosen so that

$$y_i = g(u) \approx \int_{-\infty}^{u} f_s(v) dv, \tag{10}$$

where f_s is the density of s.

In practice, it is not very necessary that this equation is true. For signals with a super-Gaussian distribution (kurtosis > 0), it did not pose as a problem to separate

them using (8). In the case of sub-Gaussian signals, Cichocki and his colleagues [12] suggested [12] the following equation

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mu_k(\mathbf{I} - \mathbf{z}_k\mathbf{y}_k^T)\mathbf{B}_k\,, \tag{11}$$

An interesting discussion about this topic was carried out by Amari [4].

3.1 Indefinition of the Solution

Because the system works in a *blind* manner, $\mathbf{B}$ does not necessarily converge to the inverse of $\mathbf{A}$. We can only affirm that $\mathbf{C} = \mathbf{DP}$, where $\mathbf{D}$ is a diagonal, and $\mathbf{P}$ is a permutation matrix [13]. Without any *a priori* information, which is the case of blind source separation, nothing can be done concerning to the permutation, but we can still normalize the weight matrix to avoid the problem of random scaling. An interesting solution is to preserve the energy of the input signal, normalizing the weights by [14]

$$\mathbf{W} = |\det(\mathbf{W})|^{\frac{1}{n}}\mathbf{W}. \tag{12}$$

3.2 Equivariance Property

An *equivariant* estimator $\Lambda(\cdot)$ for an invertible $n\times n$ matrix M is defined as [9]

$$\Lambda(M\,\mathbf{z}_k) = M\Lambda(\mathbf{z}_k). \tag{13}$$

This property can be applied to relative gradient algorithms. Multiplying both sides of (7) by $\mathbf{A}$ yields

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \mu_k\{\mathbf{C}_k - N(\mathbf{C}_k\mathbf{s}_k)\}. \tag{14}$$

Therefore, the trajectory of the global system $\mathbf{C} = \mathbf{BA}$ is *independent* of $\mathbf{A}$.

3.3 Filtering

Sometimes a filtering operation can be very useful in some ill-conditioned mixing problems. Here we discuss about a filtering operation that *preserves* the mixing matrix. With this, one can modify the mixing (ill-conditioned) condition by some pre-processing so that the signals can be separated. If the matrix is preserved and its inverse (or a scaled version of it) can be estimated by ICA, then one can easily recover the source signals.

A causal filter for the mixed vector $\mathbf{x}$ with impulse response $\mathbf{H}(t,\tau)$, can be described by

$$\mathbf{y}(t) = \int_{-\infty}^{t}\mathbf{H}(t,\tau)\mathbf{x}(\tau)\mathrm{d}\tau = \int_{-\infty}^{t}\mathbf{H}(t,\tau)\mathbf{A}\mathbf{s}(\tau)\mathrm{d}\tau. \tag{15}$$

We assume that this first-order linear system may be *time-variant*. Moreover, we should find a $\mathbf{H}(t,\tau)$ so that the following holds

$$\mathbf{y}(t) = \mathbf{A}\int_{-\infty}^{t}\mathbf{H}(t,\tau)\mathbf{s}(\tau)\mathrm{d}\tau. \tag{16}$$

In other words, the filtering operation should not alter the structure of matrix $\mathbf{A}$. For this to happen, the impulse response $\mathbf{H}(t,\tau)$ should be a diagonal matrix with the same elements, i.e., $\mathbf{H}(t,\tau) = h(t,\tau)\mathbf{I}$, which implies that the elements of vector $\mathbf{x}$ should be passed through the same filter.

4 Time Varying Step-size for ICA Algorithms

From (7), the output correlation matrix will be given by

$$\mathbf{T}_k = E[\mathbf{z}_k\mathbf{z}_k^T] = E[\mathbf{B}_k\mathbf{x}_k\mathbf{x}_k^T\mathbf{B}_k^T]. \tag{17}$$

In this analysis, we make use of the independent assumption. This effectively implies that $\mathbf{x}_k$ is independent of former values and that the elements of $\mathbf{B}_k$ are mutually independent. This is very common in the field of adaptive filtering to use this assumption, even though it is rarely true in practice.

If the fluctuations in the elements of $\mathbf{B}_k$ are small, we can thus rewrite (17) as[2]

$$\mathbf{T}_k = E[\mathbf{B}_k]\mathbf{R}E[\mathbf{B}_k^T] \tag{18}$$

where $\mathbf{x}$ is assumed to be stationary constant, in other words, the input vector $\mathbf{R} = E[\mathbf{x}_k\mathbf{x}_k^T]$ is constant. However, the same cannot be said about $\mathbf{T}_k = E[\mathbf{z}_k\mathbf{z}_k^T]$, thus it is assumed to be non-stationary.

Using (7) and (18), we can write

$$\mathbf{T}_{k+1} = E[(1+\mu_k)\mathbf{I} - \mu_k\mathbf{y}_k\mathbf{z}_k^T]\mathbf{T}_k E[(1+\mu_k)\mathbf{I} - \mu_k\mathbf{z}_k\mathbf{y}_k^T]$$
$$(1+\mu_k)^2\mathbf{T}_k - \mu_k(1+\mu_k)\mathbf{T}_k\mathbf{P}_k - \mu_k(1+\mu_k)\mathbf{P}_k^T\mathbf{T}_k + \mu_k^2\mathbf{P}_k^T\mathbf{T}_k\mathbf{P}$$

If we assume that the variation of $\mathbf{z}_k$ is bounded to the interval $[-1,1]$, we can then say that in this limit $\mathbf{y}_k \approx \mathbf{z}_k$ and $\mathbf{P}_k \approx \mathbf{T}_k$. Then, (19) can be written as

$$\mathbf{T}_{k+1} = (1+\mu_k)^2\mathbf{T}_k - 2\mu_k(1+\mu_k)\mathbf{T}_k^2 + \mu_k^2\mathbf{T}_k^3. \tag{20}$$

There is a unitary matrix $\mathbf{Q}$ that diagonalizes $\mathbf{T}_k$ so that $\Lambda_i = \mathbf{Q}^T\mathbf{T}_i\mathbf{Q}$ and $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Thus, we can rewrite (20) as

$$\Lambda_{k+1} = (1+\mu_k)^2\Lambda_k - 2\mu_k(1+\mu_k)\Lambda_k^2 + \mu_k^2\Lambda_k^3. \tag{21}$$

Notice that when deriving (21) from (20) the orthogonal property of $\mathbf{Q}$ was used[3].

[2] The following steps are similar to the one carried out by Douglas and Cichocki[15].

[3] For example $\mathbf{Q}^T\mathbf{T}\mathbf{V}_k\mathbf{Q} = \mathbf{Q}^T\mathbf{T}\mathbf{Q}\mathbf{Q}^T\mathbf{V}_k\mathbf{Q}$.

From (21), the eigenvalues of $\mathbf{T}_k$ are the elements of Λ_k, and are given by

$$\lambda_{k+1,i} = (1+\mu_k)^2\lambda_{k,i} - 2\mu_k(1+\mu_k)\lambda_{k,i}^2 + \mu_k^2\lambda_{k,i}^3. \quad (22)$$

For uniform convergence in a mean-squared sense, it is required that $\lambda_{k+1,i} < \lambda_{k,i}$, which yields the following bounds for the step-size[4]

$$0 < \mu_k < \frac{2}{\lambda_{k,i} - 1}. \quad (23)$$

From (23), the optimum step-size which will give one-step convergence is

$$\mu_{opt} = \frac{2}{\lambda_{k,i} - 1} \quad (24)$$

Using (23) and the assumption that $\mathbf{y}_k \approx \mathbf{z}_k$, we propose then to use the following step-size to update the weight matrix

$$\mu_k = \frac{2}{\mathbf{y}_k^T\mathbf{z}_k + 1} \quad (25)$$

When proposing the step-size above, we had in mind the following

$$\frac{2}{\sum_i \lambda_{k,i} + 1} < \frac{2}{\lambda_{k,i} + 1} < \frac{2}{\lambda_{k,i} - 1} \quad (26)$$

5 A Network For Fast Blind Separation

In this work, we are mainly interested in using ICA to filter noises that are possibly overlapping in frequency the cardiac signal. However, we do not want that this use of ICA implies in a longer time of convergence. This is because ICA is usually slower to converge than the LMS, because it is also estimating higher-order moments. Thus, we propose an architecture to deal with this matter.

It is very common among researchers to use a whitening filter before the ICA algorithm itself. The reason is that the whitening carries out a decorrelation between the input signals. Then, the ICA work is reduced to estimate the moments higher than two, with this, one gains in speed.

Here, we propose a different reasoning. Instead of using only second-order statistics, we suggest the use of a network that substitutes the whitening by an ICA algorithm itself. With this, we are not only estimating the second, but also higher-order moments. We will see that this simple substitution implies in a much *faster* convergence.

The architecture includes other points to improve the speed of convergence as shown in Fig. 1 . In resume, they are given below.

- Pre-process the mixed signals by a high-pass filter operation that obeys (16). Later we will discuss why this is important.
- Use a time-varying step-size for faster convergence as in (25).
- Use a two-layer network. The two layers are cascaded in series and the first layer is only used for convergence. It is *turned off* after a given number of iterations.
- Use batch update. This is because the block size alters the convergence of the algorithm.

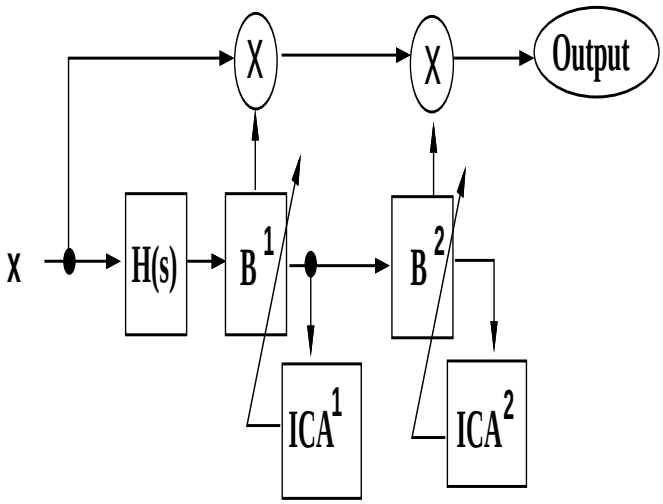

Figure 1: Block diagram of the proposed method.The signal is inputed into a high-pass filter, then to the first layer ($\mathbf{ICA}^1$), which is updated only up to a number of iterations. The other layer ($\mathbf{ICA}^2$) is always "turned on".

6 Simulations

We have carried out simulations to test the validity of the proposed method. The simulation consisted in mixing actual ECG and electrode motion artifact (usually the result of intermittent mechanical forces acting on the electrodes) signals. We used signals from the MIT-BIH noise stress test database, which are standard for testing ECG analysers. Their power spectrum is shown in Fig. 2. Notice that the fundamental frequency of the ECG signal is overlapped in frequency by the electrode artifact one. The mixing was carried out using different random matrices[5]. The first layer ($\mathbf{ICA}^1$) in Fig.1 was turned off after 1000

[4] This derivation is carried out simply by substituting $\lambda_{k+1,i} < \lambda_{k,i}$ in (22).

[5] By the equivariance property, we can conclude that this would be equivalent to keeping the mixing matrix constant and changing the weight initial value.

iterations. For all cases, we initialized the weight matrix by the identity matrix. The filter cutoff frequency was 2 Hz, and the weights were updated every block size of 50 iterations.

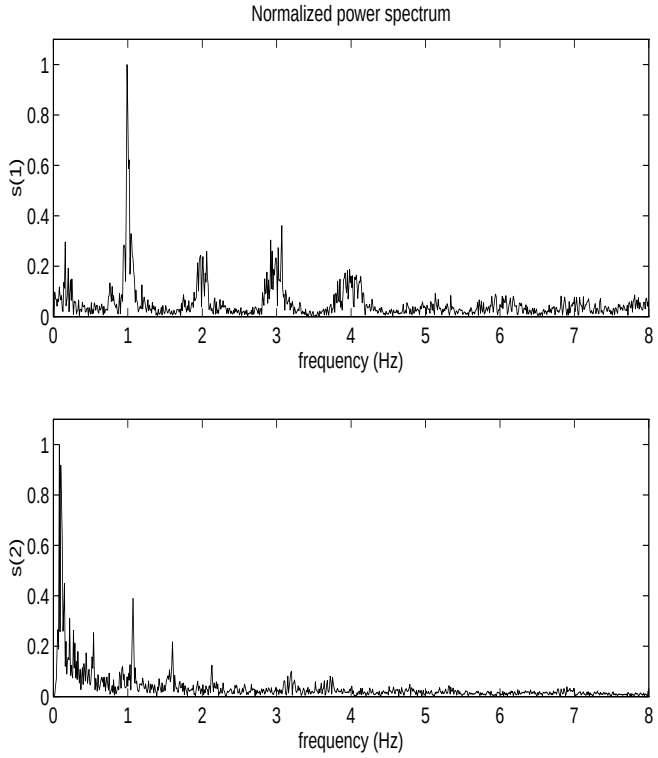

Figure 2: Power spectrum of the two source signals: ECG and respiration. Notice that the first harmonic of the ECG signal is overlapped in frequency by the respiratory one.

7 Results

We used as figure of merit to measure the quality of separation at the k-th iteration the following equation

$$\wp_k = 100 \sum_{j=1,2} (\tilde{c}(k)_j - 0.5),$$

$$\tilde{c}(k)_j = max\{\frac{|c_{i,j}|}{\sum_i |c_{i,j}|}\}, \quad \text{for} \quad j = 1,2. \qquad (27)$$

With this index, we are measuring how far the matrix $\mathbf{C}$ is from the solution $\mathbf{DP}$ at each iteration. When $\mathbf{C} = \mathbf{DP}$, only one element at each column/line is different from zero. The index $\wp$ will be 100 for the best case, and will be *null* for the worst.

We have extensively run the proposed algorithm for different randomly mixed vectors. Fig.3 shows examples of the results for three different initial matrices **A1**, **A2** and **A3**. The red, blue and green lines correspond respectively to the simulation carried out using:

- The proposed method as in Fig.1, with two ICA algorithms an in (11) and filtering;
- The same configuration of the item above, but instead of an ICA algorithm, we used a *whitening* one in the first layer. This was realized by substituting the nonlinear function in (8) by $\mathbf{y} = \text{erf}(\mathbf{z})$ [6]
- One ICA algorithm as in (11).

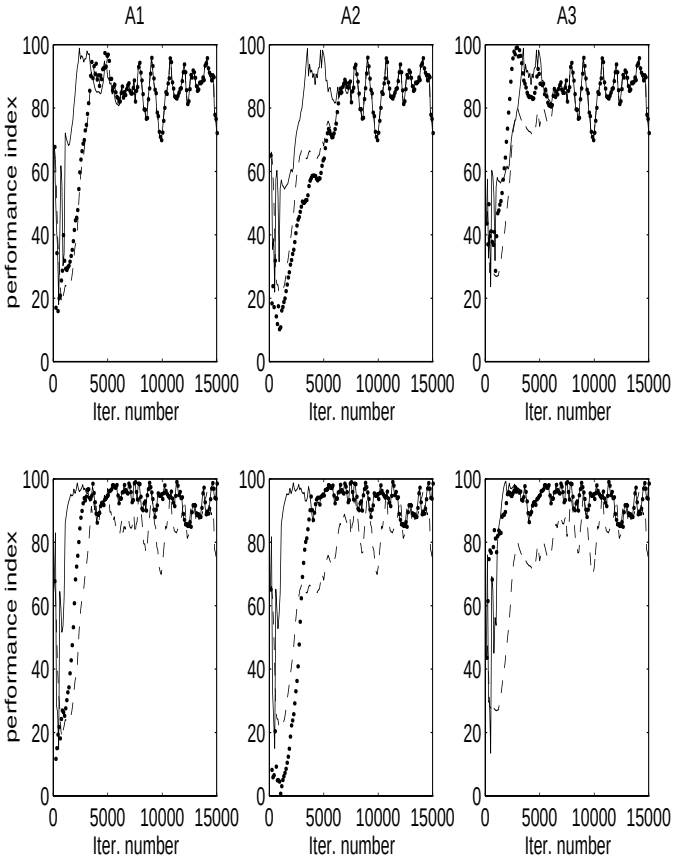

Figure 3: Values of $\wp$ as in (27) for weight matrices **A1**, **A2** and **A3**. The labels are as follows: **solid**: ICA+ICA; **dotted**: whitening+ICA; **dashed**: only ICA. "100" indicates optimum separation. In the top line, *no filtering was used* while in the bottom one, filtering was used for the case of ICA+ICA and whitening+ICA. The ICA was printed for the sake of comparison.

The index as in (27) was calculated for each simulation, ant they are shown in Fig.3. The signals recovered by the proposed network are shown in Fig.4. The "recovered signals" were obtained after normalizing the weights as in (12).

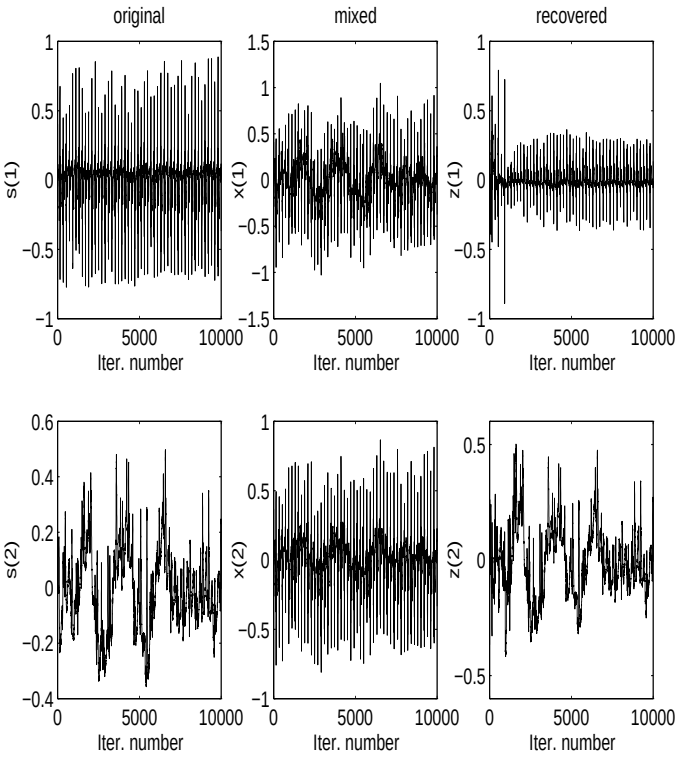

Figure 4: An example of the original, mixed and recovered signal by the proposed network. In this case, $\mathbf{A} = [11; 0.91]$

shows this result. We can see that the self-adaptive learning rate allowed a much faster learning, without diverging.

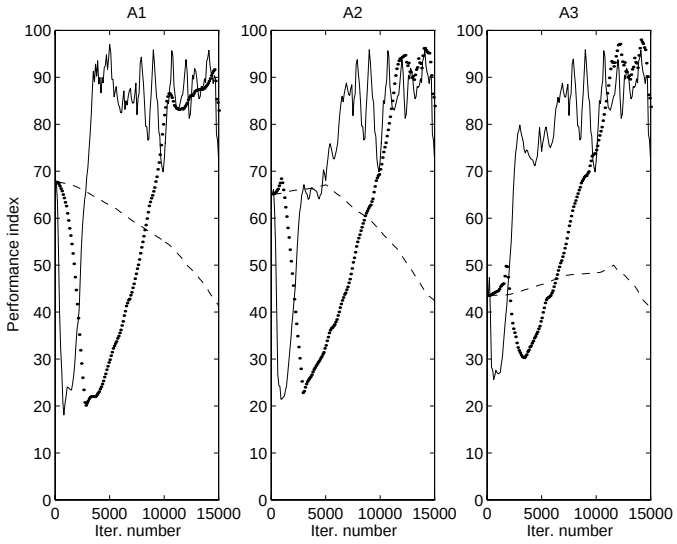

Figure 5: Values of $\wp$ as in (27) for weight matrices **A1**, **A2** and **A3**, as in Fig. 4. The labels are as follows: **solid**: ICA with adptive learning rate; **dotted**: ICA with a constant learning rate of 2×10^{-4}; **dashed**: ICA with a constant learning rate of 2×10^{-5}.

8 Discussion

By looking at Fig.3 we can take the following conclusions:

- The filtering is important in order to have less variance after convergence. This is because the lower frequency signal (trend) was removed. The trend is usually accounted in the literature as non-stationary mean.
- The two-layer ICA network performed better in general than the network with pre-whitening. While for matrix **A3** as in the top row of Fig.3, the whitening showed a slightly better performance, for the *with filtering* case, as in the bottom rows, the two-layer network outperformed the two others.

Another point that should be emphasized is that of the adaptive step size. When we started using ICA, the greatest problem in our point of view was that of step-size. Since we wanted a fast convergence, we had to fix the step-size at some upper value, otherwise the algorithm would not converge. Therefore, for these ECG and electrode noise data, we found *heuristically* an upper bound of 2×10^{-4} for the learning rate. We compared then this learning rate with the adaptive one, derived here. Fig. 5

Some words are necessary about the two-layers network. Some works, e.g., [7], [16] have proposed to carry out whitening before the ICA processing in order to orthogonalize the input components. In the same way, we have used the first layer to *force* the algorithm to search for independent components. We argue, however, that the cascading two ICA algorithms is a stronger principle because the first layer looks for an *independent* rather than an orthogonal solution. Contrary to whitening, that uses only second order statistics, ICA makes also use of higher order moments. This can be observed in Fig.3, where filtering was used and the data is stationary. By looking at the red line, we can see that the the convergence with two-layers was *much better* than that with whitening which, as with one ICA layer only, sometimes did not converge.

Probably the reader is asking why we did not use the two layers in the whole trajectory, but rather, we switched it off after some iterations. We carried it out, but the variance after convergence for such configuration was higher. Therefore, roughly speaking, the first layer works as a *propulsion* to *put the algorithm in the way* to converge to one of the solutions **C=DP**.

9 Conclusions

In this work, we proposed an architecture to blindly separate linearly mixed signals, based on the independent component analysis principle. The architecture consisted of a high-pass filter, a two-layer network based on ICA algorithm and a self-adaptive step-size. The self-adaptive step-size was theoreticaly derived from the mean behavior of the output signal.

The proposed network composed of two ICA algorithms converged faster than the one composed of whitening plus an ICA algorithm, where *whitening* stands for an algorithm designed to orthogonalize the input signals. We argued that the two-layer network of ICA algorithm behaves better because the first layer is searching for an independent solution, rather than an orthogonal. This conclusion was confirmed by simulations. The proposed self-adaptive step-size also leaded to a fast convergence, though with a greater error.

References

[1] S. Amari, A.Cichocki, H. H. Yang, "A new learning algorithm for blind signal separation", *Advances in Neural Information Processing Systems*, 8, MIT press, 1996.

[2] S. Amari, A.Cichocki, H. H. Yang, "Gradient learning in structured parameter spaces: adaptive blind separation of signal sources", *WCNN*, 1996.

[3] S. Amari, "Information geometry of the EM and em algorithms for neural networks", *Neural Networks*, Vol. 8, No. 9, pp. 1379 - 1408, 1995.

[4] S. Amari, ""Stability Analysis of Adaptive Blind Source Separation" *Neural Networks*, to appear.

[5] A. K. Barros and N. Ohnishi "MSE Behavior of Biomedical Event-Related Filters" *IEEE Trans. Biomed. Eng.*, (to appear) 1997.

[6] A.J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution" *Neural Computation*, 7, pp. 1129 - 1159, 1995.

[7] A.J. Bell and T. J. Sejnowski, "Fast blind separation based on information theory" *Proc. Intern. Symp. on Nonlinear Theory and Applications (NOLTA)*, vol. 1, 43-47, Las Vegas, Dec. 1995.

[8] Belouchrani A., Cichocki and Abed Meraim K."A Blind Identification and Separation Technique via Multi-layer Neural Networks" *Proc. ICONIP'96*, Springer-Verlag Singapore Ltd, Vol. 2, 1195-1200, 1996.

[9] J-F. Cardoso."On the performance of source separation algorithms" In Proc. EUSIPCO, pp. 776-779, Edinburgh, September 1994

[10] J-F. Cardoso and B. H. Laheld."Equivariant adaptive source separation" *IEEE Trans. Signal Process.*, SP-44 , pp.3017 - 3030, 1996.

[11] A.Cichocki , S. Amari, M. Adachi and W. Kasprzak. "Self-adaptive neural networks for blind separation of sources", *Proc. Intern. Symposium on Circuits and Systems*, Vol. 2, No. 4, pp. 157-160, 1996.

[12] A.Cichocki, S. Amari, M. Adachi and W. Kasprzak. "Local adaptive learning algorithms for blind separation of natural images", *Neural Network World*, Vol. 6, No. 4, pp. 515-523, 1996.

[13] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, 24, pp. 287 - 314, 1994.

[14] G. Deco and W. Brauer, "Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures" *Neural Networks*, Vol. 8, pp. 525 - 535, 1995.

[15] S. Douglas and A. Cichocki , "Neural networks for blind decorrelation of signals" *IEEE Trans. Signal Processing*, (to appear).

[16] Karhunen, J."Neural Approaches to Independent Component Analysis and Source Separation.", *Proc. 4th European Symposium on Artificial Neural Networks (ESANN'96)*, 1996, Bruges, Belgium.

[17] P. Laguna, R. Janè, O. Meste, P. Poon, P. Caminal, H. Rix and N. V. Thakor., "Adaptive filter for event-related bioelectric signals using an impulse correlated reference input: Comparison with signal averaging techniques" *IEEE Trans. Biomed. Eng.*, BME-39 , pp.1032-1043, 1992.

[18] S. Makeig, A. J. Bell, T-P Jung, T. Sejnowski "Independent component analysis of electroencephalographic data" *Advances in Neural Information Processing Systems*, 8, MIT press, 1996.

[19] A. Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill, 1991.

[20] B. A. Pearlmutter and L. C. Parra, "A Context-Sensitive Generalization of ICA", *International Conference on Neural Information Processing*, Hong Kong, September 1996.

[21] C. Vaz, X. Kong and N.V. Thakor, "An adaptive estimation of periodic signals using a Fourier linear combiner"*IEEE Trans. Sig. Process.*,ASSP-42, pp.1-10,1994.