Proportionate Algorithms for Blind Source Separation

Michele Scarpiniti, Danilo Comminiello, Simone Scardapane,
Raffaele Parisi, and Aurelio Uncini

Department of Information Engineering, Electronics and Telecommunications (DIET),
"Sapienza" University of Rome, via Eudossiana 18, 00184, Rome
{michele.scarpiniti,danilo.comminiello,simone.scardapane,
raffaele.parisi,aurelio.uncini}@uniroma1.it

Abstract. In this paper we propose an extension of time-domain Blind Source Separation algorithms by applying the well known proportionate and improved proportionate adaptive algorithms. These algorithms, known in the context of adaptive filtering, are able to use the sparseness of acoustic impulse responses of mixing environments and give better performances than standard algorithms. Some preliminary experimental results show the effectiveness of the proposed approach in terms of convergence speed.

Keywords: Blind Source Separation, Independent Component Analysis, Proportionate algorithms, Improved proportionate.

1 Introduction

Blind Source Separation (BSS) applied to speech and audio signals is an attractive research topic in the field of adaptive signal processing [1,2]. The problem is to recover original sources from a set of mixtures recorded in an unknown environment. While several well-performing approaches exist when the mixing environment is instantaneous, some problems arise in convolutive environments.

Several solutions were proposed to solve BSS in a convolutive environment [3,2]. Some of these solutions work in time domain, others in frequency domain. Each of them have some advantages and disadvantages, but there is not a unique winner approach [4].

In addition, when working with speech and audio signals, convergence speed is an important task to be performed. Since impulse responses of standard environments, e.g. office rooms, are quite sparse, some authors have proposed to incorporate sparseness in the learning algorithm [5,6]. The idea is to introduce a weighting matrix in the update equation, that can give more emphasis to the most important part of the impulse response. In particular a proportionate [5] and an improved proportionate [7] algorithms were proposed for supervised signal processing applications, like acoustic echo cancellation.

In this paper we aim to extend these proportionate algorithms to the BSS problem, in the hope that they can be effective also for unsupervised case. Hence a proportionate and an improved proportionate version of the well-known time-domain Torkkola's algorithm [8,9] will be proposed. Some preliminary results, that demonstrate the effectiveness of the proposed idea in terms of convergence speed, are also presented.

S. Bassis et al. (eds.), *Recent Advances of Neural Network Models and Applications*,
Smart Innovation, Systems and Technologies 26,
DOI: 10.1007/978-3-319-04129-2_10,

The rest of the paper is organized as follows: Section 2 introduces the BSS problem in convolutive environments. Then Section 3 describes the proposed algorithm, while Section 4 shows some experimental results. Finally Section 5 draws our conclusions.

2 Blind Source Separation for Convolutive Mixtures

Let us consider a set of N unknown and independent sources denoted as $\mathbf{s}[n] = [s_1[n], \ldots, s_N[n]]^T$, such that the components $s_i[n]$ are zero-mean and mutually independent. Signals received by an array of M sensors are denoted by $\mathbf{x}[n] = [x_1[n], \ldots, x_M[n]]^T$ and are called mixtures. For simplicity we consider the case of $N = M$.

The convolutive model introduces the following relation between the i-th mixed signal and the original source signals

$$x_i[n] = \sum_{j=1}^{N} \sum_{k=0}^{K-1} a_{ij}[k] s_j[n-k], \qquad i = 1, \ldots, M \tag{1}$$

The mixed signal is a linear mixture of filtered versions of the source signals, $a_{ij}[k]$ represents the k-th mixing filter coefficient and K is the number of filter taps. The task is to estimate the independent components from the observations without resorting to *a priori* knowledge about the mixing system and obtaining an estimate $\mathbf{u}[n]$ of the original source vector $\mathbf{s}[n]$:

$$u_i[n] = \sum_{j=1}^{M} \sum_{l=0}^{L-1} w_{ij}[l] x_j[n-l], \qquad i = 1, \ldots, N \tag{2}$$

where $w_{ij}[l]$ denotes the l-th mixing filter coefficient and L is the number of filter taps.

The weights $w_{ij}[l]$ can be adapted by minimizing some suitable cost function. A particular good choice is to maximize joint entropy or, equivalently, to minimize the mutual information [1,3]. Different approaches can be used, for example implementing the de-mixing algorithm in the time domain or in the frequency domain. In this paper we adopt the time-domain approach, using the algorithm proposed by Torkkola in [8] and based on the feedback network shown in Figure 1 (for the particular case of $M = N = 2$) and described mathematically by:

$$u_i[n] = \sum_{l=0}^{L-1} w_{ii}[l] x_i[n-l] + \sum_{j=1, j \neq i}^{N} \sum_{l=1}^{L-1} w_{ij}[l] u_j[n-l], \qquad i = 1, \ldots, N \tag{3}$$

This latter will be used in the paper for achieving source separation in time domain by maximizing the joint entropy of a nonlinear transformation of network output: $y_i[n] = f(u_i[n])$, with $f(\cdot)$ a suitable nonlinear function, very close to the source cumulative density function [1,8]. In this work we use $f(\cdot) = \tanh(\cdot)$. The k-th weight of the de-mixing filter $w_{ij}[k]$ is adapted by using the general rule:

$$w_{ij}^{p+1}[k] = w_{ij}^{p}[k] + \mu \Delta w_{ij}^{p}[k], \tag{4}$$

where in particular the stochastic gradient method can be used, μ is the learning rate and p is the iteration index.

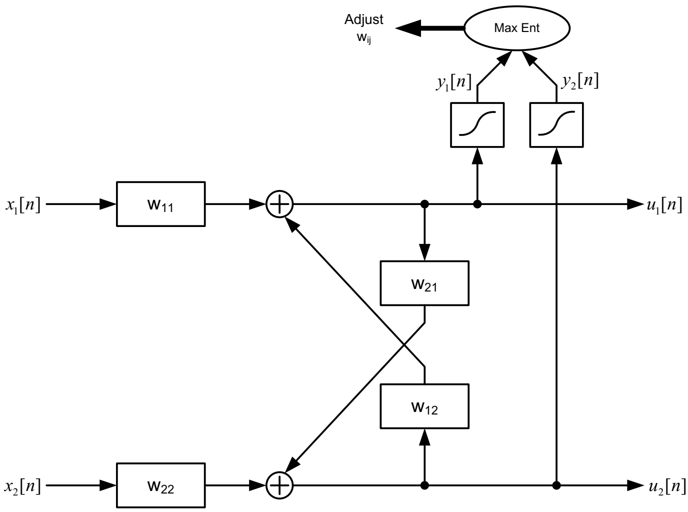

Fig. 1. The architecture for BSS in time domain in the particular case of $M = N = 2$

3 The Proposed Algorithm

Standard BSS algorithms were proposed so far in several works [3,9,10,8]. More recently, some authors have underlined the importance of sparseness of the impulse response $a_{ij}[n]$ [6]. In order to improve the performance of standard adaptive algorithms (LMS and NLMS [11]), that do not take into account sparseness, these authors proposed some modifications introducing the so-called Proportionate NLMS (PNLMS) [5] and Improved Proportionate NLMS (IPNLMS) [7]. The resulting algorithms derive from a more general class of regularized gradient adaptive algorithms [12,13].

In this kind of algorithms, the update term is simply multiplied with a matrix $\mathbf{G}[k]$, whose entries $g_{ij}[k]$ are chosen using different criteria (NLMS, PNLMS and IPNLMS) and take into account the sparseness of the impulse response: the parameters variation is proportional to the impulse response itself

$$\mathbf{w}^{p+1} = \mathbf{w}^{p} + \mu \Delta \mathbf{w}^{p} \quad \Rightarrow \quad \mathbf{w}^{p+1} = \mathbf{w}^{p} + \mu \mathbf{G}\left\{\mathbf{w}^{p}\right\} \Delta \mathbf{w}^{p}.$$

The aim of this section is to extend the previous ideas on proportionate adaptive algorithms to time-domain BSS algorithm proposed by Torkkola [8], deriving a new updating rule for the de-mixing filter matrix $\mathbf{W}[k]$.

Based on [8] and [6], the proposed modification to the Torkkola's algorithm results in the following algorithm:

$$\begin{aligned}
\Delta w_{ii}^{p}[0] &\propto g_{ii}[0]\left(\frac{1}{w_{ii}[0]} - 2y_i[n]x_i[n]\right), \\
\Delta w_{ii}^{p}[k] &\propto -2g_{ii}[k]y_i[n]x_i[n-k], \quad \text{for } k \geq 1 \\
\Delta w_{ij}^{p}[k] &\propto -2g_{ij}[k]y_i[n]u_j[n-k], \quad \text{for } k \geq 1 \text{ and } i \neq j
\end{aligned} \qquad (5)$$

where $y_i[n] = f\left(u_i[n]\right) = \tanh\left(u_i[n]\right)$.

The k-th parameter $g_{ij}[k]$ in the case of Proportionate BSS (PBSS) are chosen as follows

$$g_{ij}[k] = \frac{\gamma_{ij}[k]}{\left\|\gamma_{ij}\right\|_1}, \tag{6}$$

$$\begin{aligned} \gamma_{ij}[k] &= \max\left\{\rho \max\left[\delta_k, \left|w_{ij}[0]\right|, \ldots, \left|w_{ij}[L-1]\right|\right], \left|w_{ij}[k]\right|\right\}, \\ \gamma_{ij} &= \left[\gamma_{ij}[0], \gamma_{ij}[1], \ldots, \gamma_{ij}[L-1]\right]^T, \end{aligned} \tag{7}$$

with ρ and δ_k suitable constants.

A second proposal is the following Improved PBSS (IPBSS) choice for the k-th parameter $g_{ij}[k]$:

$$g_{ij}[k] = \frac{1-\beta}{2L} + (1+\beta)\frac{\left|w_{ij}[k]\right|}{2\left\|\mathbf{w}_{ij}\right\|_1}, \tag{8}$$

$$\mathbf{w}_{ij} = \left[w_{ij}[0], w_{ij}[1], \ldots, w_{ij}[L-1]\right]^T, \tag{9}$$

where $-1 \le \beta \le 1$ is a constant.

4 Experimental Results

Some experimental results are proposed using two speech signals sampled at 8 kHz. Three different synthetic mixing weight sets were proposed and a total of 5000 samples are used.

The first set of weights is a very simple and sparse set given by

$$\begin{array}{c} a_{11}[0] = 1, \quad a_{22}[0] = \frac{5}{6}, \ a_{12}[10] = \frac{4}{6}, \\ a_{21}[10] = \frac{3}{6}, \ a_{12}[40] = \frac{2}{6}, \ a_{21}[40] = \frac{1}{6}. \end{array} \tag{10}$$

A more dense set of weights is used in the second set, perhaps more realistically in a room that produces notable reverberation, where it was imagined that you have one microphone closer to the first source and another microphone close to the other, which results in:

$$\begin{array}{l} a_{11}[0] = a_{22}[0] = 1 \\ a_{ij}[5n-1] = \exp(-n), \quad \text{for } n = 1, \ldots, 120 \text{ and } i, j = 1, 2 \end{array} \tag{11}$$

that is, we have the strongest input respectively from the two sources in the input, with subsequent echoes every five taps of the filter that decay exponentially.

The third set has only non-zero elements for the length of the filter, but is otherwise similar to the previous weights, and is meant to simulate the same situation:

$$\begin{array}{l} a_{11}[0] = a_{22}[0] = 1 \\ a_{ij}[n] = \exp\left(-(1+0.4n)\right), \quad \text{for } n = 1, \ldots, 200 \text{ and } i, j = 1, 2 \end{array} \tag{12}$$

Furthermore, the de-mixing weights are not updated for each iteration of the learning rule, rather we sum the contributions from the learning rule over 150 iterations between each update of the filter for a robust estimation. When using the **G** matrix from the PBSS algorithm, we always use $\rho = 0.01$ and $\delta_k = 0.01$, as suggested as good values by [6]. When using IPBSS, we always use the parameter $\beta = 0.1$. In both cases we use a de-mixing filter length of $L = 200$ samples and the learning rate is set to $\mu = 1.5 \times 10^{-5}$. We will face the more complicated problem of separating real world mixtures in a future work.

Now, we ask ourself the natural question: Does the proposed method using a **G** matrix from either the PBSS method or the IPBSS method improve the results from the standard algorithm? The answer found, is that it depends on the filter length. Or likely, more generally, how sparse the filter is.

In order to evaluate the convergence of the algorithm, an estimate I of the mutual information is used as convergence and performance index [14,15].

4.1 Results of Proportionate Algorithm

As we will see in this section, there are improvements in the convergence speed over the standard algorithm. Figure 2 shows the convergence history using the estimated mutual information, and we can clearly see that there is a considerable speed-up in convergence: already at epoch number 10, we can see that using the PBSS has almost reached convergence, while the standard algorithm does not reach this level until around epoch number 30. At epoch number 20, using PBSS, the algorithm has practically reached convergence, while convergence is reached for the standard algorithm before around epoch number 80.

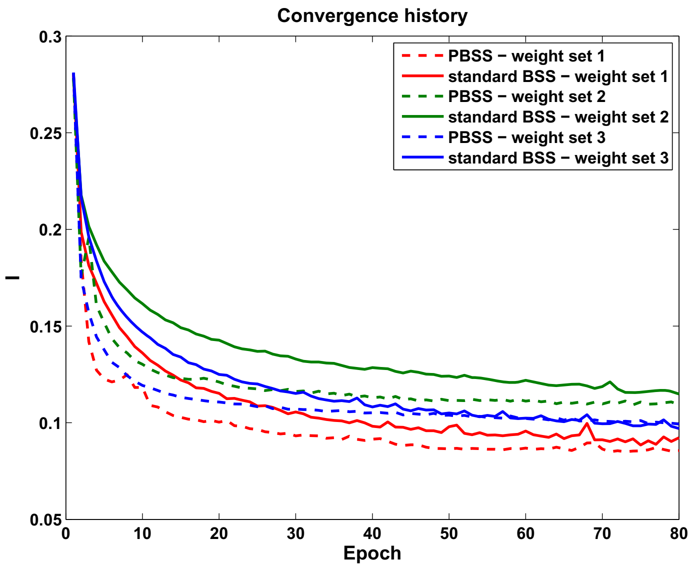

Fig. 2. Comparison of the standard algorithm to the algorithm with the **G** matrix from the PBSS method

This suggests that the algorithm converges between 3 and 4 times faster when using the **G** matrix from the PBSS method, which shows that by adaptively changing the learning rate for the different taps of the filter, we get a significantly better performance. As it can be seen in Figure 2, at convergence the solution of both algorithms reaches the same level of mutual information. It would be interesting to see if there are any differences, looking at the recovered de-mixing filters directly. Figure 3 shows the solution at convergence for both methods using the first weight set, and we easily see that the main features of both solutions are present in both filters and are approximately the same filter taps, albeit at slightly different scalings. Except for that, the are other small differences, but it seems to be mostly noise in the filter. Qualitatively, from hearing the de-mixing of the the mixed sound files, we could discern no differences between the two solutions at convergence. Thus it seems reasonable to conclude that both methods eventually converge to the same solution, but that using the **G** matrix from the PBSS method, we obtain convergence and a good solution significantly faster.

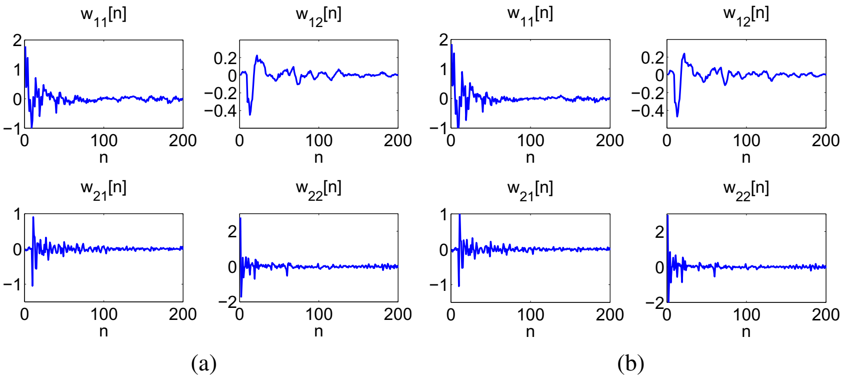

Fig. 3. Solution at convergence using the first set of weights, for a) standard algorithm, and b) PBSS algorithm

4.2 Results of Improved Proportionate Algorithm

Similar results can be obtained from the IPBSS method. Figure 4 shows the convergence history of the standard algorithm versus the proposed algorithm using the **G** matrix from the IPBSS method. As we can see, the convergence is significantly faster than what is possible with the standard algorithm. It is interesting to note also that there is more noise in the mutual information at convergence with respect to the standard algorithm. However, looking at the converging solution, there is once again not difference with the standard algorithm. Qualitatively from listening to the solutions, the authors could not hear any difference. We have therefore concluded that the difference is negligible and might have been caused by the noise in the mutual information at convergence.

The IPBSS algorithm turned out to be slightly faster for these examples than the PBSS one. From Figure 4 we can see from the value of the mutual information of the proposed algorithm using IPBSS from the first epoch, that the same value is reached for the standard algorithm between epoch 4 and 5, while for PBSS the value of the mutual

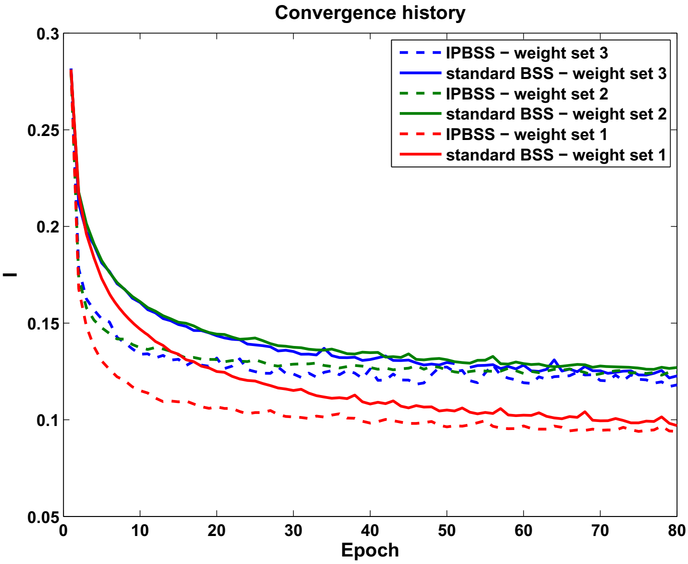

Fig. 4. Comparison of the standard algorithm to the algorithm with the **G** matrix from the IPBSS method

information at the first iteration is reached between epoch 2 and 3. A similar reasoning can be made for other epochs of iterations.

As concluded in the previous section, that indicates that the proposed algorithm using the PBSS method increases the speed of convergence of the first iterations of 3-4 times. From this, we can also infer that the IPBSS method further increases the speed of convergence, suggesting an improvement of 20-25% over the PBSS algorithm.

5 Conclusions

In this paper some preliminary results on a proportionate and improved proportionate version of the well-known time-domain Torkkola BSS algorithm have been proposed. The proposed idea is to use the sparseness of the acoustic impulse responses of the mixing environment, by applying some proportionate algorithm known in the context of speech adaptive filtering. Some experimental results have shown the effectiveness of the proposed approach in terms of convergence speed and encourage us for a more theoretical introduction of these novel classes of separation algorithms.

Acknowledgment. The authors would like to thank Mr. Gaute Halvorsen for his precious help in performing some of the presented experimental results.

References

1. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing. John Wiley (2002)
2. Makino, S., Lee, T.W., Sawada, H.: Blind Speech Separation. Springer (2007)
3. Choi, S., Cichocki, A., Park, H.M., Lee, S.Y.: Blind source separation and independent component analysis: a review. Neural Information Processing - Letters and Reviews 6(1), 1–57 (2005)
4. Araki, S., Mukai, R., Makino, S., Nishikawa, T., Saruwatari, H.: The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. IEEE Transactions on Speech and Audio Processing 11(2), 109–116 (2003)
5. Duttweiler, D.L.: Proportionate normalized least-mean-square adaptation in echo cancelers. IEEE Transactions on Speech and Audio Processing 8, 508–518 (2000)
6. Huang, Y., Benesty, J., Chen, J.: Acoustic MIMO Signal Processing. Springer (2006)
7. Benesty, J., Gay, S.L.: An improved PNLMS algorithm. In: Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002), pp. 1881–1884 (2002)
8. Torkkola, K.: Blind separation of convolved sources based on information maximization. In: Proc. of the 1996 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing, September 4-6, pp. 423–432 (1996)
9. Torkkola, K.: Blind deconvolution, information maximization and recursive filters. In: Proc. of 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1997), April 21-24, pp. 3301–3304 (1997)
10. Torkkola, K.: Blind separation of delayed sources based on information maximization. In: Proc. of 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1996), May 7-10, pp. 3509–3512 (1996)
11. Haykin, S.: Adaptive Filter Theory, 4th edn. Prentice-Hall (2001)
12. Ouedraogo, W.S.B., Jaidane, M., Souloumiac, A., Jutten, C.: Regularized gradient algorithm for non-negative independent component analysis. In: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011), Prague, Czech Republic, May 22-27, pp. 2524–2527 (2011)
13. Boulmezaoud, T.Z., El Rhabi, M., Fenniri, H., Moreau, E.: On convolutive blind source separation in a noisy context and a total variation regularization. In: Proc. of IEEE Eleventh International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2010), Marrakech, June 20-23, pp. 1–5 (2010)
14. Masulli, F., Valentini, G.: Mutual information methods for evaluating dependence among outputs in learning machines. Technical Report TR-01-02, Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova (2001)
15. Torkkola, K.: Learning feature transforms is an easier problem than feature selection. In: Proc. of 16th International Conference on Pattern Recognition, August 11-15, pp. 104–107 (2002)