A RECURSIVE LEAST-SQUARES EXTENSION OF THE NATURAL GRADIENT ALGORITHM FOR BLIND SIGNAL SEPARATION OF AUDIO MIXTURES

M. Elsabrouty, T. Aboulnasr and M. Bouchard
School of Information Technology and Engineering, University of Ottawa, 800 King Edward,
Ottawa, Ontario, K1N 6N5 {melsabro, aboulnas, bouchard}@site.uottawa.ca

1. INTRODUCTION

Blind Signal Separation (BSS) refers to the operation of recovering a set of sources that are as independent as possible from another set of observed linear or non-linear mixtures. The term blind indicates that this operation is done without knowing the mixing coefficients. Although this problem is more difficult than the classical filtering problem, a solution is still feasible provided that some information about the original independent components is provided. Major attention in the literature has been focused on different criteria to be used as objective functions for BSS. However, most of the existing on-line methods can be categorized as using a stochastic gradient. The need for second-order based algorithms for BSS can be easily revealed, as the fast convergence rate of the Recursive-Least-Squares (RLS) based algorithms can be advantageous for many applications. Among the many existing objective functions for Blind Signal Separation, Maximum Likelihood and Negentropy stand as strong criteria which are well justified, as they minimize the mutual information between the original independent components [1],[2]. Maximum likelihood proved especially suitable for heavy tailed distribution signals, such as audio data [3].

In the case of pre-whitened inputs, the separating de-mixing matrix is constrained to be orthogonal, lying on a Stiefel manifold [4]. It is widely recognized that applying the constraints of the Stiefel manifold to an optimization problem leads to a better performance of the algorithm. Following the Stiefel geometry leads to a modified gradient rather than the ordinary gradient of the objective function [4]. The new constrained gradient is referred to as the natural gradient, for the pre-whitened case [4],[5]. A natural question would be how to extend and combine the performance of the natural gradient with second-order based RLS algorithm. It is the objective of this paper.

2. EXISTING MAXIMUM-LIKELIHOOD BASED ALGORITHMS

Let $s_i(n), i=1,2,\dots,N$ be scalar inputs (or sources) to the blind signal separation model at a time n. For simplicity, it is assumed that the mixing is linear and that the mixing matrix is square, i.e. the number of inputs N is equal to the number of mixtures $x_i(n), i=1,2,\dots,N$. Therefore, the mixing matrix A is a square matrix of size $N\times N$. The mixing model can be expressed as:

$$\mathbf{x}(n)=\mathbf{A}\times\mathbf{s}(n) \qquad (1).$$

The mixture $\mathbf{x}$ is then applied to a whitening matrix $\mathbf{V}$. The resulting whitened mixtures in $\mathbf{z}$ are expressed as:

$$\mathbf{z}(n)=\mathbf{V}\times\mathbf{x}(n)=\mathbf{V}\times\mathbf{A}\times\mathbf{s}(n)=\mathbf{B}\times\mathbf{s}(n) \qquad (2),$$

where $\mathbf{B}$ is the resulting mixing matrix after the whitening stage. The purpose of the blind signal separation algorithms is to estimate a matrix $\mathbf{W}$ such that $\mathbf{W}\times\mathbf{B}=\mathbf{I}_{N\times N}$, where $\mathbf{I}$ is an identity matrix. Then the outputs of the separation process referred to as $y_i(n)$ would be identical to the source inputs $s_i(n)$. Maximum Likelihood targets a separation via increasing the likelihood between the outputs $y_i(n)$ and the inputs $s_i(n)$ [5]. In the case of pre-whitened inputs, the cost function of the log-likelihood $L(\mathbf{W})$ of the de-mixing matrix $\mathbf{W}$ can be expressed as:

$$L(\mathbf{W})=E\left\{\sum_{i=1}^{N}\log p_i(\mathbf{w}_i\mathbf{z})\right\} \qquad (3),$$

where $E\{\ \}$ refers to the expected value, $\mathbf{w}_i$ is the i^{th} row of the matrix $\mathbf{W}$ and $p_i(\ )$ is a probability density function. The above cost function has the gradient $\nabla L(\mathbf{W})$ as:

$$\nabla L(\mathbf{W})=E\left\{\sum_{i=1}^{N}\log \mathrm{g}(\mathbf{y})\mathbf{z}^T\right\} \qquad (4),$$

where $g(y_i)=\frac{p_i'}{p_i}$ and is usually set to 2 $\tanh(y_i)$ for super-gaussian data, such as audio data. Pre-whitening also constrains the matrix $\mathbf{W}$ to be orthogonal, meaning that $\mathbf{W}\,\mathbf{W}^T=\mathbf{I}_{N\times N}$. This constraint places the optimization of the cost function on a Stiefel manifold, where the knowledge of the differential geometry can be used to adjust the original gradient $\nabla L(\mathbf{W})$ to the natural gradient $\tilde{\nabla} L(\mathbf{W})$ which follows the geometry of the manifold [6],[7]:

$$\begin{aligned}\tilde{\nabla}L(\mathbf{W})&=\nabla L(\mathbf{W})-\mathbf{W}\nabla L(\mathbf{W})\ \mathbf{W}\\&=(\mathrm{g}(\mathbf{y})\mathbf{y}^T-\mathbf{y}\,\mathrm{g}(\mathbf{y})^T)\ \mathbf{W}\end{aligned} \qquad (5).$$

The natural gradient algorithm is based on taking the instantaneous value of the update for the above gradient, so that the update equation becomes:

$$\mathbf{W}_{new} = \mathbf{W}_{old} + \mu\,(\mathbf{y}\,\mathrm{g}(\mathbf{y})^T - \mathrm{g}(\mathbf{y})\,\mathbf{y}^T)\;\mathbf{W}_{old} \quad (6),$$

where μ is the step size adequately set to 0.0005.

3. THE QUASI-RLS STIEFEL ALGORITHM

A time averaged approximation of the Hessian of the cost function is (except for a scaling factor of $1-\lambda$):

$$\mathbf{R}_{new} = \lambda \times \mathbf{R}_{old} + \tilde{\nabla} L(\mathbf{W})_{vec} \times \tilde{\nabla} L(\mathbf{W})^T{}_{vec}\,\} \quad (7),$$

where $\tilde{\nabla} L(\mathbf{W})_{vec}$ is a re-arrangement of the $N \times N$ matrix $\tilde{\nabla} L(\mathbf{W})$ as a vector of size $N \times 1$, via column stacking. The resulting $\mathbf{R}$ is thus a matrix of size $N^2 \times N^2$. λ is a forgetting factor close to 1, common in RLS algorithms. The update can thus be calculated as:

$$\Delta \mathbf{W}_{vec} = \mathbf{R}^{-1} \times \tilde{\nabla} L(\mathbf{W})_{vec} \quad (8).$$

The inverse of the quasi-Hessian matrix $\mathbf{R}$, which is required for the above update, can be calculated recursively using the matrix inversion lemma as:

$$\mathbf{R}^{-1}{}_{new} = \frac{1}{\lambda}\left[\mathbf{R}^{-1}{}_{old} - \frac{\mathbf{R}^{-1}{}_{old}\,\tilde{\nabla} L(\mathbf{W}) \tilde{\nabla} L(\mathbf{W})^T\,\mathbf{R}^{-1}{}_{old}}{\lambda + \tilde{\nabla} L(\mathbf{W})^T\,\mathbf{R}^{-1}{}_{old} \tilde{\nabla} L(\mathbf{W})}\right] (9).$$

To enhance the robustness of the algorithm, the update calculated in (8) is projected on the Stiefel manifold as:

$$\tilde{\Delta} \mathbf{W}_{vec} = \mathbf{W}\mathbf{W}^T \Delta \mathbf{W}_{vec} - \mathbf{W} \Delta \mathbf{W}_{vec} \mathbf{W} \quad (10).$$

The new update is rearranged into a matrix $\tilde{\Delta} \mathbf{W}$ of the original size $N \times N$ to be added to the current estimate of the matrix $\mathbf{W}$:

$$\mathbf{W}_{new} = \mathbf{W}_{old} + \mu_{RLS}\;\tilde{\Delta} \mathbf{W} \quad (11).$$

4. SIMULATION RESULTS AND PERFORMANCE COMPARISON

To compare the different BSS algorithms, tests were performed on a mixture of audio data files (speech) sampled at 8 kHz and of duration 3.7 sec. There were 4 sources considered, two male files and two female files. The choice of the speech files duration was made short to emphasize on the fast convergence property of the new algorithm. The mixing matrix chosen for this aural scene is rather harsh. The comparison of the proposed algorithm is performed with the Natural Gradient based on Maximum Likelihood (NAG, step size μ set to 0.0005 [8]), and with the RLS-modified Natural Gradient (RLS-NAG). The RLS-NAG proposed in [8] suggests an RLS-update of the Natural Gradient algorithm by modifying the update from

$$\Delta \mathbf{W} = \mu\left[\mathbf{I} - \mathrm{g}(\mathbf{y})\,\mathbf{y}^T\right]\mathbf{W} \quad (12)$$

to

$$\Delta \mathbf{W} = \mu_r\left[\mathbf{I} - \mathrm{g}(\mathbf{y})\,\mathbf{y}^T\right]\mathbf{Q}^{-1}\mathbf{W} \quad (13)$$

with $\mathbf{Q} = \mathrm{g}(\mathbf{y})\,\mathbf{y}^T$ and has the form of a covariance matrix. This update is applied on each element of $\mathbf{W}$ individually and convergence is obtained when $\mathbf{Q}$ is a diagonal matrix, i.e. when the mutual information between $\mathrm{g}(\mathbf{y})$ and $\mathbf{y}$ is minimized. This algorithm works efficiently and provides in most cases a better performance than the ordinary Natural Gradient algorithm [8]. For the RLS-NAG algorithm μ_r is set to 0.008 while the forgetting factor is set to 0.991. For the new proposed algorithm, the step size μ_{RLS} is set to 0.12 and the forgetting factor λ is set to be time varying starting at 0.9993 at time n=1 and ending at 0.9996 at n=10000. To evaluate the effectiveness of the separation algorithms, the PESQ scores (from ITU-T P.862 [9]) of the separated outputs $y_i(n)$ were computed. PESQ scores have values varying between -0.5 to 4.5, and higher values indicate a higher speech quality. The results of the above test are provided in Table 1. From this table, it can be seen that the proposed algorithm converges much faster than the other algorithms, as shown by the significant difference in the PESQ scores achieved by the different algorithms.

5. CONCLUSION

This paper presented a new algorithm named the Quasi-RLS Stiefel. This algorithm combines the principles of natural-gradient on differential manifolds (Stiefel manifold in our case) and RLS-based algorithms. Simulation results quantified the good on-line convergence speed of the proposed algorithm and proved that the algorithm is very suitable for real-time Blind Signal Separation.

Table 1. PESQ scores for a mixture of 4 speech files

File	New algorithm	NAG	NAG-RLS
Female1	3.244	2.311	1.293
Female2	3.310	1.732	1.641
Male1	3.643	1.807	2.288
Male2	3.201	1.753	1.475

REFERENCES

[1] Amari, S. (1998). Natural gradient works efficiently in learning. Neural Computation, vol. 10, 251-276.

[2] Hyvarinen, A. , Karhunen, J. and Oja, E. (2001). In: Independent Component Analysis, John Wiley & Sons.

[3] Kidmose, P. (2001). Blind Separation of Heavy Tail Signals. Ph.D. thesis, Technical University of Denmark.

[4] Amari, S. (1999) Natural gradient learning for over- and under-complete bases in ICA. Neural Comput., vol.11, 1875–1883.

[5] Cichocki, A. and Amari, S. (2002) Adaptive Blind and Image Processing, John Wiley & Sons.

[6] Douglas, S.C. (2000). Self-stabilized gradient algorithms for blind source separation with orthogonality constraints. IEEE Trans. Neural Networks, vol. 11, 1490–1497.

[7] Edelman, A., Arias, T. and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl., vol. 20, 303–353.

[8] Benesty, J. (2000). An Introduction to Blind Source Separation of Speech Signals. In: Acoustic Signal Processing for Telecommunication, Kluwer Academic Publisher, 13-61.

[9] ITU-T P.862 (2000). Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone network and speech codecs. International Telecommunication Union.