SPSA FOR NOISY NON-STATIONARY BLIND SOURCE SEPARATION

Tariq Durrani[1] *Gordon Morison*[1]

gordon@spd.eee.strath.ac.uk durrani@spd.eee.strath.ac.uk
[1]Department of Electronic and Electrical Engineering
University of Strathclyde, Glasgow G1 1XW UK

ABSTRACT

In this paper a novel application of the simultaneous perturbation stochastic approximation algorithm (SPSA) to the noisy non-stationary blind source separation problem is presented and described. The proposed approach demonstrates the algorithm with a second order cost function suitable for applications to non-stationary data. Some extensions to the algorithm that are currently being investigated are also described in the paper, and the algorithm performance is demonstrated via simulation.

1. INTRODUCTION

Blind source separation deals with the problem of separating a number of mixed statistically independent signals impinging on an array of sensors. The term blind arises from the fact that no information relating to the mixing process is available.

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \upsilon(t) \qquad (1)$$

where $\boldsymbol{x}(t) = (\boldsymbol{x}_1(t)....\boldsymbol{x}_n(t))$ represents an n-dimensional vector of measurement signals, $\boldsymbol{A}$ is the $(n \times m)$ unknown mixing matrix, $\boldsymbol{s}(t) = (\boldsymbol{s}_1(t)....\boldsymbol{s}_n(t))$ represents the m-dimensional unknown source signal vector and $\upsilon(t) = (\upsilon_1(t)...\upsilon_n(t))$ is the sensor noise with $\boldsymbol{t}$ representing the time index for all quantities.
The objective of the process is using only the information available at the sensor outputs to perform a linear transformation on the data in order that the sensor outputs are as independent as possible. This may be described as follows:

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{x}(t) \qquad (2)$$

where $\boldsymbol{y}(t) = (\boldsymbol{y}_1(t)....\boldsymbol{y}_n(t))^{\mathrm{T}}$ represents the n-dimensional vector of input source signal estimates and $\boldsymbol{W}$ is the $(n \times m)$ demixing matrix.

As is standard with the source separation problem the sources may only be estimated up to a scale and permutation factor, the permutation ambiguity arises dues to the fact that as no assumptions are made on the channel or the sources then ordering itself is arbitrary, the scaling ambiguity arises as the exchange of a fixed scalar between the signal and the corresponding row in the mixing matrix has no effect on the observed sensor signal output.

$$\boldsymbol{x}(t) = \sum_{i=1}^{n} \frac{a_i}{\alpha_i} \alpha_i \boldsymbol{s}_i(t) + \upsilon(t) \qquad (3)$$

Throughout this paper the following assumptions are made.

(AS1) The mixing matrix $\boldsymbol{A}$ is full column rank.
(AS2) The sources are spatially and temporally uncorrelated.
(AS3) The variances of the sources are time varying i.e. they are second order non-stationary signals.
(AS4) The sensor noise $\upsilon(t)$ is additive Gaussian white noise spatially uncorrelated with the sources.

2. SPSA ALGORITHM

The SPSA algorithm was developed by Spall [1] and since its development has been used for a number of differing fields e.g. neural network training [5], traffic management [6] and recently to the blind source separation problem [2].

The goal of the algorithm is to minimize a scalar valued cost function J($\boldsymbol{W}$), assuming only that noisy measurements of the cost function are available, and that the cost function is differentiable. Differing from standard Robbins Monro SA algorithms that require either explicit calculation or measurement of the gradient of the cost function, the SPSA algorithm utilizes only two measurements of the cost function and a random simultaneous perturbation to estimate the gradient of the cost function. This is calculated as follows:

$$\Delta \boldsymbol{W}_{ij} = \frac{\boldsymbol{J}(\boldsymbol{W}_{ij} + c\xi_{ij}) - \boldsymbol{J}(\boldsymbol{W}_{ij})}{c\xi_{ij}} \quad (4)$$

where c a small perturbation value, and ξ_{ij} is the element in the i[th] row and j[th] column of a sign matrix with elements taking a value +1 or –1 generated from a Bernoulli distribution. The matrix elements ξ_{ij} have the following properties.

$$E\{\xi_{ij}\} = 0 \quad (5)$$

Taking expectations of the Taylor series expansion of equation (4)

$$\Delta \boldsymbol{W}_{ij} = \xi^T \frac{dJ(\boldsymbol{W})}{d\boldsymbol{W}} \xi_{ij} + \frac{c\xi^T}{2!} \frac{d^2 J(\boldsymbol{W})}{d\boldsymbol{W}^2} \xi\xi_{ij} \quad (6)$$

from equations (4), (5) and (6) we have

$$E\{\Delta \boldsymbol{W}_{ij}\} = \frac{dJ(\boldsymbol{W})}{d\boldsymbol{W}_{ij}} \quad (7)$$

Thus to obtain more accurate estimates of the required gradient a number of estimates of the gradient are made and the expectation of the gradient estimates is used in a standard stochastic approximation gradient update equation. The sample expectation is taken over N values of the gradient estimates.

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k - a_k \left\{ N^{-1} \sum_{n=1}^{N} (\Delta \hat{\boldsymbol{W}}_k)_n \right\} \quad (8)$$

a_k represents the step size parameter for the system. To improve system convergence the step size is exponentially decayed with increasing data points as follows:

$$a_k = \frac{a}{(k + D)^\alpha} \quad (9)$$

with a and D chosen as initially as small values.

3. SECOND ORDER BSS ALGORITHMS

Over the last 15 years there has been an extensive interest into the topic of blind source separation. Commonly it is assumed that the signals are independent and identically distributed (i.i.d) and the separation is performed using either implicitly [10] or explicitly [11] higher order statistics of the sensor output signals. This is known as Independent Component Analysis (ICA) [11]. When the source signal data represents that of a time series then the signals can be separated using second order statistics.
For the problem posed in this paper the above assumptions imply

$$\boldsymbol{R}_{ss} = E\{\boldsymbol{s}_k \boldsymbol{s}_k^T\} = diag\{\sigma_{s1}^2 .... \sigma_{sn}^2\} \quad (10)$$

$$\boldsymbol{R}_{\upsilon\upsilon} = E\{\upsilon_k \upsilon_k^T\} = diag\{\sigma_{\upsilon 1}^2 .... \sigma_{\upsilon n}^2\} \quad (11)$$

using these results and without loss of generality assuming the variance of the source signals is unity and the noise variances are identical equal to σ_υ^2. The covariance matrix of the sensor outputs is

$$\boldsymbol{R}_{xx} = E\{\boldsymbol{x}_k \boldsymbol{x}_k^T\} = \boldsymbol{A}\boldsymbol{A}^T + \sigma_\upsilon^2 \boldsymbol{I} \quad (12)$$

Using one of the standard second order blind source separation algorithms (e.g. SOBI [3], TDSEP [4], AMUSE [7]) will result in biased estimates of the demixing matrix. To avoid this the noise covariance matrix must be included in the algorithm [8]. A number of differing criterion are desirable when designing the separation algorithm e.g. output source decorrelation, output source power control etc. Recently in [9] a linear combination of elementary cost functions was defined.

$$J(\boldsymbol{W}, \boldsymbol{R}_{nn}) = \sum_n \alpha_n J_i(\boldsymbol{W}, \boldsymbol{R}_{nn}) \quad (13)$$

4. BSS USING SPSA

The cost function used for the signal separation in this paper was as follows [9]:

$$\boldsymbol{J}_1 = \left\| off\left(\boldsymbol{W}(\boldsymbol{R}_{xx} - \boldsymbol{R}_{nn})\boldsymbol{W}^H\right)\right\|_F^2 \quad (14)$$

The operator *off* in the above equation represents the off-diagonal terms of a matrix. This function gives a measure of the signal decorrelation in the noise free case. This makes the implicit assumption that the variance of the noise is known a priori or that the variance can be estimated from the data.
Yet the problem with equation (14) is that the trivial solution $W = 0$ also minimizes this function. Therefore in accordance with [9] a constraint is also placed upon the separating matrix W to avoid the trivial solution. The constraint is as follows.

$$\boldsymbol{J}_2 = \left\| ddiag(\boldsymbol{W}\boldsymbol{W}^H - I) \right\|_F^2 \qquad (15)$$

Where $ddiag(\boldsymbol{W}) = \boldsymbol{W} - diag(\boldsymbol{W})$.

Thus the overall cost function used is a linear combination of $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$.

$$\boldsymbol{J}_T(\boldsymbol{W}, \boldsymbol{R}_{nn}) = \boldsymbol{J}_1 + \alpha \boldsymbol{J}_2 \qquad (16)$$

5. SIMULATIONS

In the following an example illustrating the described algorithm is demonstrated. It is known that the closer the global system matrix $\boldsymbol{G} = \boldsymbol{WA}$ is to a permutation matrix the better the separation. During simulations the mixing matrix is known and is used to assess the performance of the system. The performance measure used is the average interference [4]:

$$SIR(G) = \frac{1}{n}\sum_{i-1}^{n}\left(\frac{1}{n-1}\sum_{j=1}^{n-1}\frac{|g_{ij}|}{\max_k |g_{ik}|} - 1 \right) \qquad (17)$$

Where g_{ij} represents the element in the ith row and jth column of a $(n \times n)$ matrix $\boldsymbol{G}$. The input sources for the system were 3 speech signals of 5000 samples duration, sampled at 8kHz. The mixing and demixing matrices were chosen randomly from a uniform distribution. The variance of the noise was $\sigma^2 = 0.1$, a_k was set initially to 0.1 and c_k was given an initial value of 0.05. The gradient calculations were averaged over 10 iterations.

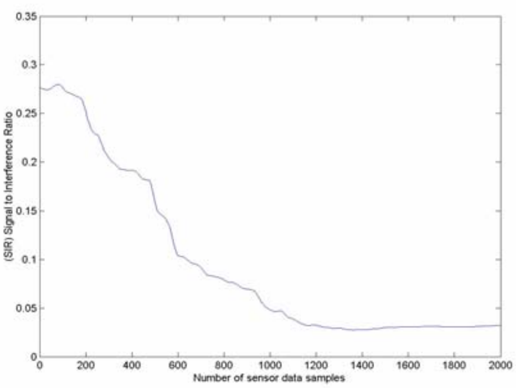

Fig 1. Learning curve showing the Signal to Interference Ratio (SIR)

It can be seen from the above diagram that the algorithm had converged after approximately 1200 iterations.

6. EXTENSIONS

A number of new algorithms have been investigated. These include a projection scaling algorithm, the use of simulated annealing, and a linear constraint algorithm. The projection algorithm drives the iterative solution towards a collinear solution along the gradient direction, and can be expressed as:

$$\boldsymbol{W}_{k+1} = P_k(\boldsymbol{W}_k + a_k \Delta J(\boldsymbol{W}_k)) + \mathrm{b}\boldsymbol{I} \qquad (18)$$

where

$$\boldsymbol{P}_{k+1} = \boldsymbol{I} - \boldsymbol{W}_k(\boldsymbol{W}_k \boldsymbol{W}_k^T)^{-1}\boldsymbol{W}_k \qquad (19)$$

Given the convergence value of

$$\boldsymbol{P}_{k+1} \equiv \boldsymbol{P}_1 \equiv \boldsymbol{P}_2 \equiv .... \equiv \boldsymbol{P}_k \qquad (20)$$

and b is a regularization parameter that ensures the avoidance of the trivial solution $W = 0$ while striving towards the minimal value of the cost function.

In the simulated annealing algorithm the perturbation values and the associated scaling matrix are chosen from a Boltzman distribution based on gradient direction. This will be presented at the conference.

In the linear constraint algorithm the cost function is augmented by a linear (planar) constraint, to drive the solution towards a constraint optimum away from the trivial solution.

7. CONCLUSION

In this paper the SPSA algorithm has been utilised for the optimization of a second order cost function with application to the blind separation of non-stationary sources. Numerical simulations have been included demonstrating the operation of the algorithm for linearly mixed uncorrelated speech signals. The application of the algorithm to the offline separation of convolutively mixed signals is an area of our current research.

Acknowledgements: This work was funded in part by Epson Scotland Design Center.

8. REFERENCES

[1] J. Spall, 'Multivariate Stochastic Approximation using a Simultaneous Perturbation Stochastic Gradient Approximation' IEEE Trans. on Automatic Control, Vol 37, No 3, March 1992.

[2] Y. Maeda, K. Tsushio, ‘Blind Signal Separation via Simultaneous Perturbation Method’ IJCNN 2002.

[3] A. Belouchrani, K Abed Meraim, J.-F Cardoso and E. Moulines, ‘A blind source separation technique based on second order statistics’ IEEE Trans. on Signal Processing, Vol 45, No 2, February 1997.

[4] A. Ziehe, K.-R Muller, ‘TDSEP – and efficient algorithm for blind separation using time structure’, Proc Int. Conf on Artificial Neural Networks (ICANN’98), Skovde, Sweden 1998.

[5] Y. Maeda, H. Hirano, Y. Kanata, ‘A learning rule of Neural Networks via Simultaneous Perturbation and its hardware implementation’, Neural Networks 251-259 1995.

[6] J. Spall, D.C Chin, ‘Traffic-Responsive Signal Timing for System Wide Traffic Control’, Transp. Res Part C 5, 153-163 1997.

[7] L. Tong, R.-W. Liu, V.C. Soon, Y.-F. Huang, ‘Indeterminacy and identifiability of blind identification’ IEEE Trans. of Circuits and Systems 38:499-509, 1991 .

[8] S.C. Douglas, A. Cichocki, S. Amari, ‘Bias removal technique for blind source separation with noisy measurements’ Electronic Lett. Vol 34. No 14 July 1998.

[9] M. Joho, R.H. Lambert, H. Mathias, ‘Elementary cost function for blind separation of non-stationary source signals’, ICASSP 2001, Salt Lake City, UT, USA, Vol 5, pp 2793-2796 2001.

[10] S. Amari, A. Cichocki, H.H. Yang, ‘A new learning algorithm for blind source separation’, In Advances in Neural Information Processing Systems 8, pages 757-763, MIT Press, 1996.

[11] P. Comon ‘Independent Component Analysis – a new concept?’ Signal Processing 36:287-314, 1994.