A COMBINED KALMAN FILTER AND NATURAL GRADIENT ALGORITHM APPROACH FOR BLIND SEPARATION OF BINARY DISTRIBUTED SOURCES IN TIME-VARYING CHANNELS

*M.G. Jafari, H.W. Seah, and J.A. Chambers**

Dept. of Electrical & Electronic Engineering, Imperial College, London, SW7 2BT, U.K.
Dept. of Electonic & Electrical Engineering, University of Bath, Bath, BA2 7AY, U.K.*
E-mail: m.jafari@ic.ac.uk

ABSTRACT

A combined Kalman filter (KF) and natural gradient algorithm (NGA) approach is proposed to address the problem of blind source separation (BSS) in time-varying environments, in particular for binary distributed signals. In situations where the mixing channel is non-stationary, the performance of NGA is often poor. Typically, in such cases, an adaptive learning rate is used to help NGA track the changes in the environment. The Kalman filter, on the other hand, is the optimal minimum mean square error method for tracking certain non-stationarity. Experimental results are presented, and suggest that the combined approach performs significantly better than NGA in the presence of both continuous and abrupt non-stationarities.

1. INTRODUCTION

Blind source separation has recently received much research attention, due to its wide range of potential applications, which include wireless communications, geophysical exploration, speech and image processing, and medical signal processing [1, 2, 3]. It is concerned with recovering the original source signals (sources), given only the observed signals (sensors), which arise when the sources are mixed by an unknown medium. When dealing with signals recorded in a real environment, BSS is complicated by additive noise, propagation delays, time-varying environments, and non-stationary sources. To make the problem more tractable, it is common practice to assume that stationary sources are instantaneously mixed by a constant environment, and that the mixtures are not corrupted by noise. In this paper, the problem of BSS in non-stationary environments, when the sources have binary distributions, is addressed. To this end, we propose to combine the Kalman filter with the NG on-line algorithm, proposed by Amari et al. [4]. The main advantage of NGA is that it has *equivariant* property, which implies that its asymptotic convergence properties are independent of the condition number of the mixing matrix and scaling factors of the source signals [1, 4]. Nevertheless, most of the research effort devoted to extending NGA to the time-varying case has been oriented toward the formulation of adaptive algorithms that update the step-size parameter, so-called learning of the learning rate. Although an adaptive learning rate does give better performance in non-stationary situations, simulation results show that the tracking performance of NGA remains limited in such cases. The Kalman filter, on the other hand, is the optimal filter for tracking certain non-stationarity, provided the dynamics of the environment it attempts to track can be modelled by the state evolution equation [5, 6]. The BSS problem, for the time-varying channel case, is introduced in section 2. The combined approach is explained in section 3, together with a brief description of the KF and NGA techniques. The performance of the proposed approach is shown by simulation in section 4, while conclusions are drawn in section 5.

This work was supported by the Engineering and Physical Sciences Research Council of the U.K.

2. PROBLEM STATEMENT

When n real sources are mixed by an instantaneous non-stationary channel, and no noise is present, the m observed signals are given by [1]

$$\mathbf{x}(k) = \mathbf{A}(k)\mathbf{s}(k) \tag{1}$$

where $\mathbf{x}(k) = [x_1(k), \ldots, x_m(k)]^T$ is the m-dimensional vector of observed signals, $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ is the vector of source signals which are assumed to be zero-mean and mutually independent, and $[.]^T$ denotes vector transpose. $\mathbf{A}(k)$ is an unknown, full column rank, mxn mixing matrix, and typically it is assumed that there are at least as many sensors as sources, that is $m \geq n$. The sources are recovered using the following linear separating system

$$\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k) \tag{2}$$

where $\mathbf{y}(k) = [y_1(k), \ldots, y_n(k)]^T$ is an estimate of $\mathbf{s}(k)$, and $\mathbf{W}(k)$ is the nxm separating matrix. However, it is only possible to recover the sources up to a multiplicative constant, and their order cannot be predetermined. These ambiguities are inherent to the BSS problem, and imply that the exact inverse of the mixing matrix cannot be obtained, so that perfect separation is achieved when the global mixing-separating matrix, defined as

$$\mathbf{P}(k) = \mathbf{W}(k)\mathbf{A}(k) \tag{3}$$

tends toward a matrix with only one non-zero term in each row and column [1], and is given by

$$\mathbf{P}(k) = \mathbf{J}\mathbf{D} \tag{4}$$

where $\mathbf{J}$ is an nxn permutation matrix which models the ambiguity relating to the ordering of the sources, and $\mathbf{D}$ is an nxn diagonal matrix which accounts for the indeterminacy of scaling.
In addition to the statistical independence of the sources and the non-singularity of the mixing matrix, BSS also assumes that at most one source has Gaussian distribution because, for Gaussian random variables, uncorrelatedness corresponds to independence [7]. In this paper we also assume that the sources have binary distribution. This assumption allows us to introduce additional a priori information about the source signals, which will help the tracking of the Kalman filter.

3. COMBINED KF AND NGA APPROACH

The proposed approach uses NGA as the basic BSS block. This updates adaptively the separating matrix, thus estimating the source signals. Algorithm tracking ability is provided by the KF technique, which uses the recovered sources and the observed signals, to estimate the mixing matrix.

3.1. Kalman Filter

Using a similar method as in [5], the following cost function is minimised

$$J_{KF} = E\{||\mathbf{h}(k) - \hat{\mathbf{h}}_K(k)||_2^2 | \mathbf{x}(k)\} \tag{5}$$

where $\hat{\mathbf{h}}_K(k)$ represents the estimate of the vector $\mathbf{h}(k)$, $\mathbf{x}(k)$ is the observation vector, and $||.||_2$ denotes the 2-norm. This leads to the following expressions, describing KF [6]

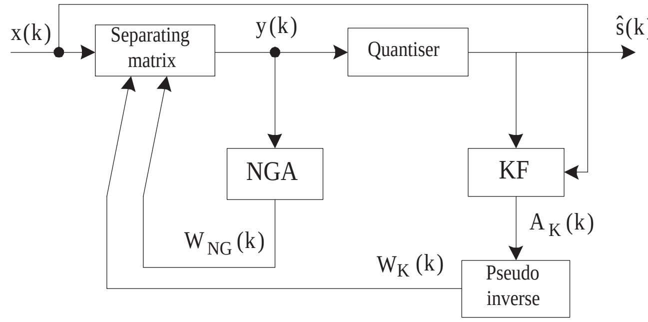

Figure 1: Structure of combined KF and NGA approach.

$$\begin{aligned}
\mathbf{h}_K^p(k) &= \mathbf{T}\mathbf{h}_K^c(k-1) && (6)\\
\mathbf{M}(k) &= \mathbf{T}\mathbf{M}(k-1)\mathbf{T}^T + \mathbf{Q} && (7)\\
\mathbf{K}(k) &= \mathbf{M}(k)\mathbf{H}^T(k)(\mathbf{C}(k) + \mathbf{H}(k)\mathbf{M}(k)\hat{\mathbf{H}}^T(k))^{-1} && (8)\\
\mathbf{h}_K^c(k) &= \mathbf{h}_K^p(k) + \mathbf{K}(k)(\mathbf{x}(k) - \mathbf{H}(k)\mathbf{h}_K^p) && (9)\\
\mathbf{M}(k) &= (\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k))\mathbf{M}(k) && (10)
\end{aligned}$$

where $\mathbf{h}_K^p(k)$ and $\mathbf{h}_K^c(k)$ denote respectively the predicted and corrected estimate of the vector $\mathbf{h}(k)$. $\mathbf{H}(k)$ and $\mathbf{T}$ are, respectively, the known observation matrix and state transition matrix, $\mathbf{Q}$ and $\mathbf{C}(k)$ are respectively the covariance matrices of the process noise, and of the measurement noise. The Kalman gain is the matrix $\mathbf{K}(k)$, $\mathbf{M}(k)$ represents the parameter error covariance matrix, and $\mathbf{I}$ is the identity matrix.

3.2. Natural Gradient Algorithm

The natural gradient algorithm update equation is given by the following expression [4, 1]

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(k)[\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)]\mathbf{W}(k) \tag{11}$$

where $\mathbf{I}_n$ is the identity matrix, $\mathbf{f}(\mathbf{y}(k))$ is an odd non-linear function of the output $\mathbf{y}(k)$, called the activation function, and $\eta(k)$ is a positive learning parameter. Usually the learning rate is assumed to be a very small positive constant which is either fixed or decreases exponentially to zero [8, 1]. However, when the algorithm is required to track a time-varying environment, neither approaches are suitable. In [8], the learning rate is self-adaptive, and changes according to a non-linear function of the mean values of the gradient components:

$$\begin{aligned}
\hat{\mathbf{g}}(k) &= (1-\rho_2)\hat{\mathbf{g}}(k-1) + \rho_2\tilde{\mathbf{g}}(k) && (12)\\
\eta(k) &= (1-\rho_1)\eta(k-1) + \rho_1\beta\phi(||\hat{\mathbf{g}}(k)||) && (13)
\end{aligned}$$

where $0 < \rho_1 < 1$, $0 < \rho_2 < 1$, and $\beta > 0$ are fixed coefficients, $\tilde{\mathbf{g}}(k) = -(\mathbf{I}_n - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k))\mathbf{W}(k)$, is the gradient at time k, and $\phi(||\hat{\mathbf{g}}(k)||)$ is a non-linear function defined in [8] as $\phi(||\hat{\mathbf{g}}(k)||) = (1/m)\sum_{i=1}^{m} |\hat{g}_i(k)|$ or $\phi(||\hat{\mathbf{g}}(k)||) = tanh((1/n)\sum_{i=1}^{n} \hat{g}_i^2(k))$, which is introduced to limit the maximum value of the gradient.

3.3. The Combined Approach

The Kalman filter, used to estimate the mixing matrix coefficients, requires the knowledge of a vector representing the desired response, and an observation matrix. Thus, the vector of sensor measurements $\mathbf{x}(k)$ is taken as the desired response of the filter and, in the absence of a known observation matrix, the source estimates generated by NGA, are quantised to ± 1, and fed to KF, as shown in Fig. 1. The use of a quantiser implies that we take advantage of *a priori* knowledge about the sources, which is found to improve significantly the performance of the Kalman filter. For our implementation, we re-arrange the mixing matrix into an mn-dimensional column vector, defined as

$$\mathbf{h}(k) = vec(\mathbf{A}^T(k)) \tag{14}$$

Hence, the combined approach can be formulated as

$$\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k) \tag{15}$$

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(k)(\mathbf{I} - \mathbf{f}(k)\mathbf{y}^T(k))\mathbf{W}(k) \tag{16}$$

where $\mathbf{W}(k)$ represents the separating matrix estimated with NGA, and $\eta(k)$ is the self-adaptive learning rate defined in (12)-(13). The mixing coefficients vector is then estimated by KF, which takes the form in (6)-(10), with the observation matrix $\mathbf{H}(k)$ in (8)-(10) replaced by the mxmn matrix $\hat{\mathbf{S}}(k)$, defined as

$$\hat{\mathbf{S}}(k) = \mathbf{I}_m \otimes \hat{\mathbf{s}}^T(k) \tag{17}$$

where $\mathbf{I}_m$ is the m-dimensional identity matrix, and $\otimes$ denotes the Kronecker product. $\hat{\mathbf{s}}(k)$ is given by

$$\hat{\mathbf{s}}(k) = \mathbf{g}(\mathbf{y}(k)) \tag{18}$$

where $\mathbf{y}(k)$ is the nx1 source signal vector estimated by NGA, and quantised by function $\mathbf{g}(.)$. The estimated mixing coefficient vector is subsequently re-arranged into the mxn matrix $\mathbf{A}_K(k)$, and its pseudo inverse, defined as

$$\mathbf{A}_K^{\dagger}(k) = \mathbf{A}_K^T(k)(\mathbf{A}_K(k)\mathbf{A}_K^T(k))^{-1} \tag{19}$$

generates an additional separating matrix, $\mathbf{W}_K(k)$, that updates periodically, every T_p samples, the NGA estimate, i.e. if $k \; mod \; T_p = 0$, $\mathbf{W}(k+1) = \mathbf{W}_K(k)$; else $\mathbf{W}(k+1)$ is updated by (16).

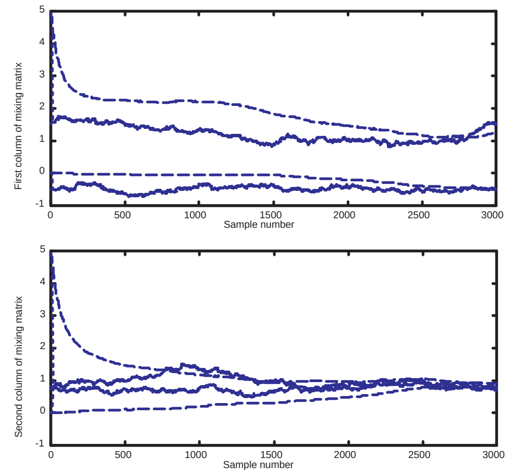

Figure 2: Tracking of the two columns of the mixing matrix, when the channel is non-stationary: actual coefficients (solid lines), estimated with NGA alone (dashed lines), and estimated with the combined approach (dotted lines, coincident with the solid lines).

4. SIMULATIONS

In this section we present computer simulations which demonstrate the performance of the combined approach. We consider two independent Bernoulli sources, mixed by a time-varying mixing channel, whose elements change according to independent first order Gauss-Markov models. We assume that the number of sensors equals the number of sources (n=m=2), and that no additive noise is present; the activation function in (11) is chosen as $\mathbf{f}(\mathbf{y}(k)) = [y_1^3(k), \ldots, y_n^3(k)]^T$, and the fixed parameters in (12)-(13), the learning rate update equations, are $\rho_1 = \rho_2 = 0.01$, and $\beta = 0.005$. The separating matrix generated by NGA, $\mathbf{W}(k)$, is initialised to $\mathbf{W}_K(k)$ every 10 samples, i.e. $T_p = 10$. Figs. 2 and 3 show, respectively, the tracking of the two columns of the mixing matrix, by NGA only and by the combined approach, in the case of continuous and abrupt non-stationarity. The results are illustrative of the indeterminacy of scaling, as both techniques track the negative of the second column of the mixing matrix, whose actual values have not been plotted in either figures for the sake of clarity. It should be noted that it is not possible to discriminate between the true mixing coefficients, and the estimates produced by the combined approach, as these are coincident on these plots. Thus, the combined approach is fast enough to track the changes in the channel that cannot be followed by NGA alone. The performance of a BSS method can also be assessed by plotting the following performance

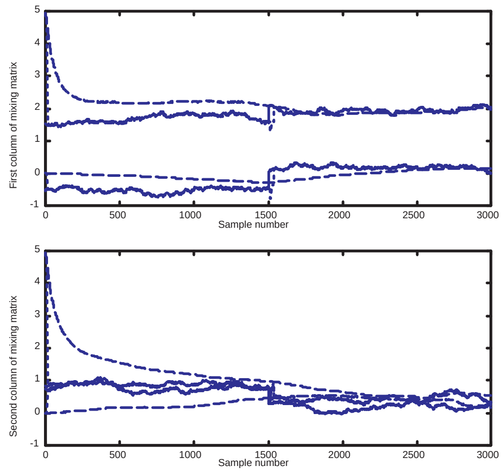

Figure 3: Tracking of the two columns of the mixing matrix, when the channel is non-stationary and changes abruptly: actual coefficients (solid lines), estimated with NGA alone (dashed lines), and estimated with the combined approach (dotted lines, coincident with the solid lines).

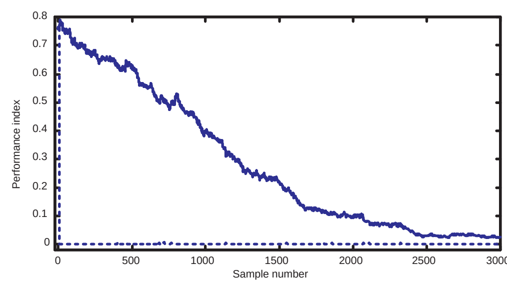

Figure 4: Evolution of the average performance index for NGA (solid line) and for the combined approach (dotted line).

index (PI)

$$\mathbf{PI}(k) = \frac{1}{m}\sum_{i=1}^{m}\left\{\sum_{j=1}^{m}\frac{|p_{ij}|^2}{max_q|p_{iq}|^2-1}\right\} + \frac{1}{m}\sum_{j=1}^{m}\left\{\sum_{i=1}^{m}\frac{|p_{ij}|^2}{max_q|p_{qj}|^2-1}\right\} \quad (20)$$

where $\mathbf{P}(k) = [p_{ij}] = \mathbf{W}(k)\mathbf{A}(k)$, and m is the number of source signals. Generally, a low PI indicates better performance. Thus, we separate the sources with NGA and the combined approach in 30 independent trials. The average performance indices for the two methods are compared in Fig. 4. It illustrates that the combined approach has a much faster initial convergence speed than NGA, and its good tracking capability results in a lower PI following initial convergence.

5. CONCLUSIONS

A combined KF and NGA approach for blind source separation of binary distributed signals, mixed by a non-stationary channel has been presented. Simulation results have shown that the KF tracks the mixing coefficients quite accurately when the mixing channel is non-stationary, with and without abrupt changes. Thus, the combined approach can quickly follow the changes in the environment, resulting in fast convergence speed, and good tracking capabilities. On-going work is considering extension to arbitrary sources.

REFERENCES

[1] S. Amari and A. Cichocki, "Adaptive blind signal processing - neural network approaches," *Proceedings of the IEEE*, vol. 86, pp. 2026–2048, 1998.

[2] C. James, K. Kobayashi, and D. Lowe, "Isolating epileptiform discharges in the unaveraged EEG using independent component analysis," in *IEE Colloq. on Medical applications for signal processing*, pp. 2/1–2/6, 1999.

[3] T.-W. Lee, A. Bell, and R. Orglmeister, "Blind source separation of real world signals," in *Proc. of Int. Conference on Neural Networks*, vol. 4, pp. 2129–2135, 1997.

[4] S. Amari, A. Cichocki, and H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, vol. 8, pp. 752–763, 1996.

[5] W. Pora, *Algorithms and structures for spatial equalisation in TDMA mobile communications.* PhD thesis, Imperial College of Science, Technology and Medicine, 2000.

[6] S. Kay, *Fundamentals of statistical signal processing estimation theory.* Prentice Hall, 1993.

[7] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.

[8] A. Cichocki, B. Orsier, A. Back, and S. Amari, "On-line adaptive algorithms in non-stationary environments using a modified comjugate gradient approach," *Proc. of the IEEE Workshop Neural Networks for Signal Processing*, pp. 316–325, 1997.