An Effective Method to Improve Convergence for Sequential Blind Source Separation

L. Yuan[1], Enfang. Sang[1], W. Wang[2], and J.A. Chambers[2]

[1] Lianxi Yuan and Enfang Sang , Harbin Acoustic Engineering College, Harbin Engineering University, China
yuan_lianxi@hotmail.com
sangef@vip.sina.com
[2] W. Wang and J. A. Chambers, the Centre of Digital Signal Processing, Cardiff University, United Kingdom,
wangw2@cf.ac.uk, chambersj@cf.ac.uk

Abstract. Based on conventional natural gradient algorithm (NGA) and equivariant adaptive separation via independence algorithm (EASI), a novel sign algorithm for on-line blind separation of independent sources is presented. A sign operator for the adaptation of the separation model is obtained from the derivation of a generalized dynamic separation model. A variable step-size sign algorithm rooted in NGA is also derived to better match the dynamics of the input signals and unmixing matrix. The proposed algorithms are appealing in practice due to their computational simplicity. Experimental results verify the superior convergence performance over conventional NGA and EASI algorithm in both stationary and non-stationary environments.

1 Introduction

Blind signal separation (BSS) is concerned with recovering the original unknown sources from their observed mixtures without. The algorithm operates blindly in the sense that except for statistical independence, no a prior information about either the sources or the transmission medium is available. BSS algorithms separate the sources by forcing the dependent mixed signals to become independent. This method has several applications in communications and signal processing. Suppose n unknown statistically independent zero mean source signals, with at most one having a Gaussian distribution, contained within $s \in \Re^n$ pass through an unknown mixing channel $A \in \Re^{m\times n}$ $(m \geq n)$, such that m mixed signals $x \in \Re^m$ are therefore observed which can be modeled as $x = As + e$, where $e \in \Re^m$ is the possible contaminating noise vector, which is usually ignored for simplicity in this study. The objective of BSS is to recover the original sources given only the observed mixtures, using the separation model $y = Wx$, where $y \notin \Re^n$ is an estimate of s to within the well-known permutation and scaling ambiguities, and $W \in \Re^{n\times m}$ is the separation matrix. The crucial assumption with conventional BSS is that the source signals are statistically independent-

L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3610, pp. 199 – 208, 2005.

ent. In this paper, we further assume that the sources have unit variance and the number of sources matches that of the number of mixtures, i.e. $m = n$, the exactly determined problem. To recover the source signals, it is frequently necessary to estimate an unmixing channel which performs the inverse operation of the mixing process, as subsequently used in the separation model. In this paper, we are particularly concerning with a family of sequential BSS algorithms. Fig.1 shows a block diagram of sequential BSS. The separating coefficients $W(k)$ are updated iteratively according to some estimate of the independence between the estimated signal components in $y(k)$. The sensor signal components in $x(k)$ are fed into the algorithm in order to estimate iteratively the source signal components, i.e. $y(k)$. Compared with block (batch)-based BSS algorithms, sequential approaches have particular practical advantage due to their computational simplicity and potentially improved performance in tracking a nonstationary environment [2]. The focus of this study is therefore the natural gradient algorithm (NGA) [1],[7] and the equivariant adaptive separation via independence algorithm (EASI)[6].

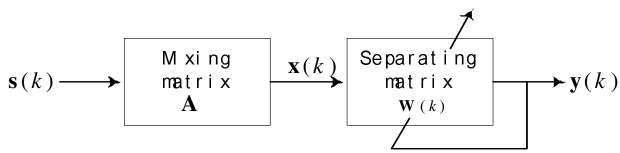

Fig. 1. Diagram of sequential blind source separation

Among important issues affecting the performance of sequential algorithms such as equation (1) are the convergence rate and the misadjustment in steady state [3]. A fixed step-size can restrict the convergence rate and can lead to poor tracking performance [2]. In contrast, an adaptive step-size can exploit the on-line measurements of the state of the separation system, from the outputs and the parameter updates. This means, the step-size can be increased for a higher convergence rate, but can be systematically decreased for reducing any misadjustment of the parameters away from their optimum settings. To improve the convergence rate, we consider using a normalization technique (leading to a sign algorithm) together with gradient-based time-varying step-size (leading to a variable step-size algorithm) in the updating process. Both techniques are shown to increase the convergence speed of the algorithm, and the sign operation can simultaneously reduce the computational complexity of the whole algorithm, additionally introduced by the adaptive step-size. The remainder of this paper is organized as follows. A sign algorithm using a normalization technique based on the standard NGA algorithm is proposed in section 2. Section 3 is dedicated to deriving a variable step-size algorithm for NGA, where the step-size is estimated from the input data and the separation matrix. Following both of the section, S-EASI algorithm was introduced. Then numerical experiments are presented in section 5 to compare the convergence performance of the proposed algorithms with that of the conventional NGA. Finally, section 6 concludes the paper.

2 Sign NGA (S-NGA)

Gradient techniques are established and well known methods for adjusting a set of parameters to minimize or maximize a chosen cost function [4]. However, simple standard gradient descent techniques is usually very slow. On these years, many novel gradient algorithms have been proposed and their better performance properties which can improve convergence speed have been proved. Here, we expect to propose a new sign-algorithm, which is based on NGA. In NGA algorithm, the discrete-time on-line updating equation of the separation matrix is denoted as

$$W(k+1) = W(k) + \mu[I - \psi(k)]W(k) \tag{1}$$

where k is the discrete-time index, μ is a positive parameter known generally as the step-size, I is an identity matrix, and $\psi(k)$ is given by

$$\psi(k) = f(y(k))y^T(k) \tag{2}$$

where $f(y(k))$ is an odd nonlinear function which acts element-wise on the output vector $y(k)$, and $(.)^T$ is the vector transpose operator.

In this section, we consider using normalization of the output vector $y(k)$ for the off-diagonal terms of $\psi(k)$. This thereby results in a sign operation on the elements of $Q(k)\psi$ which restricts the norm of the matrix $W(k)$. Our expectation is that, this will lead to faster convergence and better robustness in the adaptation. For mathematical formulation, let us consider a continuous matrix dynamic system

$$\frac{d}{dt}W(k) = -\mu\frac{\partial J(y(k),W(k))}{\partial W(k)}W^T(k)\Pi(y(k))W(k) \tag{3}$$

where $J(.)$ is a cost function from which NGA is derived, and $\Pi(y)$ is a diagonal matrix with positive elements. Equation (3) can be deemed as an extension of the standard NGA [4], since (1) is a result of $\Pi(y) = I$. By a straightforward differential matrix calculation as in [1], we obtain

$$\frac{d}{dt}W(k) = \mu\Pi(y(k)[I - \Pi^{-1}(y(k)\tilde{f}(y(k))y^T(k)\Pi(y(k))]W(k) \tag{4}$$

where $f(y(k))$ is a vector of nonlinear activation functions. Defining $\Pi^{-1}(y(k))f(y(k) = f(y(k))$ and $\mu\Pi(y(k)) = \mu(k)$, we have

$$\frac{d}{dt}W(k) = \mu(k)[I - f(y(k))y^T(k)\Pi(y(k))]W(k) \tag{5}$$

In parallel with (1), from (5), we have

$$\psi(k) \equiv f(y(k))y^T(k)\Pi(y(k)) \tag{6}$$

Denote by $f_i(y_i)$ and y_i, $i=1,\cdots,n$, the entries of $f(y)$, and y, and by π_{ij} the elements of Π, $\psi(k)$ can be re-written element-wise as

$$\psi_{ij}(k) = f_i(y_i)y_j\pi_{ij} \tag{7}$$

If π_{ij} takes the form of the normalization by y_j, i.e. $\pi_{ij} = |y_j|^{-1}$ then (6) is reduced to

$$\psi(k) \equiv f(y(k))[sign(y(k))]^T \tag{8}$$

where $sign(y(k)) = [sign(y_1(k),\cdots,sign(y_n(k)]^T$, and

$$sign\ (z) = \begin{cases} 1 & z > 0 \\ -1 & z < 0 \\ 0 & z = 0 \end{cases} \tag{9}$$

Note that, (8) could be deemed as a degenerate form of the median learning rule discussed in [4]. The introduced normalization could potentially lead to faster convergence rate because of the resulting sign activation function of the output data y increasing the magnitude of small values, which could, on the other hand, reduce the accuracy of statistics within the adaptation process, leading to inaccurate separation. To optimize both the convergence rate and separation performance, we suggest to use different normalization schemes for the elements of $\psi(k)$. Particularly, Π does not hold fixed values at its diagonal elements, but these change according to the association between $f(y(k))$ and $y(k)$. That is, (7) is re-written in the discrete-time form as

$$\psi_{ij} = \begin{cases} f_i(y_i(k))y_i(k) & i=j \\ f_i(y_i(k))sign(y_j(k)) & i\neq j \end{cases} \tag{10}$$

Using the Kronecker dot product Θ (element-wise product of matrices), we have the following concise expression

$$\psi(k) \equiv f(y(k))y^T(k)\Theta\Phi(y(k)) \tag{11}$$

where $\Phi(y(k))$ is derived from Π and (10), i.e. the entries of Φ are denoted as

$$\varphi_{ij} = \begin{cases} 1 & i=j \\ |y_j|^{-1} & i\neq j \end{cases} \tag{12}$$

Note that, (11) can also be written as

$$\Phi(k) \equiv diag[f(y(k))y^T(k)] + off[f(y(k))sign(y^T(k))] \tag{13}$$

where $diag[.]$ and $off[.]$ denote the operation of taking the diagonal elements and off-diagonal elements of a matrix respectively.

We call the adaptation procedure of using (11) and (12) the sign natural gradient algorithm (S-NGA). Compared with the NGA using (2), the sign algorithm (SA) has reduced computational complexity, i.e. n(n-1) multiplications in (2) are replaced with simple sign tests which are easily implementable. However, for each k, the off-diagonal elements of $\psi(k)$ are not continuous (see equation (10)), this where makes the analysis of such an algorithm more difficult than that of (1). However, it is straightforward to show the algorithm is Lyapunov stable. Noticing that $W^T \Pi W = (\sqrt{\Pi} W)^T (\sqrt{\Pi} W)$ in (3), where $\sqrt{\Pi}$ represents a diagonal matrix whose diagonal entries are the square root of the corresponding diagonal elements of Π, and denoting by w_{ij}, γ_{ij}, and ψ_{ij} $i, j = 1, \cdots n$, the elements of W, $\sqrt{\Pi} W$, and $\frac{\partial J}{\partial W} (\sqrt{\Pi} W)^T$, we obtain from (3) that

$$\begin{aligned} \frac{d}{dt} J(y(k), W(k)) &= \sum_{i,j} \frac{\partial J}{\partial w_{ij}} \frac{dw_{ij}}{dt} \\ &= -\sum_{i,j} \frac{\partial J}{\partial w_{ij}} \sum_{k} \psi_{ik} \gamma_{kj} \\ &= -\sum_{i,k} \psi_{ik}^2 \leq 0 \end{aligned} \quad (14)$$

where zero is obtained if and only if $dW(k)/dk = 0$, which means the solution to W W is an equilibrium of (3).

3 Variable Step-Size Sign NGA (VS-S-NGA)

It has been shown [2] that, as compared with using a fixed step-size which would restrict convergence rate, the algorithm with an adaptive step-size has an improved tracking performance for a non-stationary environment, i.e., the value of which is adjusted according to the time-varying dynamics of the input signals and the separating matrix. As another contribution, we therefore derive a gradient adaptive step-size algorithm for the NGA algorithm, which adapts the step-size in the form of

$$\mu(k) = \mu(k-1) = \rho \nabla_{\mu} J(k) \Big|_{u=u(k-1)} \quad (15)$$

where ρ is a small constant, and $J(k)$ is an instantaneous estimate of the cost function from which the NGA algorithm is derived. To proceed, we use an inner product of matrices defined as [2],

$$\langle C, D \rangle = tr(C^T D) \quad (16)$$

where $\langle . \rangle$ denotes the inner product, $tr(.)$ is the trace operator, and $C, D \in \Re^{m \times n}$ Therefore, exploiting (16), the gradient term on the right hand side of (5) can be evaluated as

$$\begin{aligned} \nabla_{\mu} J(k) |_{\mu=\mu(k-1)} &= \langle \partial J(k) / \partial W(k), \partial W(k) / \partial \mu(k-1) \rangle \\ &= tr(\partial J(k) / W(k)^T \times \partial W(k) / \partial \mu(k-1) \end{aligned} \quad (17)$$

$$\partial J(k) / \partial W(k) = -[I - f(y(k)) y^T(k)] W(k) \quad (18)$$

which is the instantaneous estimate of the natural gradient of the cost function of $J(k)$. From the equation (1), the separating matrix W at time k is obtained as

$$W(k) = W(k-1) + \mu(k-1)\left[I - f(y(k-1))y^T(k-1)\right]W(k-1) \tag{19}$$

Following the approach from [2] and [5], from the above equation, we have

$$\partial W(k) \Big/ \partial \mu(k-1) = \left[I - f(y(k-1))y^T(k-1)\right]W(k-1) \tag{20}$$

Using the notation of (2) for $\psi(k)$ in the standard NGA algorithm and denoting we have

$$\Gamma(k) \equiv \left[I - \psi(k)\right]W(k) \tag{21}$$

$$\nabla_{\mu} J(k)|_{\mu=\mu(k-1)} = -tr(\Gamma^T(k)\Gamma(k-1)) \tag{22}$$

Hence, an adaptive step-size with the form of (15) can be written as

$$\mu(k) = \mu(k-1) + \rho tr(\Gamma^T(k)\Gamma(k-1)) \tag{23}$$

which can be estimated from the input signals and the separation matrix. (21) has a similar form as the equation (7) in [2], which was derived for an equivariant adaptive source separation via independence (EASI) algorithm[6]. The separation procedure using (1), (2), (21) and (23) represents the proposed variable step-size NGA algorithm (VS-NGA). Following a similar procedure as in section 2, see (6) and (11), and as in this section, see (18) and (20), it is straightforward to derive an adaptive step-size algorithm using different normalization for the off-diagonal elements of $\psi(k)$. In this case, $\psi(k)$ takes the form of (11). We represent (1), (11), (21) and (23) the sign version of the variable step-size NGA algorithm, i.e., VS-S-NGA for notational simplicity.

4 Sign-EASI

Cardoso proposed EASI algorithm in 1996. EASI algorithm is a kind of adaptive algorithms for source separation which implements an adaptive version of equivariant estimation. It is based on the idea of serial updating: this specific form of matrix updates systematically yields algorithms with a simple structure, for both real and complex mixtures, and its performance does not depend on the mixing matrix. So convergence rates, stability conditions and interference rejection levels of EASI algorithm only depend on distributions of the source signals. In order to reduce computation complexity of the algorithm and obtain a satisfied stability, sign function is applied to this kind of algorithm. Firstly, the separating matrix update equation for EASI algorithm is given by

$$\begin{aligned} W(k+1) = W(k) + \mu[I - y(k)y^T(k) - f(y(k)y^T(k) + \\ y(k)f(y(k))^T]W(k) \end{aligned} \tag{24}$$

Here we also set a parameter $\psi(k) = f(y(k))y^T(k)$ to substitute $f(y(k))$ and $y(k)$ in the upper equation, then (24) can be rewritten as:

$$W(k+1) = W(k) + \mu\left[I - y(k)y^T(k) - \psi(k) + \psi(k)^T\right]W(k) \quad (25)$$

In order to easily understand and keep consistent with the NGA algorithm, all of the parameters in the above equation are defined just as in the section 2, i.e. $Q(k)$ in the (25) takes the same form as in the section 2:

$$\psi(k) \equiv diag[f(y(k))y^T(k)] + off[f(y(k)sign(y^T(k))] \quad (26)$$

Therefore, seeing in the section 2, we can omit some middle procedures and directly derive the final algorithm what we expect. Equation (25) and (26) are all together called the Sign EASI algorithm, namely S-EASI.

5 Numerical Experiments

In the first experiment, we mix a fixed sinusoidal signal with a randomly selected uniform source signal by using a 2-by-2 $(m = n = 2)$ matrix $A_0 = randn(m,n)$, i.e. Zero mean, independent white Gaussian noise with standard deviation 0.1 was added to the mixtures. A cubic non-linearity $f(.)$ was used as the activation function. The performance index (PI) [1], as a function of the system matrix $G = WA$, was used to evaluate the proposed algorithm

$$PI(G) = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{m}\frac{|g_{ik}|}{\max_k |g_{ik}|}\right) - 1\right] + \left[\frac{1}{m}\sum_{k=1}^{m}\left(\sum_{i=1}^{n}\frac{|g_{ik}|}{\max_i |g_{ik}|}\right) - 1\right] \quad (27)$$

where g_{ik} is the ik-th element of G. The initial value of μ for all the algorithms was set to 0.0045, $\rho = 2\times10^{-5}$, and 200 Monte Carlo trials were run for an averaged performance. The same simulation conditions were used for all the algorithms to allow fair comparison. Fig.2 shows convergence behavior of the various approaches. From Fig.2 , it is found that the proposed sign algorithms have much faster convergence speed. For example, for the fixed step size, S-NGA needs approximately 2000 samples to converge, whereas the conventional NGA needs approximately 3250 samples. Note that, we mean the convergence by the PI reduced to 0.02 (corresponding to an approximately successful separation). For the adaptive step-size, VS-S-NGA only requires approximately 1050 samples for convergence, however, VS-NGA requires approximately 1700 samples. It is clear that VS-S-NGA has the fastest convergence rate, which is a very promising property for sequential algorithms.Without any change of parameters , we continued to realize the second group of simulation with S-EASI and EASI algorithms on the same conditions. Fig.3 showed a compared result between them. S-EASI arrived its steady convergence near the approximate 1300 samples, while EASI had to need around 1800 samples to satisfy this requirement. From Fig.3, it clearly proved that the convergence rate of the S-EASI algorithm was faster than EASI. Here, we only provided the simulation results with a fixed step size. For the varying adaptive step-size, we also gained a similar conclusion, but it was not very stable. So we still need further experiments to verify it.

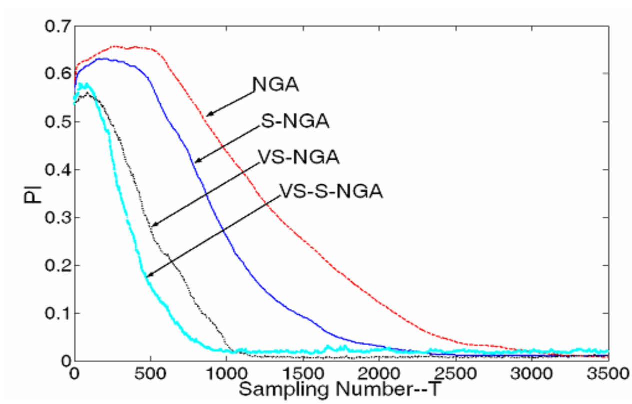

Fig. 2. Comparison of convergence rate by performance index in a stationary environment

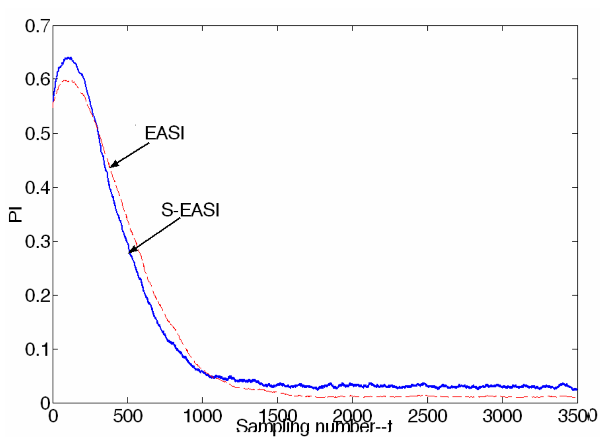

Fig. 3. Comparison of convergence rate between S-EASI and EASI in a stationary environment

In the second experiment, the different approaches were examined for a non-stationary environment. To this end, we use the following time-varying mixing matrix

$$A = A_0 + \Psi \tag{28}$$

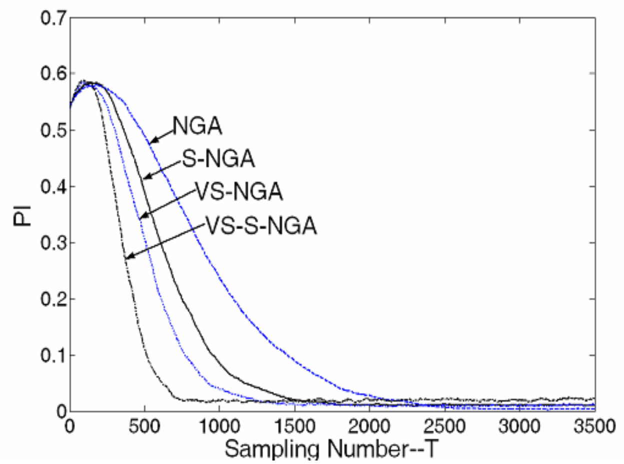

Fig. 4. Comparison of convergence rate by performance index in a non-stationary environment

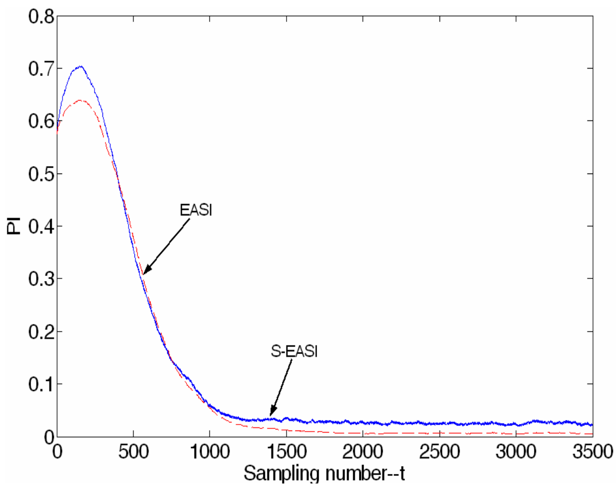

Fig. 5. Comparison of convergence rate between S-EASI and EASI in a non-stationary environment

where $\Psi = \alpha\Psi + \beta.randn(size(A,1)$, $randn(.)$ and $size(.)$ are MATLAB built-in functions, and the initial Ξ is set to a null matrix. A_0 is the same as in (27). Here α is set to 0.95 and β to 0.001. Other parameters are the same as those in the first experiment. Again, their convergence performances are compared in Fig.4 and Fig.5 respectively. For the Fig.4, we observed similar performance improvement gained for the proposed approaches in a non-stationary environment. Note that, lower PI generally indicates a better separation performance. In both Fig.2 and Fig.4, although we have not observed much difference between the final separation performance by S-NGA and VS-S-NGA in terms of *PI* measurement, the key point is that the reduced complexity improves the rate of convergence. The same conclusion on S-EASI algorithm also can be made from Fig.3 and Fig.5.

6 Conclusions

A new sign and variable step-size natural gradient algorithm for on-line blind separation of independent sources has been presented, also including a fixed step-size sign EASI algorithm. The derivation is based on the gradient calculation of a generalized dynamic equation. By applying the sign operation to NGA and EASI, these separation algorithms have been found to have much faster convergence rate as compared with the conventional natural gradient algorithm and EASI algorithm. The algorithm was shown to be Lyapunov stable. Through the results of simulations, we prove both of new algorithms can bring us a satisfied convergence rate and reduced computation complexity. Although variable step-size sign EASE algorithm need further testing, we still derived a variable step-size algorithm for the natural gradient learning which was also shown to have faster convergence rate and than using a fixed step-size algorithm.

References

1. A. Cichocki and S. Amari :Adaptive blind signal and image processing: learning algorithms and applications, John Wiley & Sons(2002)
2. J.A. Chambers, M. G. Jafari and S. McLaughlin.: Variable step-size EASI algorithm for sequential blind source separation, Elect. Lett. (2004) 393-394
3. S.C. Douglas and A. Cichocki : On-line step-size selection for training of adaptive systems, IEEE Signal Processing Magazine, no. 6(1997) 45-46
4. P. Georgiev, A. Cichocki and S. Amari :On some extensions of the natural gradient algorithm, Proc. ICA(2001) 581-585
5. V.J. Mathews and Z. Xie: A stochastic gradient adaptive filter with gradient adaptive step size, IEEE Trans. Signal Process, vol. 41, no. 6(1993) 2075-2087
6. J.-F. Cardoso and B.H. Laheld: Equivariant adaptive source separation, IEEE Trans. Signal Process, 44(1996) 3017-3030
7. S.Amari: Natural Gradient Works Efficiently in Learning, Neural Computation 10(1998) 251-276
8. S.Amari and S.C.Douglas: Why Natural Gradient, In Proc. IEEE International Conference Acoustics, Speech, Signal Processing, volume II, Seattle, WA, May (1998) 1213-1216