ADAPTIVE BLIND SOURCE SEPARATION BY SECOND ORDER STATISTICS AND NATURAL GRADIENT

Yong Xiang, Karim Abed-Meraim, and Yingbo Hua

Department of Electrical and Electronic Engineering
University of Melbourne, Parkville, Victoria 3052 Australia
E-mail:yxiang, abed, yhua@ee.mu.oz.au

ABSTRACT

Separation of sources that are mixed by an unknown (hence, "blind") mixing matrix is an important task for a wide range of applications. This paper presents an adaptive blind source separation method using second order statistics (SOS) and natural gradient. The SOS of observed data is shown to be sufficient for separating mutually uncorrelated sources provided that the temporal coherences of all sources are linearly independent of each other. By applying the natural gradient, new adaptive algorithms are derived that have a number of attractive properties such as invariance of asymptotical performance (with respect to the mixing matrix) and guaranteed local stability. Simulations suggest that the new algorithms are highly efficient and outperform some of the best existing ones.

1. INTRODUCTION

In many applications such as remote sensing, data communications, speech processing and medical diagnosis, one is interested to separate different sources that are mixed by some undesired and unknown matrix. This problem has been traditionally tackled using various kinds of information measures that are based on higher order statistics (HOS), e.g., [1]. While statistically superior provided sufficient data, the HOS approach is often too costly in computation and hardly adaptable to a fast changing environment (such as fast time-varying mixing matrix). More recently, the second order statistics (SOS) approach becomes a more attractive one, e.g., see [3, 4]. The work in [3, 4] provided a very good foundation for blind source separation using SOS. However, this paper provides a significant further development on that foundation. In particular, as shown in Figure 1, our goal here is to design an adaptive source separation method.

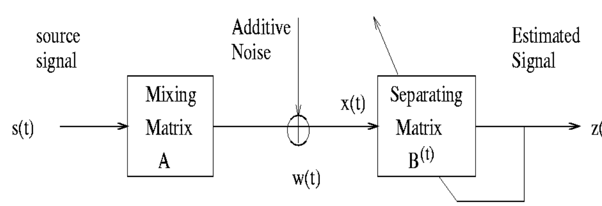

Figure 1. Adapting a separating matrix

We first generalize the result in [7] by giving necessary and sufficient identifiability conditions for BSS using a finite set of correlation coefficients. Then, necessary and sufficient separation equations are given based on which new contrast functions are formulated. To minimize the latter, two adaptive algorithms are given for the noiseless and noisy (white or colored) case, respectively. The new algorithms are shown to enjoy the uniform performance and local stability properties. Simulation results and comparison with EASI algorithm [5] are presented.

2. PROBLEM FORMULATION

Consider n mutually uncorrelated signals[1] whose n linear combinations are observed in noise:

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{w}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t)$$

where $\mathbf{s}(t) = [s_1(t), \cdots, s_n(t)]^T$ is the $n \times 1$ complex *source vector*, $\mathbf{w}(t) = [w_1(t), \cdots, w_n(t)]^T$ is the $n \times 1$ complex *noise vector*, and $\mathbf{A}$ is the $n \times n$ full rank *mixing matrix*. The source signal vector $\mathbf{s}(t)$, is assumed to be a multivariate zero-mean stationary complex stochastic process with second order moments:

$$\begin{aligned}\mathbf{S}(\tau) &\stackrel{\text{def}}{=} E\left(\mathbf{s}(t+\tau)\mathbf{s}^*(t)\right)\\ &= \text{diag}[\rho_1(\tau), \cdots, \rho_n(\tau)]\end{aligned}$$

where $\rho_i(\tau) \stackrel{\text{def}}{=} E(s_i(t+\tau)s_i^*(t))$, and the superscript $*$ denotes the conjugate transpose operator. The additive noise $\mathbf{w}(t)$ is modeled as a stationary zero-mean complex random process.

The purpose of blind source separation is to find a separating matrix, i.e., an $n \times n$ matrix $\mathbf{B}$ such that $\hat{\mathbf{s}}(t) = \mathbf{B}\mathbf{x}(t)$ is an estimate of the source signals.

Before proceeding, note that a complete identification of the separating matrix $\mathbf{B}$ (or equivalently the mixing matrix $\mathbf{A}$) is impossible in the blind context, because exchange of a scalar between the source signal and the corresponding column of $\mathbf{A}$ leaves the observations unaffected. Also note that the *numbering* of the signals is immaterial. It follows that the best that one can do is to determine $\mathbf{B}$ up to a permutation and scaling of its columns [3]. Therefore, $\mathbf{B}$ is said to be a separating matrix if

$$\mathbf{B}\mathbf{y}(t) = \mathbf{P}\mathbf{\Lambda}\mathbf{s}(t)$$

where $\mathbf{P}$ is a permutation and $\mathbf{\Lambda}$ a non-singular diagonal matrix.

3. FUNDAMENTAL RESULTS

We present here the fundamental results needed throughout the rest of the paper. More precisely, we present two separation criteria for

[1] We assume for simplicity and without loss of generality that $n = m$ (as many sources as 'sensors').

the stationary, temporally correlated source signals (see [4] for detailed proofs of the Theorems below). Consider first the noiseless case. We have the following Theorem.

Theorem 1 *(noiseless case): Let $\tau_1, \cdots, \tau_K$ be $K \geq 1$ (non-zero) time lags and define the $1 \times (K+1)$ vectors $\rho_i = [\rho_i(0), \rho_i(\tau_1), \cdots, \rho_i(\tau_K)]$, $i = 1, \cdots, n$. Then, BSS can be achieved using the output correlation matrices at time lags $0, \tau_1, \cdots, \tau_K$ if and only if:*

$$\rho_i \text{ and } \rho_j \text{ are linearly independent for } i \neq j \tag{1}$$

Assume that (1) holds and let $\mathbf{z}(t)$ be an $n \times 1$ vector given by $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$. Define $r_{ij}(k) \stackrel{\text{def}}{=} E(z_i(t+k)z_j^*(t))$. Then, $\mathbf{B}$ is a separating matrix if and only if

$$r_{ij}(k) = 0 \quad and \quad r_{ii}(0) > 0 \tag{2}$$

for all $1 \leq i \neq j \leq n$ and $k = 0, \tau_1, \cdots, \tau_K$.

In the case of temporally white additive noise (with *unknown* spatial covariance), the previous result can be extended as follows:

Theorem 2 *(noisy case): Let $\tau_1, \cdots, \tau_K$ be $K > 1$ (non-zero) time lags and define the $1 \times K$ vectors $\tilde{\rho}_i = [\rho_i(\tau_1), \cdots, \rho_i(\tau_K)]$, $i = 1, \cdots, n$. Then, BSS can be achieved using the output correlation matrices at time lags $\tau_1, \cdots, \tau_K$ if and only if:*

$$\tilde{\rho}_i \text{ and } \tilde{\rho}_j \text{ are linearly independent for } i \neq j \tag{3}$$

Assume that (3) holds and let $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$ be an $n \times 1$ vector. Then, $\mathbf{B}$ is a separating matrix if and only if

$$r_{ij}(k) = 0 \quad and \quad \sum_{k=\tau_1}^{\tau_K} |r_{ii}(k)| > 0 \tag{4}$$

for all $1 \leq i \neq j \leq n$ and $k = \tau_1, \cdots, \tau_K$.

We can see that in the trivial case where the sources show identical normalized spectra, conditions (1) and (3) cannot be satisfied and thus BSS cannot be achieved. Conversely, when the source signals have different normalized spectra, it is always possible (with probability one) to find a set of time lags $\tau_1, \cdots, \tau_K$ such that (1) (or (3)) is met. This corresponds to the second-order identifiability condition found in [7]. It is worth to point out that the condition in [7] is a necessary and sufficient condition for BSS using the whole set of SOS statistics while condition (1) (resp. (3)) is a necessary and sufficient condition for BSS using a finite set of correlation coefficients including (resp. excluding) the zero-lag one.

4. ADAPTIVE ALGORITHMS

In this section, we present SOS-based contrast functions and their corresponding adaptive optimization algorithms.

4.1. Algorithm 1: Noiseless case

To solve the separating equations (2), we consider the following least squares error criterion:

$$G_1(\mathbf{z}) \stackrel{\text{def}}{=} \sum_{k=\tau_0}^{\tau_K} \sum_{1 \leq i < j \leq m} [|r_{ij}(k) + r_{ji}(k)|^2 + |r_{ij}(k) - r_{ji}(k)|^2] + \sum_{i=1}^{m} |r_{ii}(0) - 1|^2 \tag{5}$$

(with $\tau_0 = 0$). It is easy to see that $G_1(\mathbf{z})$ is a contrast function which minimization is equivalent to solving (2). The separation criterion becomes

$$\mathbf{B} \text{ is a separating matrix} \Leftrightarrow G_1(\mathbf{z}(t)) = 0 \tag{6}$$

where $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$. Our approach to adaptive source separation may be motivated by first considering *batch* estimation. Consider the problem of estimating $\mathbf{B}$ using natural gradient technique [2].

The choice of natural gradient technique has been motivated by: **(i)** the fact that natural gradient online learning gives the Fisher efficient estimator in the sense of asymptotic statistics when the loss function is differentiable [2], **(ii)** the fact that the natural gradient at point $\mathbf{B}$ depends only on the distribution of $\mathbf{z} = \mathbf{B}\mathbf{x}$ and not on $\mathbf{B}$ itself, and as a consequence, natural gradient based algorithms enjoy uniform performance properties (see subsection 4.3), and **(iii)** the fact that it is an approximate Newton technique which can be very simply computed (no Hessian inversion) under the additional assumption that at iteration p, $\mathbf{B}^{(p)}$ is close to a separating matrix.

The solutions are obtained iteratively in the form

$$\mathbf{B}^{(p+1)} = (\mathbf{I} + \epsilon^{(p)})\mathbf{B}^{(p)} \tag{7}$$

$$\mathbf{z}^{(p+1)}(t) = (\mathbf{I} + \epsilon^{(p)})\mathbf{z}^{(p)}(t) \tag{8}$$

At iteration p, a matrix $\epsilon^{(p)}$ is determined from a local linearization of $G_1(\mathbf{B}\mathbf{x}(t))$. The procedure is illustrated as follows. By using (8), we have

$$r_{ij}^{(p+1)}(k) = r_{ij}^{(p)}(k) + \sum_{q=1}^{m} \epsilon_{jq}^{*(p)} r_{iq}^{(p)}(k) + \sum_{l=1}^{m} \epsilon_{il}^{(p)} r_{lj}^{(p)}(k) + \sum_{l,q=1}^{m} \epsilon_{il}^{(p)} \epsilon_{jq}^{*(p)} r_{lq}^{(p)}(k)$$

where

$$r_{ij}^{(p)}(k) \stackrel{\text{def}}{=} E\left(z_i^{(p)}(t+k) z_j^{*(p)}(t)\right) \tag{9}$$

Under the assumption that $\mathbf{B}^{(p)}$ is close to a separating matrix, it follows that

$$|\epsilon_{ij}^{(p)}| \ll 1, \text{ and } |r_{ij}^{(p)}(k)| \ll 1 \text{ for } i \neq j$$

and thus, a first order approximation of $r_{ij}^{(p+1)}(k)$ is given by

$$r_{ij}^{(p+1)}(k) \approx r_{ij}^{(p)}(k) + \epsilon_{ji}^{*(p)} r_{ii}^{(p)}(k) + \epsilon_{ij}^{(p)} r_{jj}^{(p)}(k) \tag{10}$$

By replacing (10) into (5), we obtain the following least squares (LS) minimization problem for $i \neq j$:

$$\min \| [\mathbf{r}_{jj}^{(p)}, \ \mathbf{r}_{ii}^{(p)}] \mathbf{E}_{ij}^{(p)} + [\frac{1}{2}(\mathbf{r}_{ij}^{(p)} + \mathbf{r}_{ji}^{(p)}), \ \frac{1}{2\mathrm{J}}(\mathbf{r}_{ij}^{(p)} - \mathbf{r}_{ji}^{(p)})] \|$$

where

$$\mathbf{E}_{ij}^{(p)} \stackrel{\text{def}}{=} \begin{bmatrix} \Re e(\epsilon_{ij}^{(p)}) & \Im m(\epsilon_{ij}^{(p)}) \\ \Re e(\epsilon_{ji}^{(p)}) & -\Im m(\epsilon_{ji}^{(p)}) \end{bmatrix} \tag{11}$$

$$\mathbf{r}_{ij}^{(p)} = [r_{ij}^{(p)}(0), r_{ij}^{(p)}(\tau_1), \cdots, r_{ij}^{(p)}(\tau_K)]^T \tag{12}$$

A solution to the LS minimization problem is given by

$$\mathbf{E}_{ij}^{(p)} = -[\mathbf{r}_{jj}^{(p)}, \mathbf{r}_{ii}^{(p)}]^{\#} [\frac{1}{2}(\mathbf{r}_{ij}^{(p)} + \mathbf{r}_{ji}^{(p)}), \frac{1}{2\mathrm{J}}(\mathbf{r}_{ij}^{(p)} - \mathbf{r}_{ji}^{(p)})] \tag{13}$$

where $\jmath = \sqrt{-1}$ and $\mathbf{A}^{\#}$ denotes the pseudo-inverse of a matrix $\mathbf{A}$. Similarly, for $i = j$ we obtain

$$\epsilon_{ii}^{(p)} = \frac{1 - r_{ii}^{(p)}(0)}{2r_{ii}^{(p)}(0)} \quad (14)$$

Now, to derive an adaptive version of the above batch algorithm we replace in the above formulae the iteration index p by the time index t and estimate adaptively the correlation coefficients $r_{ij}^{(t)}(k)$. The adaptive algorithm can be summarized as follows: At time instant $t+1$

- Update the correlation matrices, i.e., $\mathbf{R}(k) = E(\mathbf{z}(t+k)\mathbf{z}^*(t))$, $k = 0, \tau_1, \cdots, \tau_K$, using the following averaging technique:

$$\begin{aligned} \mathbf{z}(t+1) &= \mathbf{B}^{(t)}\mathbf{x}(t+1) \\ \mathbf{R}^{(t+1)}(k) &= (1-\lambda_{t+1})\mathbf{R}^{(t)}(k) + \\ & \qquad \lambda_{t+1}\mathbf{z}(t+1)\mathbf{z}^*(t+1-k) \end{aligned}$$

  where λ_t is a decreasing and positive sequence. Note that $r_{ij}^{(t)}(k)$ is the (i,j)-th entry of $\mathbf{R}^{(t)}(k)$.

- Estimate $\epsilon^{(t+1)}$ using equations (13) and (14) and the updated correlation coefficients $r_{ij}^{(t+1)}(k)$.

- Update the value of the separating matrix, the correlation matrices $\mathbf{R}(k)$, $k = 0, \tau_1, \cdots, \tau_K$, and the estimated sources $\mathbf{z}(t+2-k)$, $k = \tau_1, \cdots, \tau_K$:

$$\begin{aligned} \mathbf{B}^{(t+1)} &= (\mathbf{I} + \epsilon^{(t+1)})\mathbf{B}^{(t)} \\ \mathbf{R}^{(t+1)}(k) &= (\mathbf{I} + \epsilon^{(t+1)})\mathbf{R}^{(t+1)}(k)(\mathbf{I} + \epsilon^{(t+1)})^* \\ \mathbf{z}(t+2-k) &= (\mathbf{I} + \epsilon^{(t+1)})\mathbf{z}(t+2-k) \end{aligned}$$

4.2. Algorithm 2: Noisy case

Similarly to the approach shown in subsection 4.1, we define from Theorem 2 the following contrast function:

$$\begin{aligned} G_2(\mathbf{z}) \stackrel{\text{def}}{=} \sum_{k=\tau_1}^{\tau_K} \sum_{1 \le i < j \le m} [|r_{ij}(k) + r_{ji}(k)|^2 + \\ |r_{ij}(k) - r_{ji}(k)|^2] + \sum_{i=1}^{m} | \sum_{k=\tau_1}^{\tau_K} |r_{ii}(k)| - 1|^2 \end{aligned} \quad (15)$$

Minimizing this contrast function leads to an adaptive algorithm similar to the previous one except for equations (13) and (14) that become:

$$\epsilon_{ii}^{(p)} = \frac{1 - \sum_{k=\tau_1}^{\tau_K} |r_{ii}^{(p)}(k)|}{2\sum_{k=\tau_1}^{\tau_K} |r_{ii}^{(p)}(k)|} \quad (16)$$

$$\mathbf{E}_{ij}^{(p)} = -[\tilde{\mathbf{r}}_{jj}^{(p)}, \tilde{\mathbf{r}}_{ii}^{(p)}]^{\#}[\frac{1}{2}(\tilde{\mathbf{r}}_{ij}^{(p)} + \tilde{\mathbf{r}}_{ji}^{(p)}), \frac{1}{2\jmath}(\tilde{\mathbf{r}}_{ij}^{(p)} - \tilde{\mathbf{r}}_{ji}^{(p)})] \quad (17)$$

where

$$\tilde{\mathbf{r}}_{ij}^{(p)} = [r_{ij}^{(p)}(\tau_1), \cdots, r_{ij}^{(p)}(\tau_K)]^T \quad (18)$$

4.3. Uniform performance and stability

Uniform performance: This is an important and natural property in the context of source separation [5]. The uniform performance of an estimator is the property to have (in noiseless case) the same asymptotic performance whatever the mixing matrix is. In other words, the performance of source separation depends only on the source statistics and not on the mixing matrix.

In our case, the uniform performance is insured because of the use of Natural gradient. This can be seen in the following way: Consider the evolution of the global system $\mathbf{C}^{(t)} = \mathbf{B}^{(t)}\mathbf{A}$ given by right multiplying (7) by matrix $\mathbf{A}$:

$$\mathbf{C}^{(t+1)} = (\mathbf{I} + \epsilon(\mathbf{C}^{(t)}\mathbf{s}(t)))\mathbf{C}^{(t)}$$

The evolution of $\mathbf{C}^{(t)}$ would be same for two mixing matrices $\mathbf{A}$ and $\mathbf{A}'$ if we initialize the algorithm with $\mathbf{B}_0$ and $\mathbf{B}'_0$ such that $\mathbf{B}_0\mathbf{A} = \mathbf{B}'_0\mathbf{A}'$. Hence, with respect to the global system $\mathbf{C}^{(t)}$, changing the mixing matrix is equivalent to changing the initial value of $\mathbf{B}$ which obviously does not change the asymptotic performance of the algorithm.

Stability: We study here the stability of the proposed adaptive algorithms at the solution (optimum) points. Thanks to the uniform performance (or equivariance) property, it suffices to study the stability of the algorithms for the global system $\mathbf{C}^{(t)}$ at the optimum point $\mathbf{C}_\star = \mathbf{I}$. General results for stochastic algorithms, e.g., [6], show that $\mathbf{C}_\star$ is locally asymptotically stable if all the eigenvalues of matrix $\mathbf{\Gamma}$ defined as:

$$\mathbf{\Gamma} = -\frac{\partial \mathbf{\Psi}(\mathbf{C})}{\partial \mathbf{C}}|_{\mathbf{C}=\mathbf{C}_\star} \quad (19)$$

have positive real parts, where $\mathbf{\Psi}(\mathbf{C}) = E(\epsilon(\mathbf{C}\mathbf{s}(t))\mathbf{C})$. Simple calculations (that are omitted here for simplicity) reveal that for both algorithms the linear approximation of $\mathbf{\Psi}$ in the neighborhood of $\mathbf{C}_\star = \mathbf{I}$ is

$$\mathbf{\Psi}(\mathbf{I} + \mathbf{e}) = -\mathbf{e} + o(\mathbf{e})$$

which shows that $\mathbf{\Gamma} = \mathbf{I}$. Thus, unlike the EASI algorithm in [5], the stability is satisfied without any additional assumption.

5. NUMERICAL SIMULATIONS

In our simulation an array of $n = 2$ sensors with half wavelength spacing receives two signals in the presence of stationary complex temporally white noise. The two source signals are generated by filtering complex circular white Gaussian processes by an AR model of order 1 with coefficient $a_1 = 0.95e^{0.5\jmath}$ and $a_2 = 0.5e^{0.7\jmath}$. The time lags (delays) involved are $\tau_1 = 1$ and $\tau_2 = 2$. The sources arrive from the directions $\phi_1 = 10$ and $\phi_2 = 30$ degrees. In order to evaluate the performance of our algorithms, we define a performance index called the mean rejection level (we assume here that the permutation indeterminacy is $\mathbf{P} = \mathbf{I}$) as [3]

$$\mathcal{I}perf \stackrel{\text{def}}{=} \sum_{p \neq q} \frac{E|(BA)_{pq}|^2}{E|(BA)_{pp}|^2}$$

The signal to noise ratio is defined as $SNR = -10\log_{10}\sigma^2$, where σ^2 is the noise variance; the mean rejection level is estimated by averaging 50 independent trials.

Figure 2 shows the performance of Algorithm 1 (derived from Theorem 1) for different SNRs. The noise signal is temporally and spatially white. As we can expect it, Algorithm 1 performs well only when the signal to noise ratio is ‘sufficiently’ high.

Figure 3-a and 3-b show the performances of Algorithm 2 (derived from Theorem 2) for spatially white noise and spatially colored noise, respectively. In the latter case, the noise covariance is of the form $\mathbf{R}_n = \sqrt{n}\sigma^2 \mathbf{Q}\mathbf{Q}^H / ||\mathbf{Q}||^2$, where $\mathbf{Q}$ is given by $\mathbf{Q}_{ij} = 0.9^{|i-j|}$. Algorithm 2 performs well even at relatively low SNR but is outperformed by Algorithm 1 for high SNRs.

In Figure 4, we compare Algorithm 1 with EASI algorithm [5]. This is an adaptive HOS-based algorithm for BSS using natural gradient. In this case, the two source signals are generated by filtering QAM4 processes by the previous AR filter. The noise is a spatially white Gaussian process and its variance is $\sigma^2 = -30$dB. In this context, we can see that Algorithm 1 outperforms the EASI algorithm in both convergence speed and estimation accuracy.

In Figure 5, we check the uniform performance property of Algorithm 1 (in the noiseless case). We consider 3 mixing matrices $\mathbf{A}1 = [\mathbf{a}(10^o), \mathbf{a}(30^o)]$, $\mathbf{A}2 = [\mathbf{a}(10^o), \mathbf{a}(15^o)$, and $\mathbf{A}3$ a complex random matrix. Figure 5 shows that the performance of Algorithm 1 at the steady state nearly does not depend on the mixting matrix $\mathbf{A}$.

6. CONCLUSION

This paper presents new blind source separation methods for temporally correlated stationary sources. Two SOS-based contrast functions are introduced for noiseless and noisy cases. Adaptive algorithms (namely Algorithm 1 and Algorithm 2) based on natural gradient technique are proposed to minimize the contrast functions and perform the BSS. Unlike most existing SOS methods, Algorithm 2 can deal with spatially colored noise case. The uniform performance and the stability of the proposed algorithms are shown. The effectiveness of the new algorithms is demonstrated by numerical experiments and comparison with EASI algorithm proposed in [5].

7. REFERENCES

[1] H. Yang and S. Amari, “Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information”, *Neural Computation*, 9, pp. 1457–1482, 1997.

[2] S. Amari and S. Douglas, “Why Natural Gradient?”, *Proc. ICASSP, (Munich, Germany)*, vol. II, pp. 1213–1216, 1998.

[3] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines, “Blind source separation using second order statistics,” *IEEE-T-SP*, pp. 434–444, Feb. 1997.

[4] K. Abed-Meraim, Y. Hua, and A. Belouchrani, “A general framework for blind source separation using second order statistics,” *Proc. DSP workshop*, UTAH, USA,1998.

[5] J. F. Cardoso and B. Laheld, “Equivariant adaptive source separation,” *IEEE-T-SP*, vol. 44, pp. 3017–3030, Dec. 1996.

[6] A. Benveniste, M. Metivier, and P. Priouret, “Adaptive algorithms and stochastic approximations”, *Springer Verlag*, 1990.

[7] L. Tong, R. Liu and Y. H. V. C. Soon, “Indeterminacy and identifiability of blind identification,” *IEEE Tr. on CS*, vol. 38, pp. 499–509, May 1991.

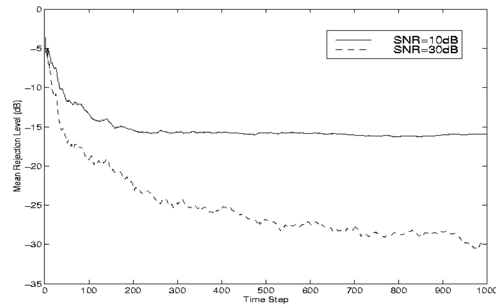

Figure 2. Performances of Algorithm 1.

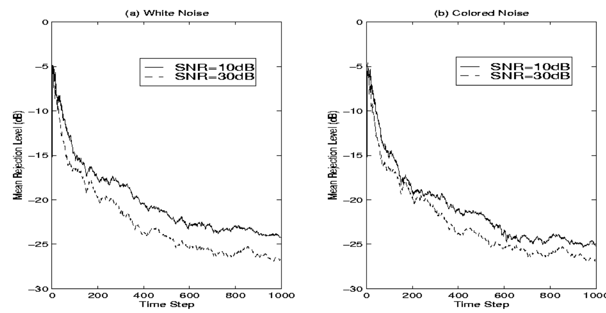

Figure 3. Performances of Algorithm 2: (a) White noise; (b) Colored noise.

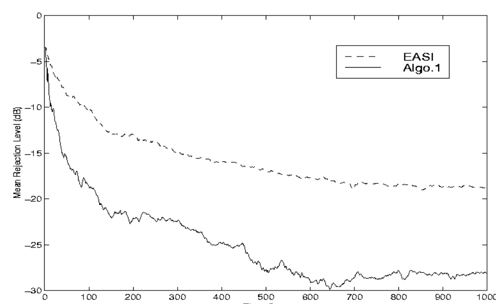

Figure 4. Performance comparison of Algorithm 1 and EASI algorithm.

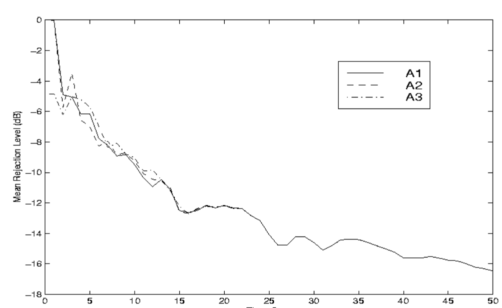

Figure 5. Uniform performance of Algorithm 1.