

Convergence analysis of a randomly perturbed infomax algorithm for blind source separation*

QI HE AND JACK XIN

We present a novel variation of the well-known infomax algorithm of blind source separation. Under natural gradient descent, the infomax algorithm converges to a stationary point of a limiting ordinary differential equation. However, due to the presence of saddle points or local minima of the corresponding likelihood function, the algorithm may be trapped around these “bad” stationary points for a long time, especially if the initial data are near them. To speed up convergence, we propose to add a sequence of random perturbations to the infomax algorithm to “shake” the iterating sequence so that it is “captured” by a path descending to a more stable stationary point. We analyze the convergence of the randomly perturbed algorithm, and illustrate its fast convergence through numerical examples on blind demixing of stochastic signals. The examples have analytical structures so that saddle points or local minima of the likelihood functions are explicit. The results may have implications for online learning algorithms in dissimilar problems.

AMS 2000 SUBJECT CLASSIFICATIONS: 34A05, 39A50, 65M12.

KEYWORDS AND PHRASES: Blind source separation, unstable equilibria, randomly perturbed infomax method.

1. Introduction

Blind source separation (BSS) aims at recovering a set of independent source signals from the observations of their mixtures without knowledge of mixing. It has been an active area of research in signal and image processing literature [1, 2, 9, 19] among others. For example, various algorithms have been developed based on minimizing mutual information (MMI) [1], information maximization (infomax) [2] and maximum likelihood (ML) approach [3]. In the instantaneous linear mixture model of d observations of d signals, the observed signals x can be represented by $x = A s$, where A is a $d \times d$ invertible mixing matrix, $s = [s_1, \dots, s_d]^T$ is a source signal with mutually

*The work was partially supported by NSF ATD grants DMS-0911277 and DMS-1222507.

independent components. Assuming that the joint probability density function (pdf) of the source is known as $r(s) = \prod_{i=1}^d r_i(s_i)$, the infomax and ML approaches provide an estimator of A , by maximizing the likelihood function

$$L(A) = E[\log r(A^{-1}x)/|\det(A)|].$$

In practice, the pdf of the source signals may not be known, so hypothetical pdf's $q_i(\cdot)$ are used as substitutes of $r_i(\cdot)$. Letting $W = A^{-1}$, the likelihood function becomes

$$J(W) = E[\log q(Wx)] + \log |\det(W)|,$$

where $q(x) = (q_1(x_1), \dots, q_d(x_d))$. Experience has shown that maximizing this alternative likelihood function still produces good demixing matrices as long as the true and hypothetical pdf's do not differ too much. Let $Y = Wx = (y_1, \dots, y_d)^T$ denote the recovered source vector and $x(i)$ denote the i -th sample of the mixture signal x . The associated algorithm is given by

$$(1.1) \quad W(n+1) = W(n) + \nu(I - F(n))W(n),$$

where $F(n) = \frac{1}{L} \sum_{i=nL+1}^{(n+1)L} f(Y^n(i))Y^n(i)^T$, $Y^n(i) = W(n)x(i)$, $f(Y) = (f_1(y_1), \dots, f_d(y_d))^T$ and $f_j(u) = -q'_j(u)/q_j(u)$. We say that W is a demixing matrix if it is such that $WA = P\Lambda$, where P is a permutation matrix and Λ is an invertible diagonal matrix. An equilibrium W_{eq} of this learning rule satisfies the steady state equation:

$$(1.2) \quad E[f(Y)Y^T] - I = E[f(Wx)(Wx)^T] - I = 0,$$

where the expectation can be theoretically carried out with pdf function of the source signals, or approximated from data x . The left hand side of (1.2) is a function of W . We define the function:

$$(1.3) \quad g(W) := E[I - f(Wx)(Wx)^T],$$

then (1.2) is just $g(W) = 0$.

To ensure the convergence to a good demixing matrix, one usually examines the stability of the limiting demixing matrix. In the literature, there are many studies on the convergence properties of the algorithm near equilibria [4]. The stability condition in the neighborhood of an equilibrium is well studied, however, the analysis of global convergence is much less known. Recently, [11] constructed examples of likelihood functions from hypothetical pdfs mimicing true pdfs of the source signals, and found that the global

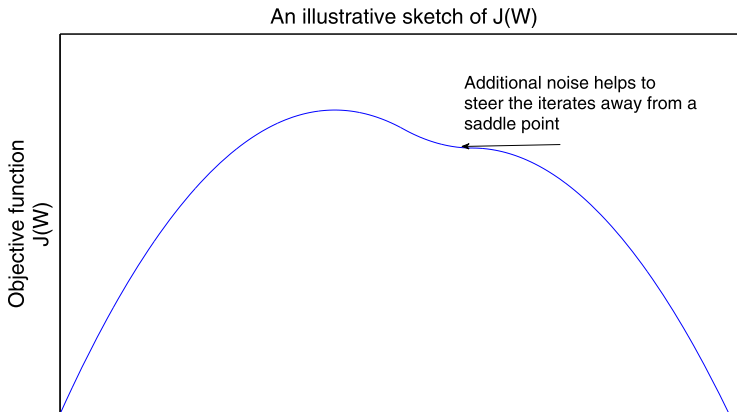


Figure 1.1: An illustration of the random perturbation method near a saddle point.

maximizers can be spurious in the sense that they do not separate the mixtures. Aside from this complication, even if the infomax algorithm converges to a separating solution under the stability condition, there may exist an unstable equilibrium to cause slow convergence. In [6], an explicit formula of equilibria is found for the two source separation problem when f is a cubic nonlinearity. An explicit formula of equilibria is given in section 3 for infomax algorithm under cubic nonlinearity with whiteness constraint. It shows that a set of saddle points always exists. Though the algorithm does not converge to these “bad” points, they tend to slow down the convergence dramatically. The reason is that the gradient descent as the driving force of the algorithm becomes very small when the iterates wander around these undesired equilibria points. To avoid such problem, one may increase the step size (learning rate) ν to help the algorithm leave these unstable points at the cost of introducing larger errors. To speed up convergence without sacrificing accuracy, some work has been done on variable step size [13]. However, this method is computationally more complex than our proposed remedy below of injecting random perturbations. Systematic expositions of random perturbation methods in the context of stochastic approximation theory can be found in [18]. The idea is to “shake the iterative sequence” until it is “captured” by a path descending to a more stable point, see Figure 1.1 for an illustration.

Random perturbations have been applied widely in optimization problems, such as multiple stochastic approximation [24], crude perturbations [12, 17], intermittent diffusion method [8] and its application in the shortest

path problem [7]; see also [15, 25] and application to image segmentation and restoration [26]. Our idea is similar to the stochastic annealing method, except that we are not looking for a global minimum. We will show that a suitable random perturbation of the basic algorithm does not alter its convergence property and equilibria, however it exhibits more robustness to initial data and fast convergence. The perturbed algorithm is found numerically to speed up convergence significantly in neighborhoods of unstable equilibria (e.g. saddle points) while maintaining the rate of convergence near stable equilibria.

The paper is organized as follows. Section 2 presents the mathematical analysis of the convergence of our randomly perturbed infomax algorithm. Section 3 gives analytical structure of equilibria of the algorithm under cubic nonlinearity. Section 4 shows some examples of the unstable equilibrium points of the algorithm. Section 5 investigates the performance of this new algorithm via numerical examples. Concluding remarks are given in section 6.

2. Analysis of convergence

We propose the following randomly perturbed infomax algorithm:

$$(2.1) \quad \begin{aligned} W(n+1) &= W(n) + \nu_n(I - F(n))W(n), n \neq \varepsilon_k, \text{ any } k, \\ W(n+1) &= W(n) + \nu_n(I - F(n))W(n) + b_k \chi_k, n = \varepsilon_k, \\ F(n) &= \frac{1}{L} \sum_{i=nL}^{(n+1)L} f(Y^n(i))Y^n(i)^T, \end{aligned}$$

where ν_n is a variable step size, ε_k is a sequence of integers going to infinity, $t_0 = 0$, $t_n = \sum_{k=0}^{n-1} \nu_k$, $T_k = t_{\varepsilon_{k+1}} - t_{\varepsilon_k}$, such that $T_k \rightarrow \infty$. Let $\{\chi_k\}$ be a sequence of independent identically distributed (i.i.d) random matrices in $\mathbb{R}^{d \times d}$ whose entries are uniformly distributed on the interval $[-1, 1]$, and let $\{b_k\}$ be a sequence of positive numbers tending to 0. We assume that the input $x(i)$ is a stationary stochastic process and that $f(\cdot)$ is a continuous nonlinear function, the case of most commonly used nonlinearities [9].

To analyze the convergence of this perturbed algorithm, we consider ν_n as a sequence satisfying $\sum_{k=1}^{\infty} \nu_k = \infty$ and $\sum_{k=1}^{\infty} \nu_k^2 < \infty$. We shall choose decreasing step size ν_k for the simplicity of convergence proof. In computation, we use constant step size ν . Without loss of generality, we also assume that $W(n)$ is uniformly bounded i.e. $|W(n)| \leq C$, for $n \geq 1$. This condition is natural in practice and can be achieved by adding proper

constraints in the algorithm, see [19, Lemma 4.1]. The boundness can also be achieved by truncation technique in stochastic approximation, see [18, Sec. 5.1]. Next we will apply stochastic approximation method to prove the convergence. Stochastic approximation(SA) method can be traced back to Robbins-Monro algorithm [20] and Kiefer-Wolfowitz algorithm [16]. It has been successfully applied to a variety of problems in the fields of numerical analysis [22], network systems [28], finance and insurance [14, 21, 27]. For more detailed and general setup of SA and its application, readers could refer to the book [18].

Let us define the associated interpolated process:

$$\begin{aligned} W^0(t) &= W(k), \quad \text{for } t_k < t < t_{k+1}, \\ W^n(t) &= W^0(t_n + t), \quad t \in (-\infty, \infty), \end{aligned}$$

where we define $t_0 = 0$, $t_n = \sum_{k=0}^{n-1} \nu_k$ and $m(t)$ to be the unique value of n such that $t_n \leq t < t_{n+1}$. Then $W^n(t)$ can be written as

$$\begin{aligned} W^n(t) &= W(n) + \sum_{i=n}^{m(t_n+t)-1} \nu_i Z_i \\ &+ \sum_{k:n \leq \varepsilon_k \leq m(t_n+t)} b_k \chi_k = W(n) + Z^n(t) + p^n(t), \end{aligned}$$

where

$$\begin{aligned} p^n(t) &= \sum_{k:n \leq \varepsilon_k \leq m(t_n+t)} b_k \chi_k, \\ Z_n &= (I - \frac{1}{L} \sum_{i=nL}^{(n+1)L} f(Y^n(i))Y^n(i)^T)W(n), \\ Z^0(t) &= \sum_{i=1}^{m(t)-1} \nu_i Z_i, \\ Z^n(t) &= Z^0(t_n + t) - Z^0(t_n) = \sum_{i=n}^{m(t_n+t)-1} \nu_i Z_i, t \geq 0. \end{aligned}$$

We have the following

Lemma 2.1. Denote $\delta M(n) = (I - \frac{1}{L} \sum_{i=nL+1}^{(n+1)L} f(Y^n(i))Y^n(i)^T)W(n) - g(W(n))W(n)$, then $\delta M(n)$ is a martingale difference, i.e., $E[\delta M_n|W(k)$,

$k \leq n] = 0$. The sequence Z_n has bounded second moment, i.e. $\sup_n E|Z_n|^2 < \infty$. The perturbation process $p^n(t)$ converges to zero.

Proof. Since $X(i)$ is a stationary process,

$$E\left[I - \frac{1}{L} \sum_{i=nL+1}^{(n+1)L} f(Y^n(i))Y^n(i)^T \mid W(k), k \leq n\right] = g(W(n)).$$

It follows that

$$\begin{aligned} & E[\delta M_n \mid W(k), k \leq n] \\ &= E\left[\left(I - \frac{1}{L} \sum_{i=nL+1}^{(n+1)L} f(Y^n(i))Y^n(i)^T\right)W(n) - g(W(n))W(n) \mid W(k), k \leq n\right] \\ &= W(n)E\left[\left(I - \frac{1}{L} \sum_{i=nL+1}^{(n+1)L} f(Y^n(i))Y^n(i)^T\right) - g(W(n)) \mid W(k), k \leq n\right] \\ &= 0. \end{aligned}$$

Hence, $\delta M(n)$ is a martingale difference. The sequence Z_n has bounded second moment follows from the fact that $W(n)$ is uniformly bounded. The perturbation sequence $p_n(t) \rightarrow 0$, since $T_k \rightarrow \infty$ and $b_k \rightarrow 0$. \square

With this lemma, we prove the following

Theorem 2.2. *There is a set N of probability zero such that for $\omega \notin N$, the set of functions $\{W^n(\omega, \cdot), n < \infty\}$ is equicontinuous. Let $W(\omega, \cdot)$ be the limit of a convergent subsequence. Then it satisfies the ordinary differential equation (ODE)*

$$(2.2) \quad \dot{W} = g(W)W.$$

The iterates $W_n(\omega)$ converge to the stationary set S of ODE (2.2). The set S is a union of finite disjoint compact subsets S_1, \dots, S_N . Moreover, $W_n(\omega)$ converge to a unique stationary point set S_i , consisting of stationary solutions of (1.2).

Proof. First, we show the equicontinuity of the interpolated process. Note

that

$$\begin{aligned}
 W^n(t) &= W(n) + \sum_{i=n}^{m(t+t_n)-1} \nu_i Z_i + p^n(t) \\
 &= W(n) + \sum_{i=n}^{m(t+t_n)-1} \nu_i g(W(i))W(i) \\
 (2.3) \quad &+ \delta M(i) + p^n(t) \\
 &= W(n) + \sum_{i=n}^{m(t+t_n)-1} \nu_i g(W(i))W(i) \\
 &+ \sum_{i=n}^{m(t+t_n)-1} \nu_i \delta M(i) + p^n(t)
 \end{aligned}$$

Since $W^n(\cdot)$ is defined as a piecewise constant function, (2.3) can be rewritten as

$$\begin{aligned}
 (2.4) \quad W^n(t) &= W(n) + \int_0^t g(W^n(s))W^n(s)ds \\
 &+ p^n(t) + M^n(t) + E^n(t),
 \end{aligned}$$

where $E^n(t)$ is the error due to the replacement of the first sum by an integral. Note that $E^n(t) = 0$ at time points $t = t_k - t_n, k > n$, at which the interpolated processes have jumps, and $E^n(t) \rightarrow 0$ uniformly in t as $n \rightarrow \infty$. In (2.4), $M^n(t)$ is defined similarly as $Z^n(t)$ by

$$\begin{aligned}
 M_n &= \sum_{i=n}^{m(t+t_n)-1} \nu_i \delta M(i), \\
 M^0(t) &= \sum_{i=1}^{m(t)-1} \nu_i M_i, \\
 M^k(t) &= M^0(t_k + t) - M^0(t_k) \\
 &= \sum_{i=k}^{m(t_k+t)-1} \nu_i M_i, t \geq 0.
 \end{aligned}$$

By the definition of M_n , it is a martingale sequence. In view of the properties

of martingale, for each $\Delta > 0$,

$$P\left(\sup_{n \geq j \geq m} |M_j - M_m| \geq \Delta\right) \leq \frac{E\left|\sum_{i=m}^{n-1} \nu_i \delta M(i)\right|^2}{\Delta^2}.$$

Note that the boundless of $W(n)$ and $\sum_{i=1}^{\infty} \nu^2$ and the fact that $E\delta M(i)\delta M^T(j) = 0$ for $i \neq j$, the right hand side of above expression is bounded, i.e. $E\left|\sum_{i=m}^{n-1} \nu_i \delta M(i)\right|^2 \leq K \sum_{i=1}^{\infty} \nu^2$. Thus, for each $\Delta > 0$,

$$(2.5) \quad \lim_m P\left(\sup_{j \geq m} |M_j - M_m| \geq \Delta\right) = 0.$$

By (2.5) and Lemma 2.1, there is a null set N such that for $\omega \notin N$, $E^n(\omega, \cdot)$ goes to zero uniformly on any bounded interval as $n \rightarrow \infty$, also $p^n(t) \rightarrow 0$ as $n \rightarrow \infty$. Let $\omega \notin N$, then the functions on the right hand side of (2.4) are equicontinuous in n with the limits of $M^n(\cdot)$ and $E^n(\cdot)$ being zero. By the Arzela-Ascoli Theorem, there is a convergent subsequence $\{W^{n_k}(\omega, \cdot)\}$ for $\omega \notin N$. We denote the limit by $W(\omega, \cdot)$. It is easily seen that the limit must satisfy the following integral equation:

$$W(\omega, t) = W(\omega, 0) + \int_0^t g(W(\omega, s))W(\omega, s)ds.$$

Define the cost function $G(W) = -J(W)$. By the derivation of the natural gradient [9], $g(W)$ can be written as the gradient of the objective function $J(W) = -G(W)$, i.e. $g(W) = -\partial G(W)/\partial W$, where $\partial G/\partial W$ is a $d \times d$ matrix whose entries are $\partial G/\partial w_{ij}$. Then the set of stationary points of (2.2) can be divided into disjoint compact and connected subsets $S_i, i = 0, \dots, N$, see [18, Sec 5.2]. The derivative of $G(W(\cdot))$ along the solution W of (2.2) is $-(\partial G/\partial W)^T(\partial G/\partial W) \leq 0$. Using $G(\cdot)$ as a Liapunov function, we can show that $W^n(t)$ must converge to some stationary point. It follows that $W^n(t)$ converge to a unique S_i , otherwise the resulting path would oscillate between distinct S_i 's, implying the existence of limit points outside of the stationary points. □

Remark 2.3. *Our proposed algorithm can also be applied to the algorithm under the whiteness constraint [5], i.e., $E[YY^T] = I$ and $E[f(Y)Y^T - Yf(Y)^T] = 0$. In this case, we just need to replace $F(n)$ in (2.1) by $F(n) = \frac{1}{L} \sum_{i=nL+1}^{(n+1)L} (f(Y^n(i))Y^n(i)^T - Y^n(i)f(Y^n(i))^T + Y^n(i)Y^n(i)^T)$. The convergence proof remains the same.*

3. Analytical structure of equilibria under cubic nonlinearity

In this section, we give analytical structure for the equilibria of the algorithm for separation of two sources under cubic nonlinearity. We know that an equilibrium of the infomax algorithm is a solution of $E[I - f(Y)^T f(Y)] = 0$, or a solution of $E[I - Y Y^T + f(Y) Y^T - Y f(Y)^T] = 0$ under whiteness constraint. The stability conditions of W are given in [4] by

$$(3.1) \quad \begin{aligned} 1 + \kappa_i &> 0, \quad \text{for } 1 \leq i \leq r \\ (1 + \kappa_i)(1 + \kappa_j) &> 1, \quad \text{for } 1 \leq i < j \leq n, \end{aligned}$$

or under whiteness constraint

$$(3.2) \quad \kappa_i + \kappa_j > 0, \quad \text{for } 1 \leq i < j \leq r,$$

where $\kappa_i = E[f'_i(y_i)] E[y_i^2] - E[f_i(y_i) y_i]$. First, let us consider separation of two independent sources by infomax algorithm under cubic nonlinearity. The direct method is to solve the equilibrium equation of W ,

$$E[I - f(Y)Y^T] = 0,$$

where $Y = WAs$ and $f(x) = x^3$. Let us set $V = WA$, and so $Y = Vs$, where s is the source signal vector.

3.1. Infomax algorithm

The equilibrium equation is

$$I - E(f(Y)Y^T) = 0,$$

where $Y = Vs = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} v_{11}s_1 + v_{12}s_2 \\ v_{21}s_1 + v_{22}s_2 \end{pmatrix}$. Simple calculations lead to a system of four equations:

$$\left\{ \begin{aligned} 1 - E[y_1^4] &= 1 - (v_1^4 \mu_1^4 + 2v_{11}^2 v_{12}^2 m + v_{12}^4 \mu_2^4) = 0 \\ E[y_1^3 y_2] &= v_{11} v_{21} (v_{11}^2 \mu_1^4 + v_{12}^2 m) \\ &\quad + v_{12} v_{22} (v_{12}^2 \mu_2^4 + v_{11}^2 m) = 0 \\ E[y_2^3 y_1] &= v_{11} v_{21} (v_{21}^2 \mu_1^4 + v_{22}^2 m) \\ &\quad + v_{12} v_{22} (v_{22}^2 \mu_2^4 + v_{21}^2 m) = 0 \\ 1 - E[y_2^4] &= 1 - (v_{21}^4 \mu_1^4 + 2v_{11}^2 v_{22}^2 m + v_{22}^4 \mu_2^4) = 0, \end{aligned} \right.$$

where $\mu_i^p = E[s_i^p]$ and $m = 3\mu_1^2\mu_2^2$. Cheung and Xu [6] give the mathematical analysis of the equilibrium equation, and found 16 solutions with explicit expressions as follows:

Solutions *A1* – *A8*:

$$V = \begin{bmatrix} \pm(\mu_1^4)^{-1/4} & 0 \\ 0 & \pm(\mu_2^4)^{-1/4} \end{bmatrix},$$

or

$$V = \begin{bmatrix} 0 & \pm(\mu_1^4)^{-1/4} \\ \pm(\mu_2^4)^{-1/4} & 0 \end{bmatrix}.$$

Solutions *B1* – *B8*:

$$V = \begin{bmatrix} s_{11}(-2c_1\eta_1)^{-1/4} & s_{12}(-2c_1\eta_2)^{-1/4} \\ s_{21}(-2c_2\eta_1)^{-1/4} & s_{22}(-2c_1\eta_2)^{-1/4} \end{bmatrix},$$

where $s_{ij} = 1$ or -1 , satisfying $s_{11}s_{12}s_{21}s_{22} = -1$, and $\eta_1 = \mu_1^4 + \sqrt{\mu_1^4/\mu_2^4}m$, $\eta_2 = \mu_2^4 + \sqrt{\mu_2^4/\mu_1^4}m$. Cheung and Xu [6] also proved that for sources satisfying:

$$\mu_1^4\mu_2^4 - [3(\mu_1^2)^2][3(\mu_2^2)^2] < 0,$$

solutions *A1* – *A8* are local minima and solutions *B1* – *B8* are saddle points of the limiting equation.

For source signals satisfying

$$\mu_1^4\mu_2^4 - [3(\mu_1^2)^2][3(\mu_2^2)^2] > 0,$$

solutions *B1* – *B8* are local minima and solutions *A1* – *A8* are saddle points of the limiting equation (2.2).

These results are consistent with the stability conditions (3.1), which show that the cubic nonlinearity is good for separation of sub-Gaussian signals. Under cubic nonlinearity, the analytic structure of the solutions of the equilibrium equation can also be given for infomax algorithm under whiteness condition. We discuss this case below.

3.2. Infomax algorithm under whiteness constraint

For simplicity, let us assume that the source signals have unit power, $E[ss^T] = I$. The equilibrium equation is

$$E[I - f(Y)Y^T + Yf(Y)^T - YY^T] = 0,$$

where $f(x) = x^3$ and

$$Y = Vs = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} v_{11}s_1 + v_{12}s_2 \\ v_{21}s_1 + v_{22}s_2 \end{pmatrix}.$$

Simplifying the above equation gives us a system of four equations:

$$(3.3) \quad 1 - E[y_1^2] = 1 - (v_{11}^2\mu_1^2 + v_{12}^2\mu_2^2) = 0,$$

$$(3.4) \quad \begin{aligned} & E[-y_1^3y_2 + y_1y_2^3 - y_1y_2] \\ & = v_{11}v_{21}[(v_{21}^2 - v_{11})\mu_1^4 + (v_{22}^2 - v_{12})m + \mu_1^4] \\ & + v_{12}v_{22}[(v_{22}^2 - v_{12})\mu_1^4 + (v_{21}^2 - v_{11})m + \mu_2^4] = 0, \end{aligned}$$

$$(3.5) \quad \begin{aligned} & E[-y_1y_2^3 + y_1^3y_2 - y_1y_2] \\ & = v_{11}v_{21}[(v_{11}^2 - v_{21})\mu_1^4 + (v_{12}^2 - v_{22})m + \mu_1^4] \\ & + v_{12}v_{22}[(v_{12}^2 - v_{22})\mu_1^4 + (v_{11}^2 - v_{21})m + \mu_2^4] = 0, \end{aligned}$$

$$(3.6) \quad 1 - E[y_2^2] = 1 - (v_{21}^2\mu_1^2 + v_{22}^2\mu_2^2) = 0,$$

where $\mu_i^p = E[s_i^p]$ and $m = 3\mu_1^2\mu_2^2$. Simplifying the system of equations by substituting $\mu_1^2 = \mu_2^2 = 1$, and adding equations (3.4) and (3.5) give:

$$(3.7) \quad \begin{cases} v_{11}^2 + v_{12}^2 = 1 \\ v_{21}^2 + v_{22}^2 = 1 \\ v_{11}v_{21} + v_{12}v_{22} = 0. \end{cases}$$

Plugging (3.7) into (3.4), we get

$$(3.8) \quad (v_{11}^2 - v_{21}^2)[v_{11}v_{21}(\mu_1^4 - m) + v_{12}v_{22}(-\mu_2^4 + m)] = 0.$$

Similar to [6], (3.8) is factorized as:

$$(3.9) \quad (v_{11}^2 - v_{21}^2)(1 \quad -1) \begin{pmatrix} \mu_1^4 & m \\ m & \mu_2^4 \end{pmatrix} \begin{pmatrix} v_{11}v_{21} \\ v_{12}v_{22} \end{pmatrix} = 0,$$

which is solved below depending on whether $v_{11}^2 - v_{21}^2 = 0$ or $v_{11}^2 - v_{21}^2 \neq 0$.

Case A: $v_{11}^2 \neq v_{12}^2$.

In this case, (3.9) can be simplified as

$$(3.10) \quad (\mu_1^4 - m \quad m - \mu_2^4) \begin{pmatrix} v_{11}v_{21} \\ v_{12}v_{22} \end{pmatrix} = 0.$$

Direct calculations and conditions (3.7) imply

$$(2m - \mu_1^4 - \mu_2^4)v_{12}v_{22} = 0.$$

By virtue of the definition of v_i and m , it is easily seen that $2m - \mu_1^4 - \mu_2^4 < 0$. Hence, we have $v_{12}v_{22} = 0$. So we get $v_{12} = 0$ or $v_{22} = 0$. Plugging these into (3.7), we found a set of 8 solutions:

A1 – A8

$$V = \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix},$$

or

$$V = \begin{pmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{pmatrix}.$$

Case B: $v_{11}^2 \neq v_{12}$. In this case, we get $v_{11} = \pm v_{12}$. Plugging these into (3.7), we get solutions:

B1 – B8

$$V = \begin{pmatrix} s_{11}\sqrt{2}/2 & s_{11}\sqrt{2}/2 \\ s_{21}\sqrt{2}/2 & s_{22}\sqrt{2}/2 \end{pmatrix},$$

where $s_{ij} = \pm 1$ such that $s_{11}s_{12}s_{21}s_{22} = -1$.

In the following, we analyze the type of the equilibrium points. Analogous to (1.3), we define the equilibrium function as

$$\bar{g}(V) := I - f(Y)Y^T + Yf(Y)^T - YY^T.$$

Then the limiting ODE (2.2) can be written as

$$\dot{V} = \bar{g}(V)V.$$

The system of ODEs for $(v_{11}, v_{12}, v_{21}, v_{22})^T$ reads:

$$(3.11) \quad \begin{cases} \dot{v}_{11} = -v_{21}(v_{11}^2 - v_{21}^2)(v_{11}v_{21}(\mu_1^4 - m) \\ \quad + v_{12}v_{22}(-\mu_2^4 + m)) \\ \dot{v}_{12} = -v_{22}(v_{11}^2 - v_{21}^2)(v_{11}v_{21}(\mu_1^4 - m) \\ \quad + v_{12}v_{22}(-\mu_2^4 + m)) \\ \dot{v}_{21} = v_{11}(v_{11}^2 - v_{21}^2)(v_{11}v_{21}(\mu_1^4 - m) \\ \quad + v_{12}v_{22}(-\mu_2^4 + m)) \\ \dot{v}_{22} = v_{12}(v_{11}^2 - v_{21}^2)(v_{11}v_{21}(\mu_1^4 - m) \\ \quad + v_{12}v_{22}(-\mu_2^4 + m)). \end{cases}$$

Table 3.1: Eigenvalues of Jacobian Matrix at Equilibria for (3.11)

	λ_1	λ_2	λ_3	λ_4
Eigenvalues at A's	-2	-2	-2	$2m - \mu_1^4 - \mu_2^4$
Eigenvalues at B's	-2	-2	-2	$\mu_1^4 + \mu_2^4 - 2m$

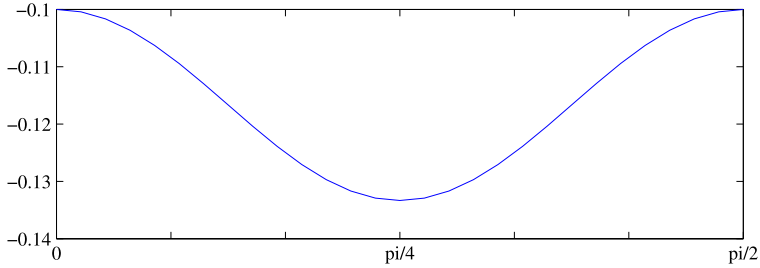


Figure 3.1: Objective function for algorithm under whiteness constraint.

We analyze the stability of the equilibria by calculating the eigenvalues of the Jacobian matrix of the right hand of (3.11) at the corresponding equilibria. The results are in Table 3.1.

We see that for signal satisfying $2m - \mu_1^4 - \mu_2^4 < 0$, the solutions $A1 - A8$ are local maxima and solutions $B1 - B8$ are saddle points. While for signal satisfying $2m - \mu_1^4 - \mu_2^4 > 0$, $A1 - A8$ are saddle points and $B1 - B8$ are local maxima. The results are consistent with the stability condition (3.2).

On the other hand, there is a geometric way to understand the equilibrium of infomax algorithm under the whiteness constraints. The transfer matrix WA is a rotation or reflection given by

$$WA = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \pm \sin(\theta) & \mp \cos(\theta) \end{pmatrix}.$$

The objective function can be written as

$$J(\theta) = E[\log q(y_1(\theta))] + E[\log q(y_2(\theta))].$$

The algorithm is separating successfully if it has local maxima only at $k\pi/2$. A sketch of the objective function is in Fig. 3.1. It can be seen that $\theta = 0, \pi/2$ are solutions corresponding to stable stationary points, while $\theta = \pi/4$ is a local minimum corresponding to an unstable stationary point.

4. Examples of unstable stationary points

In this section, we give examples of unstable equilibria. In the following examples, we consider 2 uniformly distributed source signals under cubic

Table 4.1: Equilibrium Points

	w_{11}	w_{12}	w_{21}	w_{22}
A1-A8 unstable equilibrium points	0	$15^{1/4}/2$	$15^{1/4}/2$	0
	$15^{1/4}/2$	0	0	$15^{1/4}/2$
	0	$15^{1/4}/2$	$-15^{1/4}/2$	0
	$15^{1/4}/2$	0	0	$-15^{1/4}/2$
	$-15^{1/4}/2$	0	0	$15^{1/4}/2$
	0	$-15^{1/4}/2$	$15^{1/4}/2$	0
	$-15^{1/4}/2$	0	0	$-15^{1/4}/2$
B1-B8 stable equilibrium points	0	$-15^{1/4}/2$	$-15^{1/4}/2$	0
	$-5^{1/4}/2$	$5^{1/4}/2$	$5^{1/4}/2$	$5^{1/4}/2$
	$5^{1/4}/2$	$-5^{1/4}/2$	$5^{1/4}/2$	$5^{1/4}/2$
	$5^{1/4}/2$	$5^{1/4}/2$	$-5^{1/4}/2$	$5^{1/4}/2$
	$5^{1/4}/2$	$5^{1/4}/2$	$5^{1/4}/2$	$-5^{1/4}/2$
	$-5^{1/4}/2$	$-5^{1/4}/2$	$-5^{1/4}/2$	$5^{1/4}/2$
	$5^{1/4}/2$	$-5^{1/4}/2$	$-5^{1/4}/2$	$-5^{1/4}/2$
	$-5^{1/4}/2$	$5^{1/4}/2$	$-5^{1/4}/2$	$-5^{1/4}/2$
	$-5^{1/4}/2$	$-5^{1/4}/2$	$5^{1/4}/2$	$-5^{1/4}/2$

nonlinearity. It is easily checked that the demixing matrix W is a stable equilibrium. However, there also exist unstable equilibria.

Example 1. Consider the two dimensional independent source $s = (s_1, s_2)^T$ drawn from uniform distribution on $[-1, 1]$ and the nonlinearity of the learning rule being $f(x) = x^3$. Let the mixing matrix be $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Denote the demixing matrix by $W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$. The recovered source is

$$\begin{aligned}
 Y = Wx = WAs &= \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \\
 &= \begin{pmatrix} (w_{11} - w_{12})s_1 + (w_{11} + w_{12})s_2 \\ (w_{21} - w_{22})s_1 + (w_{21} + w_{22})s_2 \end{pmatrix}.
 \end{aligned}$$

Plugging these expressions into equation (1.2) or

$$(4.1) \quad E \begin{bmatrix} 1 - y_1^4 & y_1^3 y_2 \\ y_2^3 y_1 & 1 - y_2^4 \end{bmatrix} = 0,$$

carrying out the expectations, one solves for W as in [6] or by using the results in section 3. There are 16 equilibria points listed explicitly in table 4.1. We see that there are two sets of equilibria. Set A consists of solutions which are unstable equilibria and are not separating matrices. Set B consists of

Table 4.2: Eigenvalues of Jacobian Matrix at Equilibria of Table 4.1

	λ_1	λ_2	λ_3	λ_4
Eigenvalues at A's	-4.000	-4.000	-1.500	0.500
Eigenvalues at B's	-4.000	-4.000	-2.667	-0.667

stable equilibria which are demixing matrices up to scaling and permutation. We also calculated the eigenvalues of the Jacobian from linearization of the ODE (2.2) of $(w_{11}, w_{12}, w_{21}, w_{22})^T$ at each equilibrium point. We observe that the equilibria in set A are saddle points since one of the four eigenvalues is positive. Note that equilibria A still attract the iterates in three directions except the fourth direction along which there is only a weak force to keep the iteration off. That is why the iterates can be trapped around set A for a long time.

Next, we give an example of the algorithm under whiteness constraint.

Example 2. Consider two source signals distributed uniformly on $[-\sqrt{3}, \sqrt{3}]$ with variance 1. Assume that the mixing matrix is $A = \begin{bmatrix} 1 & -4 \\ 1 & 1 \end{bmatrix}$. The nonlinearity for the learning rule is $f(x) = x^3$. Then the recovered source is

$$\begin{aligned}
 Y = Wx = WAs &= \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \cdot \begin{bmatrix} 1 & -4 \\ 1 & 1 \end{bmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \\
 &= \begin{pmatrix} (w_{11} + w_{12})s_1 + (-4w_{11} + w_{12})s_2 \\ (w_{21} + w_{22})s_1 + (-4w_{21} + w_{22})s_2 \end{pmatrix}.
 \end{aligned}$$

Plugging these expressions into equation (3.3)–(3.6) or

$$(4.2) \quad E \begin{bmatrix} 1 - y_1^2 & -y_1^3 y_2 + y_1 y_2^3 - y_1 y_2 \\ -y_1 y_2^3 + y_1^3 y_2 - y_1 y_2 & 1 - y_2^2 \end{bmatrix} = 0,$$

one can solve for W , or by using the results from section 3. There are 16 equilibria points listed explicitly in table 4.3.

We see that there are also two sets of equilibria. Set A consists of solutions which are unstable equilibria and are not separating matrices. Set B consists of stable equilibria which are demixing matrices up to scaling and permutation. Similar to table 4.2, we also calculated the eigenvalues of the Jacobian from linearization of the ODE (2.2) of $(w_{11}, w_{12}, w_{21}, w_{22})^T$ at each equilibrium point.

In this case, we observe that the equilibria in set A are saddle points since one of the four eigenvalues is positive. Note that equilibria A still attract the iterates in three directions except the fourth direction along which there

Table 4.3: Equilibrium Points

	w_{11}	w_{12}	w_{21}	w_{22}
A1-A8 unstable equilibrium points	0	$\sqrt{2}/2$	$\sqrt{2}/5$	$(3\sqrt{2})/10$
	$\sqrt{2}/5$	$(3\sqrt{2})/10$	0	$\sqrt{2}/2$
	$\sqrt{2}/5$	$(3\sqrt{2})/10$	0	$-\sqrt{2}/2$
	0	$-\sqrt{2}/2$	$\sqrt{2}/5$	$(3\sqrt{2})/10$
	$-\sqrt{2}/5$	$-(3\sqrt{2})/10$	0	$\sqrt{2}/2$
	0	$\sqrt{2}/2$	$-\sqrt{2}/5$	$-(3\sqrt{2})/10$
	0	$-\sqrt{2}/2$	$-\sqrt{2}/5$	$-(3\sqrt{2})/10$
	$-\sqrt{2}/5$	$-(3\sqrt{2})/10$	0	$-\sqrt{2}/2$
B1-B8 stable equilibrium points	1/5	4/5	1/5	-1/5
	1/5	4/5	-1/5	1/5
	1/5	-4/5	1/5	1/5
	-1/5	4/5	1/5	1/5
	1/5	-4/5	-1/5	-1/5
	-1/5	4/5	-1/5	-1/5
	-1/5	-4/5	1/5	-1/5
	-1/5	-4/5	-1/5	1/5

Table 4.4: Eigenvalues of Jacobian Matrix at Equilibria of Table 4.3

	λ_1	λ_2	λ_3	λ_4
Eigenvalues at A's	-2	-2	-2	2.4
Eigenvalues at B's	-2	-2	-2	-2

is only a weak force to keep the iteration off. So the iterates can be trapped around set A for a long time.

Example 3. In this example, we study blind separation of 3 signal sources. We shall see that saddle equilibria also exist for this higher dimension case. Since it is difficult to get explicit solutions of the equilibrium for higher dimensional cases, we consider a perturbative example. Suppose we have mixing matrix $\bar{A} = \begin{pmatrix} A & \bar{\varepsilon} \\ \bar{\varepsilon}^T & 1 \end{pmatrix}$, where $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, and $\bar{\varepsilon} = (\varepsilon, \varepsilon)^T$, for some small $\varepsilon > 0$. Assume that source signals s_1, s_2, s_3 are independent uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$. Then the inverse of \bar{A} should be of the following form.

Lemma 4.1.

$$\bar{A}^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon).$$

Proof. Write \bar{A} into the form $\bar{A} = \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & \bar{\varepsilon} \\ \bar{\varepsilon}^T & 0 \end{pmatrix}$, and apply Neuman series expansion. We have

$$\begin{aligned}
 \bar{A}^{-1} &= \begin{pmatrix} A & \bar{\varepsilon} \\ \bar{\varepsilon}^T & 0 \end{pmatrix}^{-1} \\
 &= \left[\begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & \bar{\varepsilon} \\ \bar{\varepsilon}^T & 0 \end{pmatrix} \right]^{-1} \\
 &= \left[\begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix} \left(I + \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & \bar{\varepsilon} \\ \bar{\varepsilon}^T & 0 \end{pmatrix} \right) \right]^{-1} \\
 &= \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}^{-1} \left[I - \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -\bar{\varepsilon} \\ -\bar{\varepsilon}^T & 0 \end{pmatrix} \right]^{-1} \\
 &= \begin{pmatrix} A^{-1} & 0 \\ 0 & 1 \end{pmatrix} \left[I + \begin{pmatrix} A^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -\bar{\varepsilon} \\ -\bar{\varepsilon}^T & 0 \end{pmatrix} \right. \\
 &\quad \left. + \left(\begin{pmatrix} A^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -\bar{\varepsilon} \\ -\bar{\varepsilon}^T & 0 \end{pmatrix} \right)^2 + o(\varepsilon^2) \right] \\
 &= \begin{pmatrix} A^{-1} & 0 \\ 0 & 1 \end{pmatrix} \left[I + \begin{pmatrix} 0 & -A^{-1}\bar{\varepsilon} \\ -\bar{\varepsilon}^T & 0 \end{pmatrix} \right. \\
 &\quad \left. + \begin{pmatrix} A^{-1}\bar{\varepsilon}\bar{\varepsilon}^T & 0 \\ 0 & \bar{\varepsilon}^T A^{-1}\bar{\varepsilon} \end{pmatrix} + o(\varepsilon^2) \right] \\
 &= \begin{pmatrix} A^{-1} & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -A^{-2}\bar{\varepsilon} \\ -\bar{\varepsilon}^T & 0 \end{pmatrix} \\
 &\quad + \begin{pmatrix} A^{-2}\bar{\varepsilon}\bar{\varepsilon}^T & 0 \\ 0 & \bar{\varepsilon}^T A^{-1}\bar{\varepsilon} \end{pmatrix} + o(\varepsilon^2).
 \end{aligned}$$

Due to the expression of \bar{A}^{-1} , we assume the demixing matrix $\bar{W} = \begin{pmatrix} W & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon)$, where $W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$. Then the recovered signal is

$$\begin{aligned}
 \bar{Y} &= \bar{W}\bar{A}\bar{s} \\
 &= \left(\begin{pmatrix} W & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon) \right) \left(\begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon) \right) \begin{pmatrix} s \\ s_3 \end{pmatrix} \\
 &= \begin{pmatrix} WA & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s \\ s_3 \end{pmatrix} \\
 &= \begin{pmatrix} WAs \\ s_3 \end{pmatrix} + O(\varepsilon) \\
 &= \begin{pmatrix} Y \\ s_3 \end{pmatrix} + O(\varepsilon),
 \end{aligned}$$

where $s = (s_1, s_2)^T$, $Y = WAs$ is as defined in Example 1.

The equilibria are the solutions of the following equation,

$$(4.3) \quad I - E(f(\bar{Y})\bar{Y}^T) = 0,$$

where $f(x) = x^3$. By Taylor expansions, $f(\bar{Y}) = f(\left(\begin{smallmatrix} Y \\ s_3 \end{smallmatrix}\right)) + O(\varepsilon)$. Plugging the above expression into (4.3) and equating terms without ε ,

$$I - E\left(\begin{pmatrix} f(Y) \\ f(s_3) \end{pmatrix} \left(\begin{matrix} Y^T & s_3 \end{matrix} \right)\right) = 0.$$

Simplifying the above equation,

$$I - E\left(\begin{matrix} f(Y)Y^T & f(Y)s_3 \\ f(s_3)Y^T & f(s_3)s_3 \end{matrix}\right) = 0.$$

By the independence of sources, we have

$$I - \begin{pmatrix} Ef(Y)Y^T & 0 \\ 0 & 1 \end{pmatrix} = 0,$$

which implies

$$I - Ef(Y)Y^T = 0.$$

This is exactly the equilibria equation (4.1) in Example 1. Hence, solutions \bar{W} of (4.3) can be written as

$$(4.4) \quad \bar{W} = \begin{pmatrix} W & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon),$$

where W are solutions of (4.1). Next, we examine the stability of these solutions \bar{W} . The limiting ODE of the algorithm is:

$$\dot{\bar{W}}(t) = (I - Ef(\bar{Y})\bar{Y}^T)\bar{W}(t).$$

Plugging (4.4) into the above equation and applying Taylor expansion of $f(x)$, we get

$$\begin{aligned} \begin{pmatrix} \dot{W} & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon) &= (I - (E\left(\begin{matrix} f(Y)Y^T & 0 \\ 0 & 1 \end{matrix}\right) + O(\varepsilon))) \\ &\cdot \left(\begin{pmatrix} W & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon)\right). \end{aligned}$$

By comparing the terms without ε , we get

$$\begin{pmatrix} \dot{W} & 0 \\ 0 & 1 \end{pmatrix} = (I - E \begin{pmatrix} f(Y)Y^T & 0 \\ 0 & 1 \end{pmatrix}) \begin{pmatrix} W & 0 \\ 0 & 1 \end{pmatrix},$$

which implies that

$$\dot{W}(t) = (I - Ef(Y)Y^T)W(t) =: \bar{g}(W(t)).$$

To check the stability of equilibria $\bar{W} = \begin{pmatrix} W & 0 \\ 0 & 1 \end{pmatrix} + O(\varepsilon)$, we can calculate Eigenvalues of Jacobian Matrix of $\bar{g}(\bar{W})$ at $\bar{W} = W + O(\varepsilon)$. Note that

$$\frac{\partial \bar{g}}{\partial W}(W + O(\varepsilon)) = \frac{\partial \bar{g}}{\partial W}(W) + O(\varepsilon),$$

by the continuity of \bar{g} . Hence, $\bar{\lambda}_{\bar{W}} = \lambda_W + O(\varepsilon)$ where $\bar{\lambda}_{\bar{W}}$ is the eigenvalue of Jacobian Matrix of $\bar{g}(W)$ at \bar{W} , λ is the eigenvalue of Jacobian Matrix of $g(W)$ at W , see Table 4.2. So, for small number $\varepsilon > 0$ there is a set of saddle points which is a small perturbations of unstable solution set A in Example 1.

5. Numerical simulations

From the above section, we see that there are unstable equilibria which are saddle points or local minimizers of J . In the following, we simulate the original algorithm to see how it performs around these unstable equilibria. By comparing with the unperturbed algorithm, we find that random perturbations improve the convergence very well. We shall measure the performance by inter-channel interference index (ICI) defined as

$$ICI_k = \left(\sum_{i=1}^n \sum_{j=1}^n \frac{|w_{ij}|^2}{\max_l |w_{il}(k)} \right) - n.$$

Example 1. To simulate Example 1 of section 3, we set step size $\nu = 0.001$ and $L = 10$. Let $\varepsilon_k = k(k - 1)/2 + 1$, $b_k = 1/k^2$ and let χ_k be 2×2 random matrices whose elements are drawn independently from uniform distribution on $[-1, 1]$. Let the algorithm start from identity matrix I_2 , which is common for most algorithms. However, this is not a good initial value in this example, since I_2 is close to the saddle points in table 4.1. Figure 5.1 shows the ICI index after 4000 iterations. The original algorithm does not appear

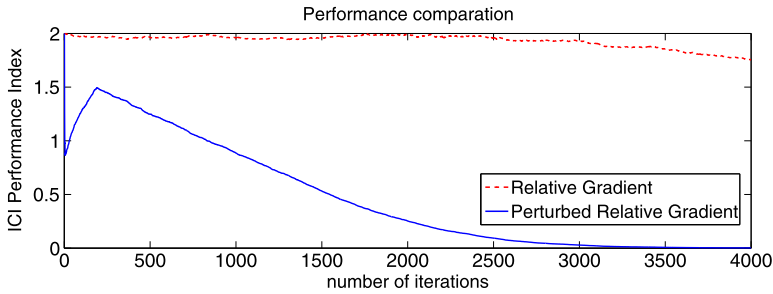


Figure 5.1: Comparison of the original and randomly perturbed infomax algorithms built from relative (natural) gradients.

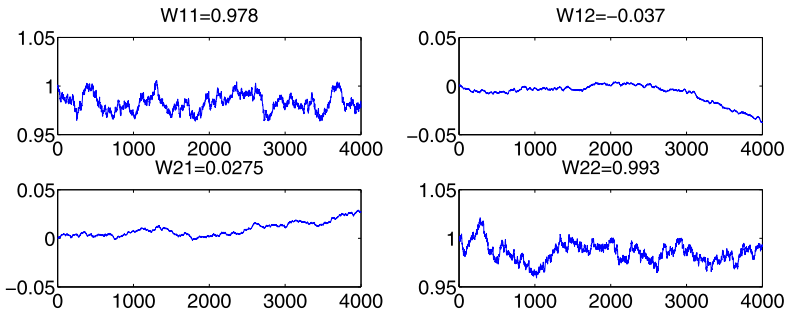
to converge in 4000 iterations (or very slow convergence). The randomly perturbed algorithm converges more rapidly after about 500 iterations. Figure 5.2 shows the convergence path for each component of W under the original algorithm and the randomly perturbed algorithm respectively.

Example 2. We simulate Example 2 of section 3 with the infomax algorithms under whiteness constraint. The step size $\nu = 0.002$, $L = 10$ and the mixing matrix $A = \begin{bmatrix} 1 & -4 \\ 1 & 1 \end{bmatrix}$. We use perturbations $\varepsilon_k = k(k-1)/2 + 1$, $b_k = 1/k$ and χ_k the 2×2 random matrices whose elements are uniformly distributed on $[-1, 1]$. The initial value $W(1) = I_2$. Figure 5.3 shows the comparison of the ICI index of the original and the randomly perturbed algorithms under whiteness constraint. The original algorithm almost does not converge, while the randomly perturbed one achieves convergence immediately after about 50 iterations.

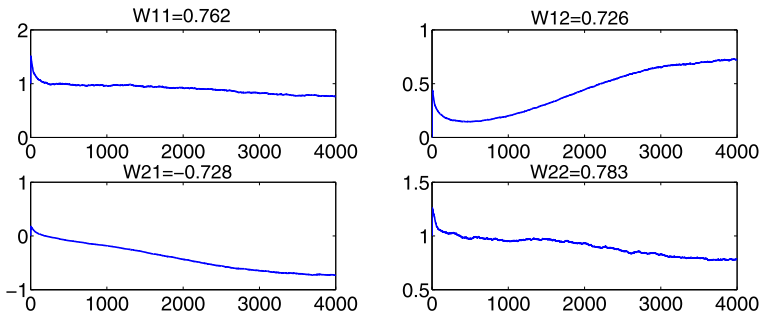
Example 3. In this example, we consider three source signals distributed uniformly on $[-\sqrt{3}, \sqrt{3}]$ with variance 1. The mixing matrix

$$A = \begin{pmatrix} 1 & 1 & 0.1 \\ -1 & 1 & 0.2 \\ 0.1 & 0.2 & 1 \end{pmatrix}.$$

We use the similar perturbations $\varepsilon_k = k(k-1)/2 + 1$, $b_k = 1/k$, and χ_k being the 3×3 random matrices whose elements are uniformly distributed on $[-1, 1]$. Set the initial value $W(1) = I_3$. By analysis in Section 4, saddle equilibria exists. Figure 5.5 shows the comparison of the ICI index of the original and the randomly perturbed algorithms. The randomly perturbed algorithm approaches convergence much faster than the original algorithm slowed down by saddle equilibria.



(a)



(b)

Figure 5.2: Convergence paths for demixing matrix W under infomax algorithm (a) and randomly perturbed infomax algorithm (b) built from relative (natural) gradients.

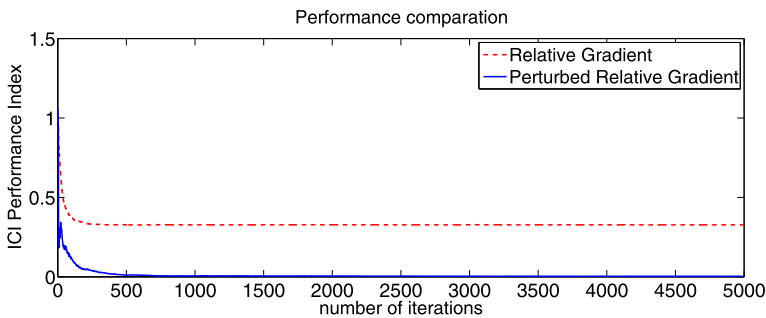


Figure 5.3: Comparison of the original (relative/natural gradient) and the randomly perturbed infomax algorithms under whiteness constraint.

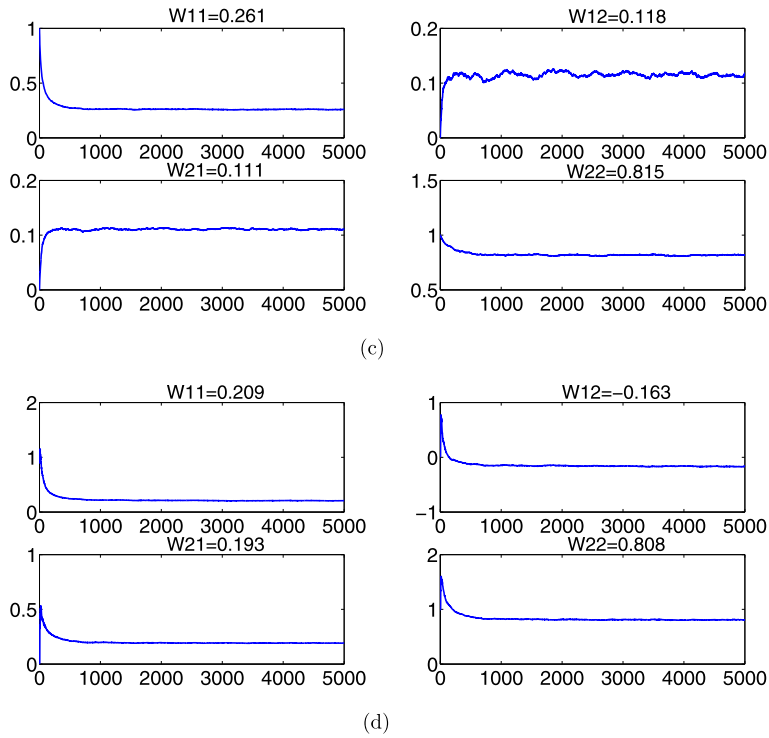


Figure 5.4: Convergence path for $W(n)$ under whiteness constraint (c) and convergence path of perturbed algorithm under whiteness algorithm (d).

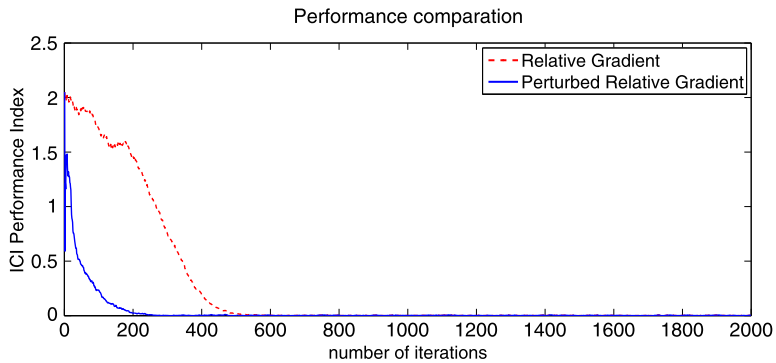


Figure 5.5: Comparison of the original (relative/natural gradient) and the randomly perturbed infomax algorithms in case of three source signals.

Conclusions

We showed that the existence of unstable equilibria may slow down the convergence of infomax learning algorithm tremendously. Adding suitable random perturbations, we proposed the randomly perturbed infomax algorithm and provided the mathematical analysis of its convergence. We showed by numerical simulations that this algorithm achieves fast convergence even when starting around unstable equilibria. The selection of noise types for fast convergence, the rate of convergence and the extension of our work to convolutive mixtures [10, 23] will be left for future research.

References

- [1] S. Amari, A. Cichocki, and H. Yang, *A new learning algorithm for blind signal separation*, in *Advances in Neural Information Processing Systems*, MIT Press, 1996, pp. 757–763.
- [2] A. Bell and T. Sejnowski, *An information-maximization approach to blind separation and blind deconvolution*, *Neural Computation*, 7, 1995, pp. 1129–1159.
- [3] J. F. Cardoso, *Blind signal separation: Statistical principles*, in *Proceedings of the IEEE*, 86, 1998, pp. 2009–2025.
- [4] J. F. Cardoso, *On the stability of source separation algorithms*, *The Journal of VLSI Signal Processing*, 26, 2000, pp. 7–14.
- [5] J. F. Cardoso and B. H. Laheld, *Equivariant adaptive source separation*, *IEEE Transactions on Signal Processing*, 44, 1996, pp. 3017–3030.
- [6] C. C. Cheung and L. Xu, *Separation of two independent sources by the information-theoretic approach with cubic nonlinearity*, in *International Conference on Neural Networks*, 4, 1997, pp. 2239–2244.
- [7] S. Chow, J. Lu, and H. Zhou, *Finding the shortest path by evolving junctions on obstacle boundaries (e-job): An initial value ODE’s approach*, *Applied and Computational Harmonic Analysis*, 35, 2013, pp. 165–176. [MR3053752](#)
- [8] S. Chow, T. Yang, and H. Zhou, *Global optimizations by intermittent diffusion*, *National Science Council Tunghai University Endowment Fund for Academic Advancement Mathematics Research Promotion Center*, 2009, pp. 121–134.

- [9] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley and Sons, Ltd., 2002.
- [10] S. C. Douglas and M. Gupta, *Scaled natural gradient algorithms for instantaneous and convolutive blind source separation*, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2, 2007, pp. II–637.
- [11] F. Ge and J. Ma, *Spurious solution of the maximum likelihood approach to ICA*, Signal Processing Letters, IEEE, 17, 2010, pp. 655–658.
- [12] S. B. Gelfand and S. K. Mitter, *Recursive stochastic algorithms for global optimization in R^d* , SIAM Journal on Control and Optimization, 29, 1991, pp. 999–1018. [MR1110084](#)
- [13] T. P. von Hoff, *On the convergence of blind separation and deconvolution*, PhD thesis, ETH, 2000.
- [14] Z. Jin, G. Yin, and C. Zhu, *Numerical solutions of optimal risk control and dividend optimization policies under a generalized singular control formulation*, Automatica, 48, 2012, pp. 1489–1501. [MR2950398](#)
- [15] R. Khas'minskii, *Application of random noise to optimization and recognition problems*, Problemy Peredachi Informatsii, 1, 1965, pp. 113–117.
- [16] J. Kiefer and J. Wolfowitz, *Stochastic estimation of the maximum of a regression function*, The Annals of Mathematical Statistics, 23, 1952, pp. 462–466. [MR0050243](#)
- [17] H. J. Kushner, *Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: Global minimization via Monte Carlo*, SIAM Journal on Applied Mathematics, 47, 1987, pp. 169–185. [MR0873242](#)
- [18] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, vol. 35, Springer, 1997. [MR1453116](#)
- [19] J. Liu, J. Xin, and Y. Qi, *A soft-constrained dynamic iterative method of blind source separation*, Multiscale Modeling & Simulation, 7, 2009, pp. 1795–1810. [MR2539199](#)
- [20] H. Robbins and S. Monro, *A stochastic approximation method*, The Annals of Mathematical Statistics, 1951, pp. 400–407. [MR0042668](#)
- [21] Q. Song and G. Yin, *Rates of convergence of numerical methods for controlled regime-switching diffusions with stopping times in the costs*,

SIAM Journal on Control and Optimization, 48, 2009, pp. 1831–1857.
[MR2516190](#)

- [22] Q. Song, G. Yin, and Z. Zhang, *Numerical solutions for stochastic differential games with regime switching*, Automatic Control, IEEE Transactions on, 53, 2008, pp. 509–521. [MR2394388](#)
- [23] J. Xin, M. Yu, Y.-Y. Qi, H.-I. Yang, and F.-G. Zeng, *A nonlocally weighted soft-constrained natural gradient algorithm for blind separation of reverberant speech*, in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2009, pp. 81–84.
- [24] S. Yakowitz, *A globally convergent stochastic approximation*, SIAM Journal on Control and Optimization, 31, 1993, pp. 30–40. [MR1200220](#)
- [25] G. Yin, *Rates of convergence for a class of global stochastic optimization algorithms*, SIAM Journal on Optimization, 10, 1999, pp. 99–120. [MR1722131](#)
- [26] G. Yin and P. Kelly, *Convergence rates of digital diffusion network algorithms for global optimization with applications to image estimation*, Journal of Global Optimization, 23, 2002, pp. 329–358. [MR1923050](#)
- [27] G. Yin, Q. Song, and H. Yang, *Stochastic optimization algorithms for barrier dividend strategies*, Journal of Computational and Applied Mathematics, 223, 2009, pp. 240–262. [MR2463114](#)
- [28] G. Yin, Q. Yuan, and L. Y. Wang, *Asynchronous stochastic approximation algorithms for networked systems: Regime-switching topologies and multi-scale structure*, Multiscale Modeling and Simulation, 11, 2013, pp. 813–839. [MR3090644](#)

QI HE

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA AT IRVINE
IRVINE, CA 92697
USA
E-mail address: qhe@math.uci.edu

JACK XIN

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA AT IRVINE
IRVINE, CA 92697
USA
E-mail address: jxin@math.uci.edu

RECEIVED JANUARY 30, 2014