Auxiliary-function-based Independent Component Analysis for Super-Gaussian Sources

Nobutaka Ono and Shigeki Miyabe

Graduate School of Information Science and Technology,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
{onono,miyabe}@hil.t.u-tokyo.ac.jp

Abstract. This paper presents new algorithms of independent component analysis (ICA) for super-Gaussian sources based on auxiliary function technique. The algorithms consist of two alternative updates: 1) update of demixing matrix and 2) update of weighted covariance matrix, which include no tuning parameters such as step size. The monotonic decrease of the objective function at each update is guaranteed. The experimental results show that the derived algorithms are robust to nonstationary data and outliers, and the convergence is faster than natural-gradient-based algorithm.

Key words: independent component analysis, Infomax, blind source separation, auxiliary function, super-Gaussian

1 Introduction

Independent component analysis (ICA) is a powerful technique to find independent components from mixtures without mixing information, which has been widely used for blind source separation [1]. One of the popular algorithms of ICA is Infomax [2], where the demixing matrix is estimated iteratively by applying natural-gradient-based update [3]. However, like other kinds of gradient-based optimization, there is a tradeoff between the convergence of speed and stability. The larger step size would make the convergence faster, but it may leads diverge. For improving robustness, modification of the natural gradient algorithm has been investigated in [4].

In this paper, we derive another kind of iterative solution for optimization of Infomax-type objective function based on auxiliary function approach, which is a framework to find efficient iterative solution for nonlinear optimization problem. In signal processing field, it has been used in well-known EM algorithm, and recently applied to various kinds of optimization problems such as nonnegative matrix factorization (NMF) [5], multi pitch analysis [6], sinusoidal parameter estimation [7], music signal separation [8], source localization [9], etc. In this paper, for a class of contrast functions related with super-Gaussianity, efficient iterative update rules of ICA are derived, which have no tuning parameters such as step size, and the monotonic decrease of the objective function is guaranteed. The experimental comparisons with existing other methods are also shown.

2 Objective Function of Infomax

In this paper, we generally assume that all variables can take complex values. A real-valued case is easily obtained by just replacing Hermitian conjugate ^{h} by transpose ^{t}. In Infomax algorithm, the demixing matrix is estimated by minimizing the following objective function:

$$J(W) = \sum_{k=1}^{K} E[G(\boldsymbol{w}_k^h \boldsymbol{x})] - \log|\det W|, \tag{1}$$

where

$$W = (\boldsymbol{w}_1 \ \cdots \ \boldsymbol{w}_K)^h \tag{2}$$

is a demixing matrix, $\boldsymbol{x} = (x_1 \ \cdots \ x_K)^t$ is an observed random vector, $E[\cdot]$ denotes expectation, and $G(y)$ is called contrast function [2]. The goal of this paper is to derive efficient iterative algorithms to find W to minimize eq. (1).

3 Auxiliary Function of Contrast Function

3.1 Auxiliary Function Approach

In order to introduce auxiliary function approach, let's consider a general optimization problem to find a parameter vector $\boldsymbol{\theta} = \boldsymbol{\theta}^\dagger$ satisfying

$$\boldsymbol{\theta}^\dagger = \mathrm{argmin}_{\boldsymbol{\theta}} \ J(\boldsymbol{\theta}), \tag{3}$$

where $J(\boldsymbol{\theta})$ is an objective function.

In the auxiliary function approach, a function $Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is designed such that it satisfies

$$J(\boldsymbol{\theta}) = \min_{\tilde{\boldsymbol{\theta}}} \ Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}). \tag{4}$$

$Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is called an auxiliary function for $J(\boldsymbol{\theta})$, and $\tilde{\boldsymbol{\theta}}$ are called auxiliary variables. Then, instead of directly minimizing the objective function $J(\boldsymbol{\theta})$, the auxiliary function $Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is minimized in terms of $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$, alternatively, the variables being iteratively updated as

$$\tilde{\boldsymbol{\theta}}^{(i+1)} = \mathrm{argmin}_{\tilde{\boldsymbol{\theta}}} \ Q(\boldsymbol{\theta}^{(i)}, \tilde{\boldsymbol{\theta}}), \tag{5}$$

$$\boldsymbol{\theta}^{(i+1)} = \mathrm{argmin}_{\boldsymbol{\theta}} \ Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}^{(i+1)}), \tag{6}$$

where i denotes the iteration index. The monotonic decrease of $J(\boldsymbol{\theta})$ under the above updates is guaranteed [5–9].

Note that even if eq. (3) has no closed-form solutions, there could exist $Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ satisfying eq. (4) such that both eq. (5) and eq. (6) have closed-form solutions. In such situations, the auxiliary function approach gives us efficient iterative update rules. However, how to find appropriate $Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ is problem-dependent.

3.2 Auxiliary Function of Contrast Function

A good candidate of auxiliary functions is given by a quadratic function because it can be easily minimized. For exploiting a quadratic function as an auxiliary function, eq. (4) requires that an objective function has to grow up more slowly than a quadratic function when $|\boldsymbol{\theta}| \to \infty$. From this point, we found a class of contrast functions which a quadratic function is applicable for, which is tightly related with super-Gaussianity. We first begin with two definitions.

Definition 1. *A set of real-valued functions of a real- or complex- variable z, S_G, is defined as*

$$S_G = \{G(z)|\ G(z) = G_R(|z|)\} \tag{7}$$

where $G_R(r)$ is a continuous and differentiable function of a real variable r satisfying that $G'_R(r)/r$ is continuous everywhere and it is monotonically decreasing in $r \geq 0$.

Definition 2. *([10] pp. 60-61) If a random variable r has a probability density of the form $p(r) = e^{-G_R(r)}$ where $G_R(r)$ is an even function which is differentiable, except possibly at the origin, $G_R(r)$ is strictly increasing in R_+, and $G'_R(r)/r$ is strictly decreasing in R_+, then, we shall say that it is super-Gaussian.*

From definition 2, it is clear that $G(z) = G_R(|z|)$ is a member of S_G if the distribution $p(r) = e^{-G_R(r)}$ is super-Gaussian and $G'_R(r)/r$ is continuous at the origin. Note that several well-used contrast functions such as $G_1(r) = \frac{1}{a_1}\log\cosh(a_1 r)$ and $G_2(r) = -\frac{1}{a_2}e^{-a_2 r^2/2}$, where $1 \leq a_1 \leq 2$, $a_2 \simeq 1$ [1], and their polar-type contrast functions [11] are included in S_G. Based on this definition of S_G, an explicit auxiliary function for the objective function of Infomax is obtained by the following two theorems.

Theorem 1. *For any $G(z) = G_R(|z|) \in S_G$,*

$$G(z) \leq \frac{G'_R(|z_0|)}{2|z_0|}|z|^2 + \left(G_R(|z_0|) - \frac{|z_0|G'_R(|z_0|)}{2}\right) \tag{8}$$

holds for any z. The equality sign is satisfied if and only if $|z| = |z_0|$.

Proof. First, let $r = |z|$ and $r_0 = |z_0|$, and consider the following function:

$$F(r) = \frac{G'_R(r_0)}{2r_0}r^2 + \left(G_R(r_0) - \frac{r_0 G'_R(r_0)}{2}\right) - G_R(r). \tag{9}$$

Differentiating $F(r)$, we have

$$F'(r) = \frac{G'_R(r_0)}{r_0}r - G'_R(r) = r\left(\frac{G'_R(r_0)}{r_0} - \frac{G'_R(r)}{r}\right). \tag{10}$$

Note that $G'_R(r)/r$ monotonically decreases in $r > 0$ and $F'(r_0) = 0$. Then, $F(r)$ has the unique minimum value at $r = r_0$ since $F(r)$ is continuous everywhere, and $F(r_0) = 0$. Consequently, it is clear that eq. (8) holds for any z. ∎

Theorem 2. *For any* $G(z) = G_R(|z|) \in S_G$*, let*

$$Q(W, \tilde{W}) = \frac{1}{2}\sum_{k=1}^{K} \boldsymbol{w}_k^h V_k \boldsymbol{w}_k - \log|\det W| + R, \tag{11}$$

where $\tilde{W} = (\tilde{\boldsymbol{w}}_1 \ \cdots \tilde{\boldsymbol{w}}_K)^h$, $r_k = |\tilde{\boldsymbol{w}}_k^h \boldsymbol{x}|$,

$$V_k = E\left[\frac{G'(r_k)}{r_k}\boldsymbol{x}\boldsymbol{x}^h\right], \tag{12}$$

and R *is a constant independent of* W*. Then,*

$$J(W) \le Q(W, \tilde{W}) \tag{13}$$

holds for any W*. The equality sign holds if and only if* $\tilde{\boldsymbol{w}}_k = e^{j\phi_k}\boldsymbol{w}_k$ *for* $^\forall k$ *where* ϕ_k *denotes an arbitrary phase.*

Proof. Eq. (13) is directly obtained by substituting $z = \boldsymbol{w}_k^h \boldsymbol{x}$ and $z_0 = \tilde{\boldsymbol{w}}_k^h \boldsymbol{x}$ into eq. (8), summing them up from $k = 1$ to $k = K$, and taking expectation. ■

4 Derivation of Update Rules

4.1 Derivative of Auxiliary Function

Based on the principle of the auxiliary function approach, update rules should be obtained by minimizing $Q(W, \tilde{W})$ in terms of W and $\tilde{W}$, alternatively. From Theorem 2, the minimization of Q in terms of $\tilde{W}$ is easily obtained by just setting $\tilde{W} = W$. Then, let's focus on minimizing Q in terms of W. It can be done by solving $\partial Q(W, \tilde{W})/\partial \boldsymbol{w}_k^* = 0$ $(1 \le k \le K)$, where * denotes complex conjugate. They yield

$$\frac{1}{2}V_k \boldsymbol{w}_k - \frac{\partial}{\partial \boldsymbol{w}_k^*}\log|\det W| = 0 \quad (1 \le k \le K). \tag{14}$$

Rearranging eq. (14) with using $(\partial/\partial W)\det W = W^{-t}\det W$, the following simultaneous vector equations are derived.

$$\boldsymbol{w}_l^h V_k \boldsymbol{w}_k = \delta_{lk} \quad (1 \le k \le K,\ 1 \le l \le K). \tag{15}$$

For updating all of $\boldsymbol{w}_k$ simultaneously, it is necessary to solve eq. (15), and especially, a closed form solution is desirable for efficient updates. If all of V_k are commutable, their matrices would share eigenvectors and they would give the solutions of eq. (15). However, V_ks are generally not commutable. In that case, to the authors' knowledge, there are no closed-form solutions for eq. (15) except when $K = 2$. Instead of simultaneously updating all of $\boldsymbol{w}_k$, we here propose two kinds of different update rules: 1) sequentially updating each of $\boldsymbol{w}_k$, and 2) sequentially updating each pair of $\boldsymbol{w}_k$s, which can be both performed in closed-form manners.

4.2 AuxICA1: Sequential Update Rules

The first method updates each of $\boldsymbol{w}_k$ sequentially. Consider an update of $\boldsymbol{w}_k$ with keeping other $\boldsymbol{w}_l$s $(l \neq k)$ fixed. In this case, $\partial Q(W, \tilde{W})/\partial \boldsymbol{w}_k^* = 0$ yields

$$\boldsymbol{w}_k^h V_k \boldsymbol{w}_k = 1, \tag{16}$$

$$\boldsymbol{w}_l^h V_k \boldsymbol{w}_k = 0 \quad (l \neq k). \tag{17}$$

From eq. (17), $\boldsymbol{w}_k$ has to be orthogonal to all of $V_k \boldsymbol{w}_l$ $(l \neq k)$ at least. Such a vector can be obtained by projecting the estimate of $\boldsymbol{w}_k$ at the previous iteration into the complementary space of the space spanned by all of $V_k \boldsymbol{w}_l$ $(l \neq k)$. Then, the normalization should be performed to satisfy eq. (16). While, the auxiliary variables $\tilde{\boldsymbol{w}}_k$ are only included in V_k. Hence, updates of the auxiliary variables $\tilde{\boldsymbol{w}}_k$ are equivalent to calculating V_k. Consequently, The algorithm is summarized as follows.

AuxICA1: The following alternative updates are applied for all k in order until convergence.

Auxiliary variable updates: Calculate the following matrices:

$$V_k = E\left[\frac{G'(r_k)}{r_k} \boldsymbol{x}\boldsymbol{x}^h\right], \tag{18}$$

$$P = V_k(\boldsymbol{w}_1 \ \cdots \ \boldsymbol{w}_{k-1} \ \boldsymbol{w}_{k+1} \ \cdots \ \boldsymbol{w}_K), \tag{19}$$

where $r_k = |\boldsymbol{w}_k^h \boldsymbol{x}|$.

Demixing matrix updates: Apply the following updates in order.

$$\boldsymbol{w}_k \leftarrow \boldsymbol{w}_k - P(P^h P)^{-1} P^h \boldsymbol{w}_k, \tag{20}$$

$$\boldsymbol{w}_k \leftarrow \boldsymbol{w}_k / \sqrt{\boldsymbol{w}_k^h V_k \boldsymbol{w}_k}. \tag{21}$$

4.3 AuxICA2: Pairwise Update Rules

The second update rule is based on the closed form solution of eq. (15) in $K = 2$. When $K = 2$, eq. (15) indicates that both of $V_1 \boldsymbol{w}_1$ and $V_2 \boldsymbol{w}_1$ are orthogonal to $\boldsymbol{w}_2$. Because the direction orthogonal to $\boldsymbol{w}_2$ is uniquely determined in the two dimensional space, $V_1 \boldsymbol{w}_1$ and $V_2 \boldsymbol{w}_1$ have to be parallel such as

$$V_1 \boldsymbol{w}_1 = \gamma V_2 \boldsymbol{w}_1, \tag{22}$$

where γ is a constant. In the same way, $V_1 \boldsymbol{w}_2$ and $V_2 \boldsymbol{w}_2$ are also parallel. Such vectors are obtained as solutions of eq. (22), which is a generalized eigenvalue problem. If V_1 and V_2 are not singular, they are simply obtained as eigen vectors of $V_2^{-1} V_1$. Even if $K > 2$, it is possible to apply this pairwise updates for all possible pairs. Then, the algorithm is summarized as follows.

AuxICA2: The following alternative updates are applied for all pairs of m and n such that $m < n$ in order until convergence.

Auxiliary variable updates: Calculate the following covariance matrices:

$$U_m = E\left[\frac{G'_R(r_m)}{r_m}\boldsymbol{u}\boldsymbol{u}^h\right], \quad U_n = E\left[\frac{G'_R(r_n)}{r_n}\boldsymbol{u}\boldsymbol{u}^h\right], \tag{23}$$

where $r_m = |\boldsymbol{w}_m^h\boldsymbol{x}|$, $r_n = |\boldsymbol{w}_n^h\boldsymbol{x}|$, and $\boldsymbol{u} = (\boldsymbol{w}_m^h\boldsymbol{x}\ \boldsymbol{w}_n^h\boldsymbol{x})^t$.

Demixing matrix updates: Find two solutions $\boldsymbol{h}_m$ and $\boldsymbol{h}_n$ of the generalized eigenvalue problem of 2×2 matrix: $U_m\boldsymbol{h} = \gamma U_n\boldsymbol{h}$. Then, apply the following updates in order.

$$\boldsymbol{h}_m \leftarrow \boldsymbol{h}_m/\sqrt{\boldsymbol{h}_m^h U_m \boldsymbol{h}_m}, \qquad \boldsymbol{h}_n \leftarrow \boldsymbol{h}_n/\sqrt{\boldsymbol{h}_n^h U_n \boldsymbol{h}_n}, \tag{24}$$

$$(\boldsymbol{w}_m\ \boldsymbol{w}_n) \leftarrow (\boldsymbol{w}_m\ \boldsymbol{w}_n)(\boldsymbol{h}_m\ \boldsymbol{h}_n). \tag{25}$$

5 Experimental Evaluations

In order to evaluate the performance of estimating demixing matrix and the convergence speed of AuxICA1 and AuxICA2, three artificial sources were prepared here. All of them were complex-valued random process with independent amplitude and phase. The phase followed the uniform distribution from 0 to 2π, and the amplitude $a \geq 0$ followed $p_1(a) = e^{-a}$, $p_2(a) = (3/4)\delta(a) + (1/4)e^{-a}$, and $p_3(a) = \arctan a_0/\pi(1 + a^2)$ $(0 \leq a \leq a_0)$ where $a_0 = 1000$, by which we intended to simulate 1) stationary source, 2) nonstationary source (each source is silent with probability 3/4), and 3) spiky source including outliers, respectively. In each case, the numbers of sources are $K = 2$ or $K = 6$ and the data length is $N = 1000$. The observed mixtures were made with instantaneous mixing matrices, where each element was independently generated by complex-valued Gaussian random process with zero mean and unity variance.

For these mixtures, we compared the proposed algorithms with popular existing methods such as polar-type Infomax [11], scaled Infomax [4] and FastICA [12]. In all of the algorithms, $G(z) = \log\cosh|z|$ was used as a contrast function. The initial value of the demixing matrix was given by data whitening.

The performance was evaluated by averaged signal to noise ratio:

$$\mathrm{SNR} = \frac{1}{K}\sum_k 10\log_{10}\frac{\sum_t |s_k(t)|^2}{\sum_t |s_k(t) - y_k(t)|^2}, \tag{26}$$

which was calculated by using demixing matrix iteratively-estimated at each iteration, where the correct permutation was given and the appropriate scaling was estimated by projection back [13]. The experiments were performed in Matlab ver. 7.9 on a laptop PC with 2.66GHz CPU. Fig. 1 shows the relationship between actual computational time and resultant SNR obtained by averaging 100 trials. AuxICA1 or AuxICA2 shows the best convergence speed in most cases, and the obtained solutions are robust to nonstationary signal and outliers.

6 Conclusion

In this paper, we presented new algorithms of ICA for super-Gaussian sources based on auxiliary function technique. The experimental results showed that the derived algorithms give the faster convergence than natural-gradient-based Infomax, and are robust to nonstationary data and outliers. Applying the proposed algorithms to blind source separation for convolutive-mixture is ongoing.

References

1. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.
2. A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
3. S. Amari, A. Cichocki, and H. H. Yang, "A New Learning Algorithm for Blind Signal Separation," in *Advances in Neural Information Processing Systems*, D. Touretzky, M. Mozer, and M. Hasselmo, Eds., MIT Press, pp. 757–763, 1996.
4. S. C. Douglas and M. Gupta, "Scaled Natural Gradient Algorithms for Instantaneous and Convolutive Blind Source Separation," *Proc. ICASSP*, pp. 637–640, 2007.
5. D. D. Lee and H. S. Seung, "Algorithms for Non-Negative Matrix Factorization" *Proc. NIPS*, pp. 556–562, 2000.
6. H. Kameoka, T. Nishimoto, and S. Sagayama, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," *IEEE Trans. ASLP*, vol. 15, no. 3, pp.982-994, Mar. 2007.
7. H. Kameoka, N. Ono, and S. Sagayama, "Auxiliary Function Approach to Parameter Estimation of Constrained Sinusoidal Model," *Proc. ICASSP*, pp.29–32, Apr. 2008.
8. N. Ono, K. Miyamoto, H. Kameoka, and S. Sagayama, "A Real-time Equalizer of Harmonic and Percussive Components in Music Signals," *Proc. ISMIR*, pp. 139–144, Sep. 2008
9. N. Ono and S. Sagayama, "R-Means Localization: A Simple Iterative Algorithm for Source Localization Based on Time Difference of Arrival," *Proc. ICASSP*, pp. 2718–2721, Mar. 2010.
10. A. Benveniste, M. M'etivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*, Springer-Verlag, 1990.
11. H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar Coordinate Based Nonlinear Function for Frequency-domain Blind Source Separation," *IEICE Trans. Fundamentals*, vol. E86-A, no. 3, pp. 590–596, Mar. 2003.
12. E. Bingham and A. Hyvärinen, "A Fast Fixed-Point Algorithm for Independent Component Analysis of Complex Valued Signals," *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, Feb. 2000.
13. N. Murata, S. Ikeda, and A. Ziehe, "An Approach to Blind Source Separation Based on Temporal Structure of Speech Signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, 2001.

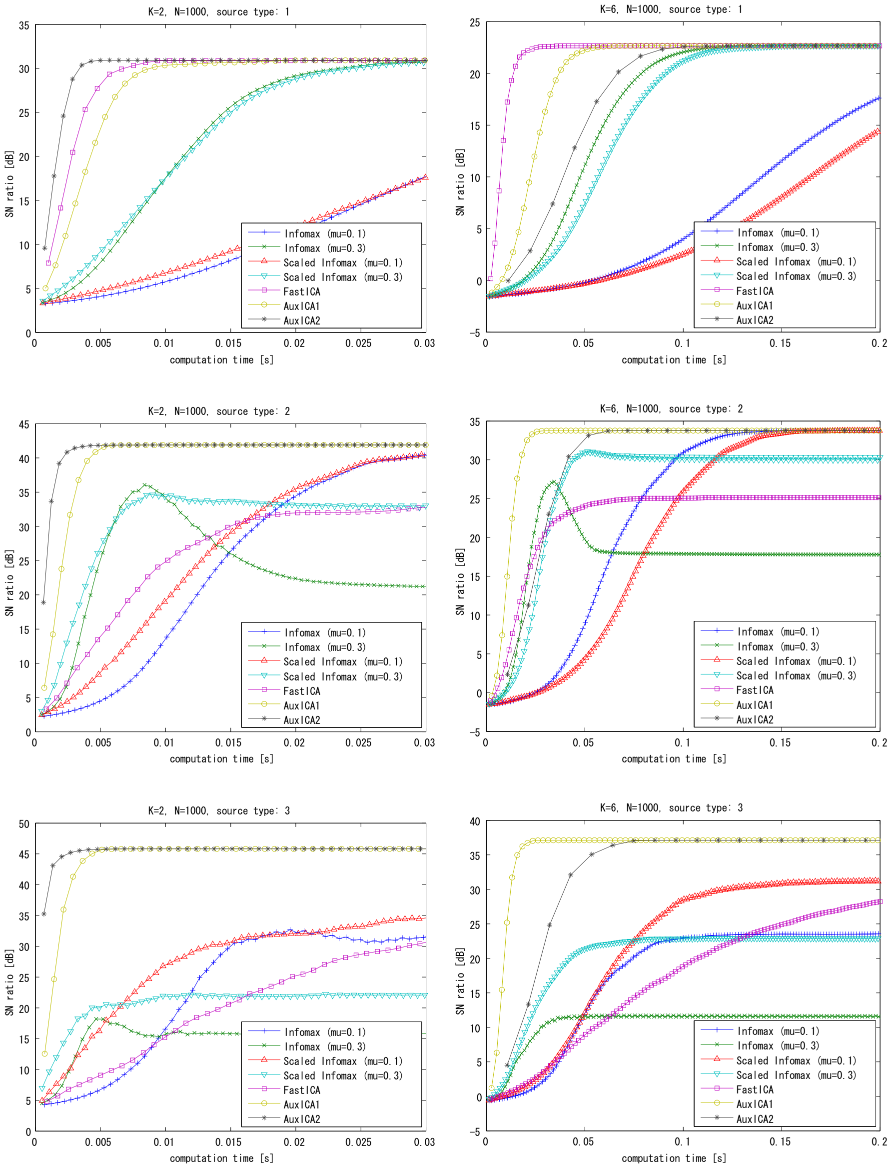

Fig. 1. The relationship between actual computational time and resultant SNR obtained by averaging 100 trials. The left and the right columns represent the results for $K = 2$ and $K = 6$, respectively, where K denotes the number of sources. The amplitude of sources follow $p_1(a)$, $p_2(a)$, and $p_3(a)$ from top to bottom, respectively. In the legend, "mu" denotes a step size. In conventional Infomax, there is a tradeoff between the convergence speed and stability. The scaled Infomax improves them, but the tradeoff still remains. FastICA gives the fast convergence for stationary signal, but it becomes rather slow for nonstationay signal or outliers. While, the proposed methods (AuxICA1 or AuxICA2) show the best convergence speed in most cases and robust to nonstationary signal and outliers.