Auxiliary function approach to independent component analysis and independent vector analysis

N. Ono

N. Ono, "Auxiliary function approach to independent component analysis and independent vector analysis," Proc. SPIE 9496, Independent Component Analyses, Compressive Sampling, Large Data Analyses (LDA), Neural Networks, Biosystems, and Nanoengineering XIII, 94960L (11 June 2015); doi: 10.1117/12.2179859

Event: SPIE Sensing Technology + Applications, 2015, Baltimore, Maryland, United States

Auxiliary function approach to independent component analysis and independent vector analysis

N. Ono*[a]
[a]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, JAPAN, 101-8430

ABSTRACT

In this paper, we review an auxiliary function approach to independent component analysis (ICA) and independent vector analysis (IVA). The derived algorithm consists of two alternative updates: 1) weighted covariance matrix update and 2) demixing matrix update, which include no tuning parameters such as a step size in the gradient descent method. The monotonic decrease of the objective function is guaranteed by the principle of the auxiliary function method. The experimental evaluation shows that the derived update rules yield faster convergence and better results than natural gradient updates. An efficient implementation on a mobile phone is also presented.

Keywords: Independent component analysis, independent vector analysis, blind source separation, auxiliary function, majorization minimization algorithm

1. INTRODUCTION

Independent component analysis (ICA) [1] is a technique for finding independent components from a mixture and it is a basis of blind source separation (BSS). Several standard algorithms such as InfoMax and FastICA have been already established and it has been applied to many kinds of signals as sound, image, bio-signals, and so on. However, the computational cost is still not small, especially for a convolutive mixture. In the frequency-domain approach for the convolutive mixture, demixing matrices are estimated at each frequency bin, and due to the larger number of the frequency bins (for example, 2049, 4097), reducing the computation time is still a major challenge in this field. Recently, the author derived a new effective algorithm of ICA by an auxiliary function approach, called AuxICA (auxiliary function-based independent component analysis) [2] and extended it to independent vector analysis (IVA) [3], implemented it on iPhone [4], and applied to user-guided source separation [5], real-time implementation [6], multi-channel non-negative matrix factorization [7] and model-based source separation [8]. In this paper, we review the basic principle, algorithm and results of source separation.

2. PROBLEM FORMULATION

Let us consider a multi-channel observation generated by a linear mixing process of multiple sources as shown in Fig. 1. The problem here is to estimate the original sources from only the observation without any knowledge of the mixing process. If the number of the observations and the number of the sources is equal, the problem is how to estimate a demixing matrix. Based on the assumption of independency between sources, ICA gives us a way to estimate it. The objective function of ICA in the InfoMax approach [9], which is one of the most standard and popular algorithms, is written as Fig. 2, where the function G(y) is called a contrast function. In InfoMax, the natural gradient method is applied for minimizing the objective function [10]. However, it includes a step size parameter and there is a trade-off between the speed and the stability of convergence depending on the step size.

3. AUXILIARY FUNCTION BASED INDEPENDENT COMPONENT ANALYSIS [2]

Let's consider a general optimization problem shown in Fig. 3. In the auxiliary function approach, which can be denoted as the majorization minimization (MM) approach, instead of directly minimizing the objective function, another function called auxiliary function is introduced. An auxiliary function has to be larger than the objective function or to be tangential to the objective function. If the auxiliary function can be minimized in a closed form, the alternative updates of parameters to be estimated and auxiliary variables can be an effective algorithm. However, how to find such a useful auxiliary function is problem-dependent. The contribution of this work is to find a useful auxiliary function for InfoMax-type ICA and derive an effective algorithm.

Independent Component Analyses, Compressive Sampling, Large Data Analyses (LDA), Neural Networks, Biosystems, and Nanoengineering XIII, edited by Harold H. Szu, Liyi Dai, Yufeng Zheng, Proc. of SPIE Vol. 9496, 94960L · © 2015 SPIE · CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2179859

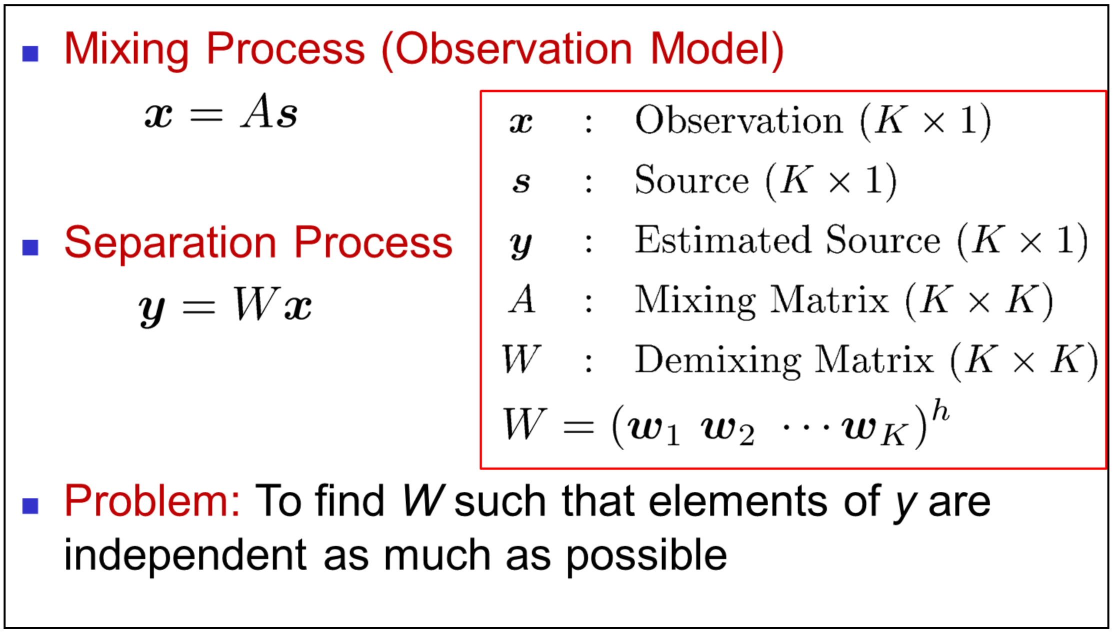


Figure 1. The problem formulation of independent component analysis

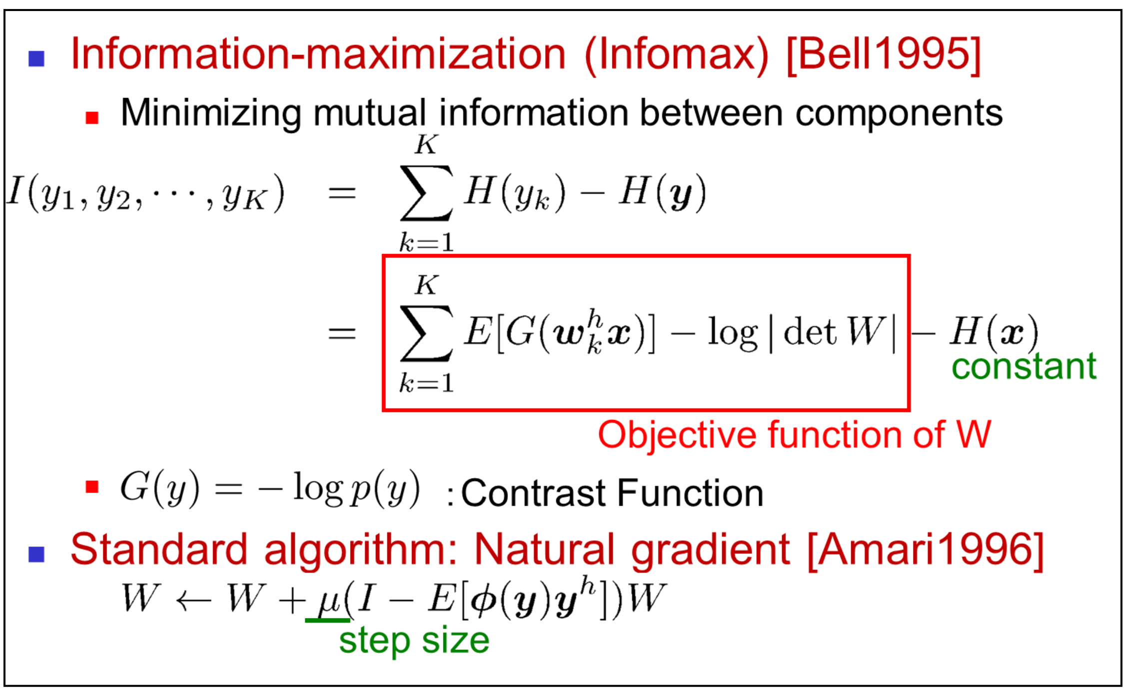


Figure 2. The objective function of ICA in the InfoMax approach and the natural gradient-based update rule

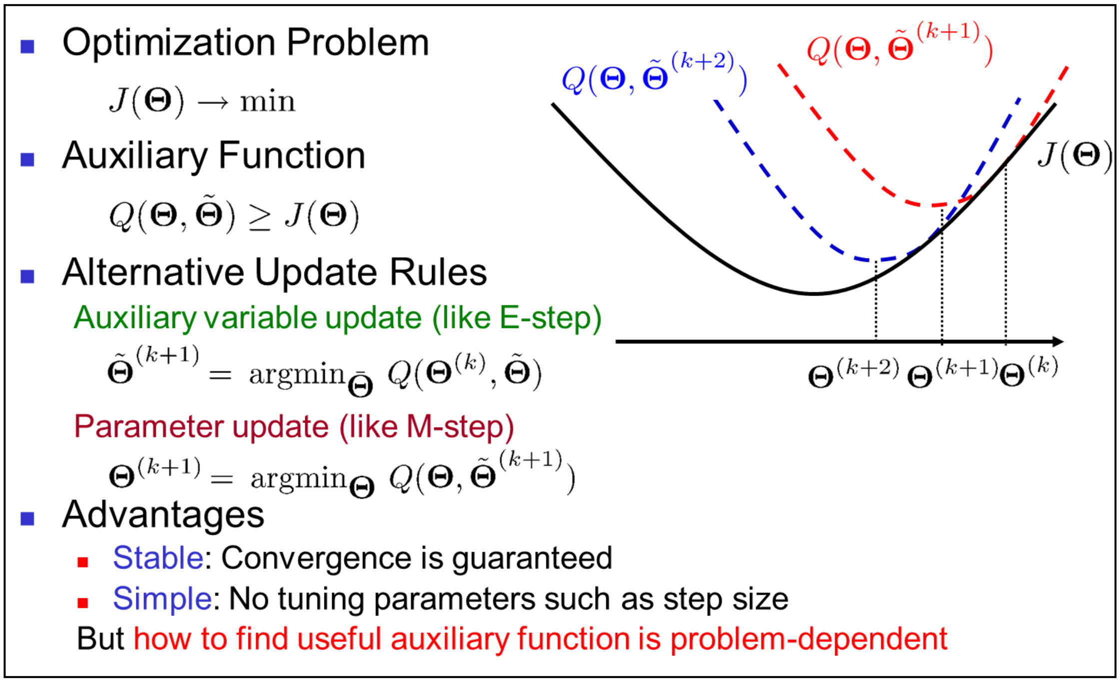


Figure 3. The auxiliary function approach for an optimization problem

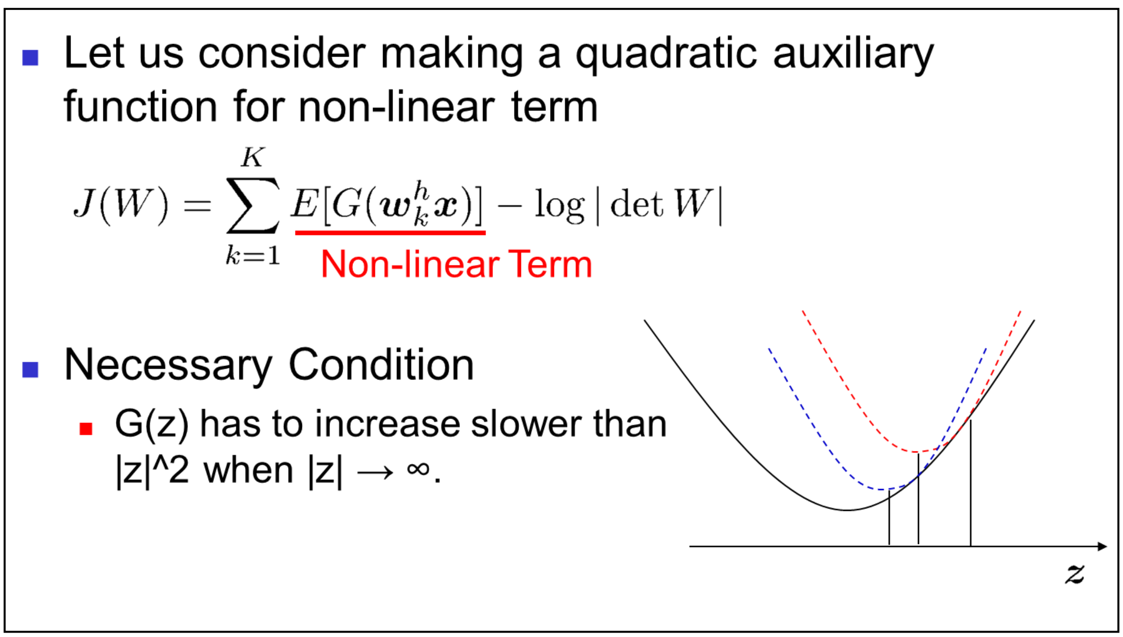


Figure 4. A basic idea of applying the auxiliary function approach to ICA.

As shown in Fig. 4, a basic concept is to majorize a non-linear term caused from a contrast function by a quadratic function. For that, the contrast function term should increase more slowly than a quadratic function when the L2 norm of a vector variable approaches infinity. Fig. 5 shows a class of the contrast function satisfying this condition. Fortunately, this class includes many well-used contrast functions and the condition indicates that the p.d.f. of source is spherical and super-Gaussian. An example is shown in Fig. 6.

- For any contrast function $G(z) \in S_G$, the following inequality holds.

$$G(z) \leq \frac{G'_R(|z_0|)}{2|z_0|}|z|^2 + R$$

The equality sign is valid iff $|z| = |z_0|$

- S_G is defined as

$$S_G = \{G(z) |\ \underbrace{G(z) = G_R(|z|)}_{\text{Spherical}}\}$$

where

- $G_R(r)$ is continuous and differentiable
- $G'_R(r)/r$ is continuous and monotonically decreasing in $r \in R_+$.

Super Gaussian

Figure 5. A class of contrast functions such that a quadratic function can be a majorizer for them.

- Source p.d.f.

$$p(y) \propto e^{-\log\cosh(y)}$$

- Contrast function

$$G(y) = \log\cosh(y)$$

- Weight function

$$G'_R(r)/r = \tanh(r)/r$$

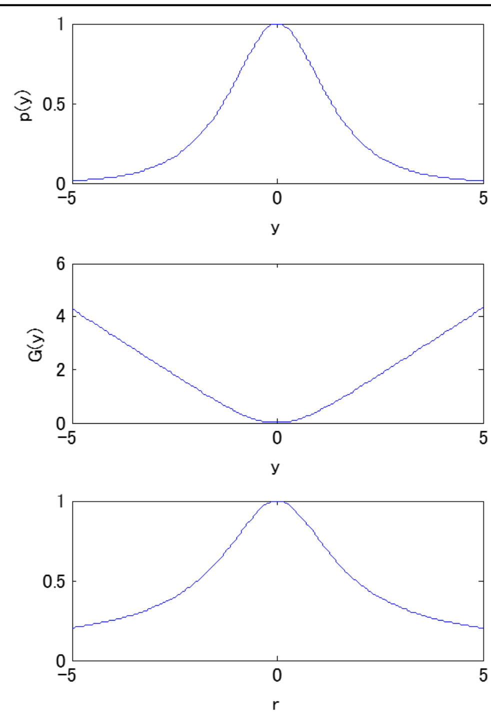


Figure 6. An example of source p.d.f., contrast function, and weight function

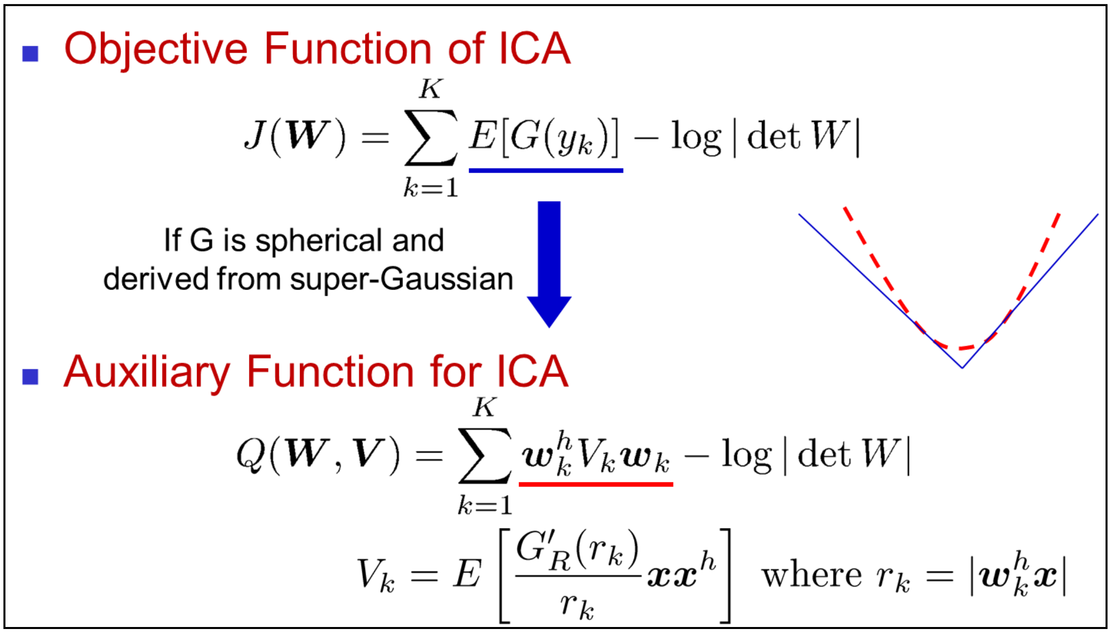


Figure 7. The derived auxiliary function for InfoMax-type ICA

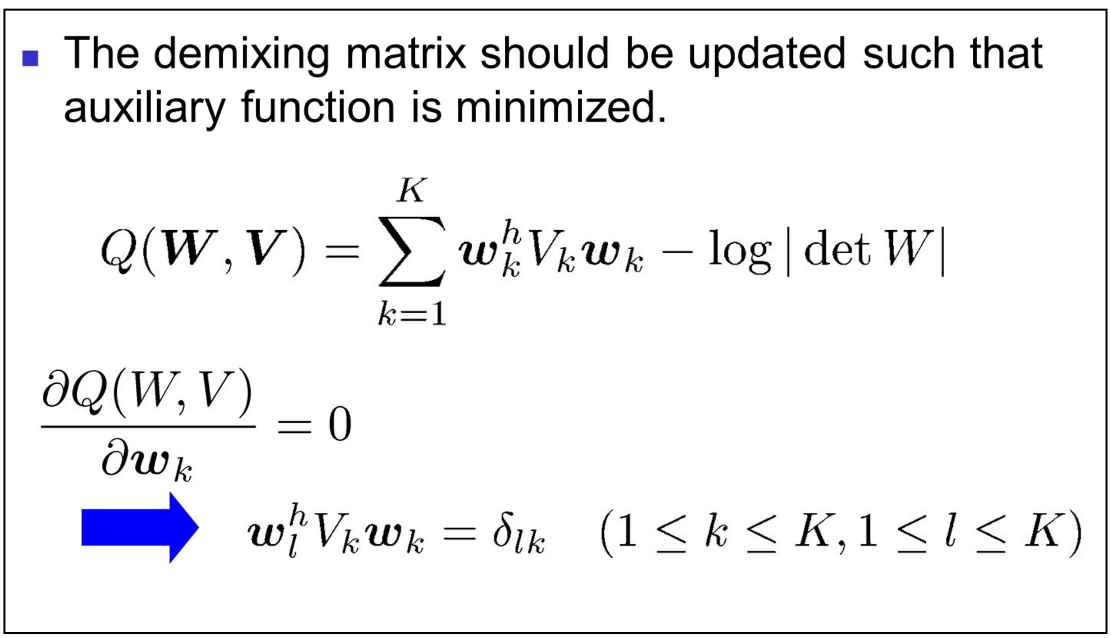


Figure 8. The derivation of update rules

Fig. 7 shows the derived auxiliary function for InfoMax-type ICA. The contrast function in the original objective function is replaced by a quadratic term. Fig. 8 shows the derivation of update rules. By taking the derivative, we have simultaneous vector equations as shown in Fig. 9. Ideally, all of w_n (a row vector of the demixing matrix) should be updated together. However, unfortunately, a closed-form solution for that has not yet been found in a general case. Instead, we can derive the closed-form update rule for each of w_n. Finally, the derived algorithm of AuxICA is shown in Fig. 10.

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

e.x.) Case of K=3

$$\begin{array}{ccc} \boldsymbol{w}_1^h V_1 \boldsymbol{w}_1 = 1 & \boldsymbol{w}_1^h V_2 \boldsymbol{w}_2 = 0 & \boldsymbol{w}_1^h V_3 \boldsymbol{w}_3 = 0 \\ \boldsymbol{w}_2^h V_1 \boldsymbol{w}_1 = 0 & \boldsymbol{w}_2^h V_2 \boldsymbol{w}_2 = 1 & \boldsymbol{w}_2^h V_3 \boldsymbol{w}_3 = 0 \\ \boldsymbol{w}_3^h V_1 \boldsymbol{w}_1 = 0 & \boldsymbol{w}_3^h V_2 \boldsymbol{w}_2 = 0 & \boldsymbol{w}_3^h V_3 \boldsymbol{w}_3 = 1 \end{array}$$

- Is there a closed-form solution? → Not yet found

1) Sequential update can be done by orthogonal projection.

2) This simultaneous equations can be reduced to generalized eigenvalue problem only when K=2.

Figure 9. The simultaneous vector equations for deriving the update rules of the demixing matrix

Weighted covariance matrix update

$$\begin{aligned} r_k &= |\boldsymbol{w}_k^h \boldsymbol{x}| \\ V_k &= E\left[\frac{G'_R(r_k)}{r_k} \boldsymbol{x}\boldsymbol{x}^h\right] \end{aligned}$$

Demixing matrix update

$$\begin{aligned} \boldsymbol{w}_k &\leftarrow (WV_k)^{-1} \underline{\boldsymbol{e}_k} \\ \boldsymbol{w}_k &\leftarrow \boldsymbol{w}_k / \sqrt{\boldsymbol{w}_k^h V_k \boldsymbol{w}_k} \end{aligned}$$

Unit vector with the *k*th element unity

Figure 10. The derived AuxICA algorithm

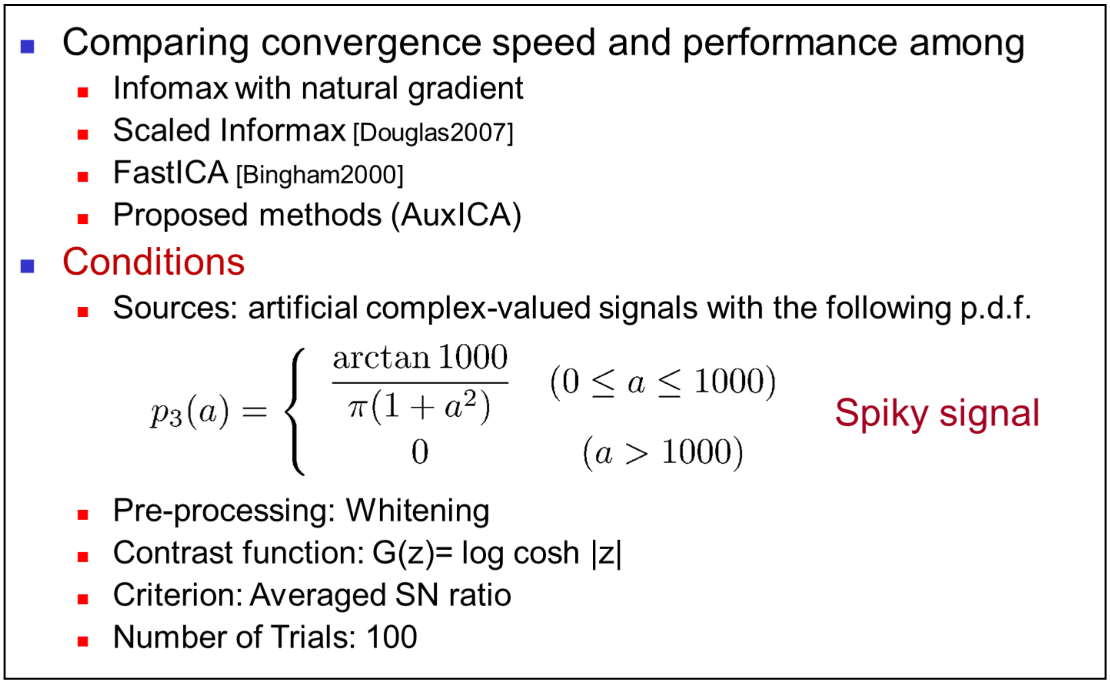


Figure 11. The experimental conditions for comparing the separation performance of AuxICA to other existing algorithms

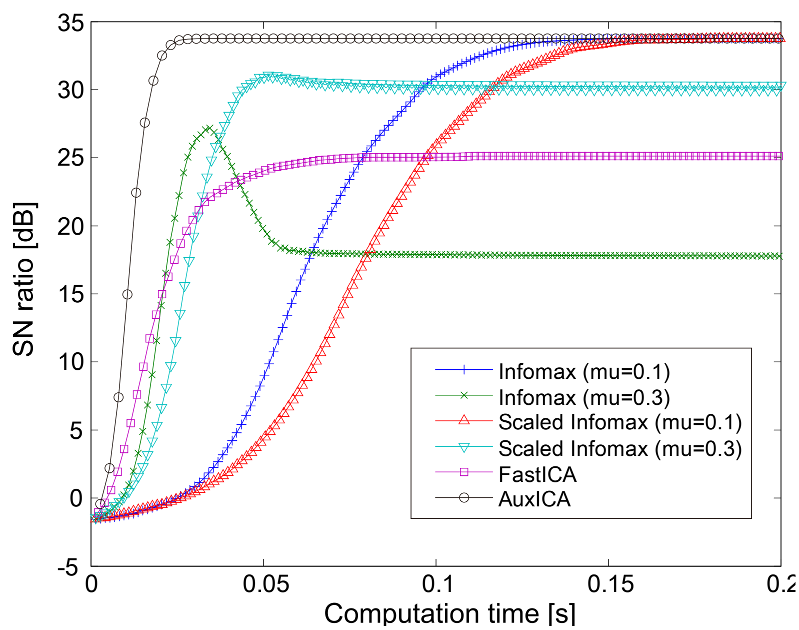


Figure 12. The experimental results for the separation performance of AuxICA compared to other existing algorithms

The experimental conditions and results of the experiment comparing the separation performance of AuxICA with other algorithms are shown in Fig. 11 and Fig. 12, respectively. It is shown that AuxICA shows the fastest convergence with the best separation performance.

4. AUXILIARY FUNCTION BASED INDEPENDENT VECTOR ANALYSIS [3]

For blind source separation of convolutive mixture, the frequency domain approach is well used. In this approach, as shown in Fig. 13, the time-domain signals are first converted to time-frequency representation typically by short-time Fourier transform (STFT). Conventionally, ICA is then applied at each frequency bin individually. However, in this case, we have to solve the well-known permutation problem. Recently, an extension of ICA to multivariate-type variables, called IVA has been proposed [11][12]. In IVA, the whole frequency components are modeled as a stochastic vector variable and simultaneously processed. Due to the model including dependencies over frequency components, IVA is theoretically not affected by the permutation ambiguity, which is its remarkable advantage.
Thanks to the resemblance of the objective function of IVA to that of ICA, applying the same idea derives an auxiliary function of IVA and the update rules called auxiliary function-based independent vector analysis (AuxIVA) as shown in Fig. 14 and Fig. 15, respectively. An example of speech separation is shown in Fig. 16.
As already mentioned, AuxICA or AuxIVA generally updates each of row vectors in a demixing matrix in order. However, in a stereo case (in the case that both the number of sources and the number of microphones is two), the simultaneous vector equations shown in Fig. 9 can be solved. In two dimensional space, if two vectors are orthogonal to the same vector, they should be in parallel. Then, the simultaneous vector equations can be deformed into a generalized eigenvalue problem as shown in Fig. 17 and the two row vectors of a demixing matrix can be updated together. The algorithm in detail is presented in [4]. The experimental conditions and results of the experiment comparing the separation performance of AuxIVA with the natural gradient-based IVA are shown in Fig. 18 and Fig. 19, respectively. The general algorithm of AuxIVA is denoted as AuxIVA1 and the special algorithm for the stereo case is denoted as AuxIVA2 in Fig. 19. It is shown that AuxIVA yields the fastest convergence with the best separation performance.

5. IMPLEMENTATION ON IPHONE [4]

As a prototype stereo BSS system on mobile phone, the whole algorithm of AuxIVA2 has been implemented on iPhone in cooperation with Redec co., Ltd. A picture of the system is shown in Figure 20. Under the conditions of 16kHz sampling frequency, 4096-point frame length, half frame shift, and 10 iterations, it took 1.7s to separate 10s input and 23.2s, which means a real time factor (RTF) of less than 1/5 was achieved. The demo movie can be shown in [13].

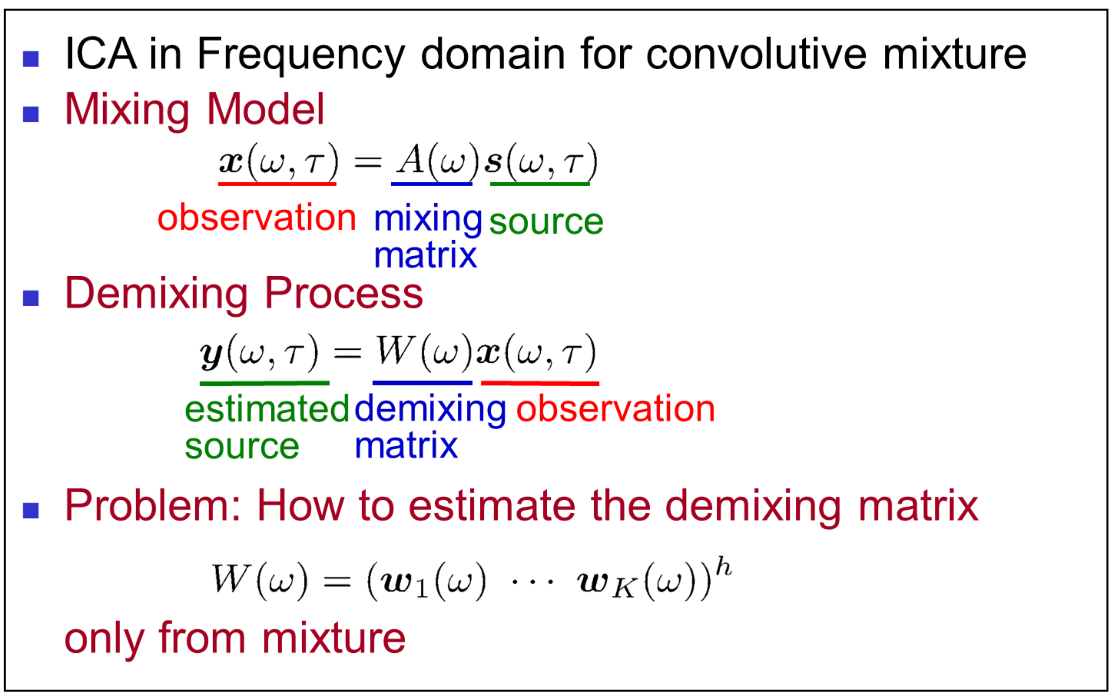
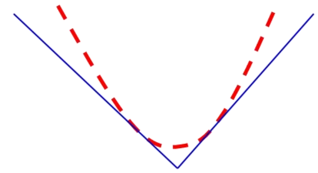


Figure 13. A formulation of frequency-domain ICA

- **Objective Function of IVA**

$$J(\boldsymbol{W}) = \frac{1}{N_\tau}\sum_{\tau=1}^{N_\tau}\sum_{k=1}^{K} \underline{G(\boldsymbol{y}_k(\tau))} - \sum_{\omega=1}^{N_\omega} \log|\det W(\omega)|$$

If G is spherical and derived from super-Gaussian



- **Auxiliary Function for IVA**

$$Q(\boldsymbol{W},\boldsymbol{V}) = \sum_{\omega=1}^{N_\omega}\left[\frac{1}{N_\tau}\sum_{\tau=1}^{N_\tau}\sum_{k=1}^{K} \underline{\boldsymbol{w}_k^h(\omega)V_k(\omega)\boldsymbol{w}_k(\omega)} - \log|\det W(\omega)|\right]$$

Figure 14. The derived auxiliary function for IVA

Weighted covariance matrix update

$$r_k = \sqrt{\sum_{\omega=1}^{N_\omega} |\boldsymbol{w}_k^h(\omega)\boldsymbol{x}(\omega)|^2}$$

Shared for all frequency bins

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

Demixing matrix update

$$\boldsymbol{w}_k(\omega) \leftarrow (W(\omega)V_k(\omega))^{-1}\underline{\boldsymbol{e}_k}$$

Unit vector with the *k*th element unity

$$\boldsymbol{w}_k(\omega) \leftarrow \boldsymbol{w}_k(\omega)/\sqrt{\boldsymbol{w}_k^h(\omega)V_k(\omega)\boldsymbol{w}_k(\omega)}$$

Figure 15. The derived AuxIVA algorithm

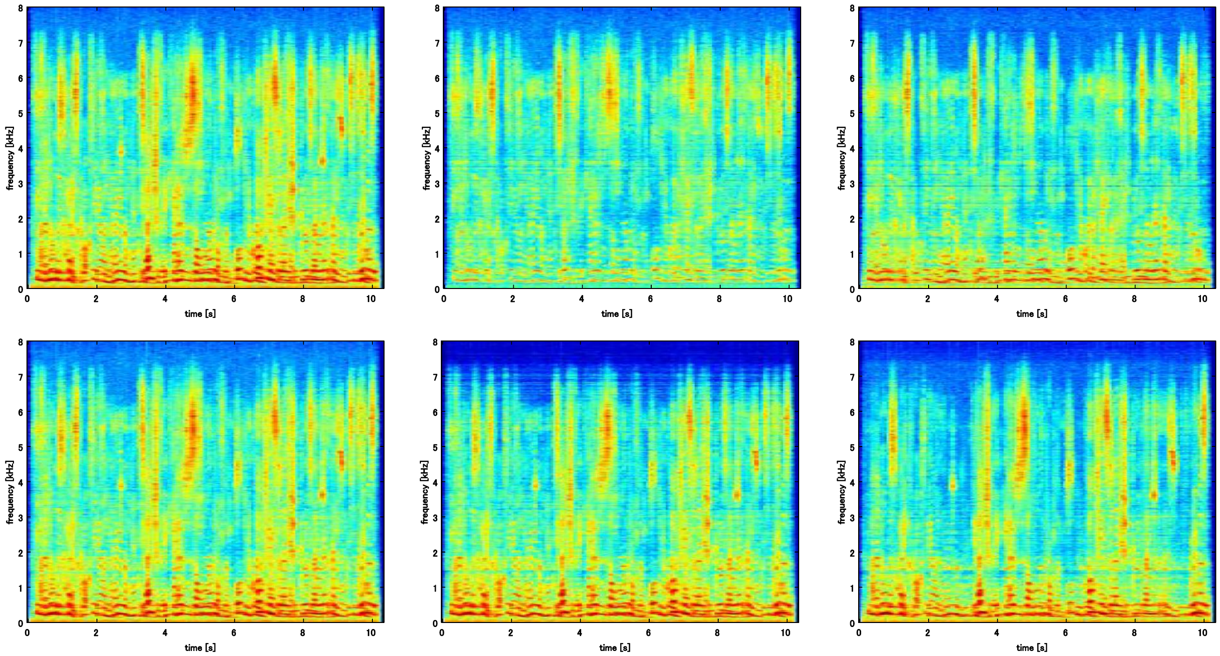


Figure 16. An example of speech separation by AuxIVA. The upper and the lower figures represent the spectrograms of the observations (left), the separated signals with 2 iterations (center), and with 10 iterations (right).

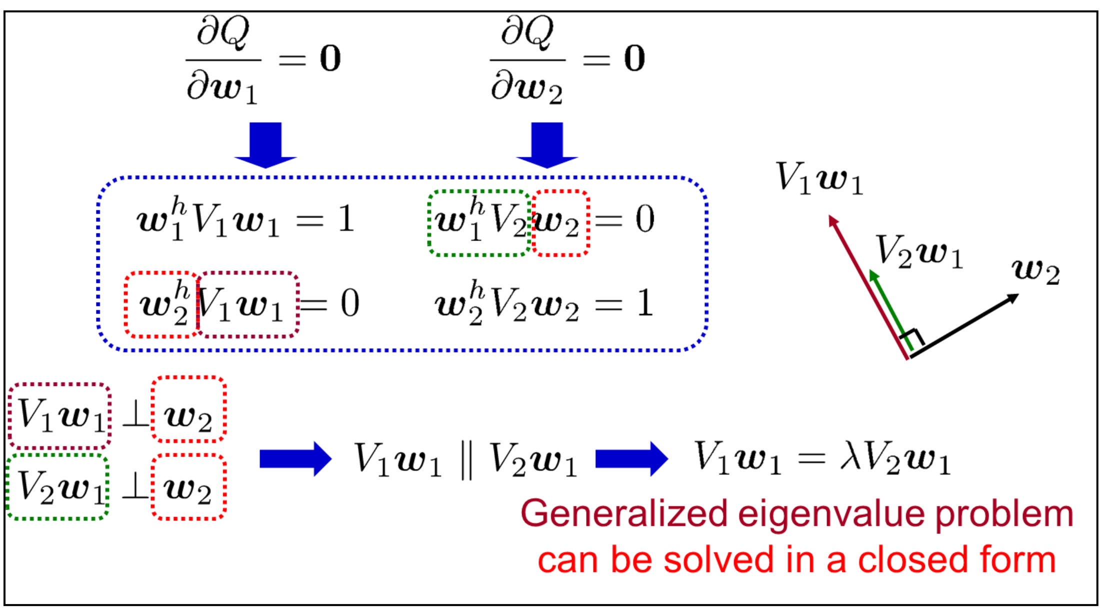


Figure 17. The closed-form solution of the simultaneous vector equation in the stereo case

- **Data Setup**
  - Source: speech from ATR B-set
  - Signal length: 10s
  - Impulse response: room E2A (R60=300ms) from RWCP sound scene database
  - Source direction: 10° , 30° , 50° , …, 170° (9 directions)
  - Sampling frequency: 16kHz
  - Number of sources/microphones: 2
- **Conditions**
  - Frame length: 4096 (256ms)
  - Frame shift: 2048 (128ms)
  - Window function: Hamming window
  - Contrast function of IVA: $G_R(r)=r$
  - Initial demixing matrix: Identity matrix
  - Evaluation criteria: average SIR improvement of ${}_9C_2=36$ trials

Figure 18. The experimental conditions for comparing the separation performance of AuxIVA to the natural gradient-based IVA

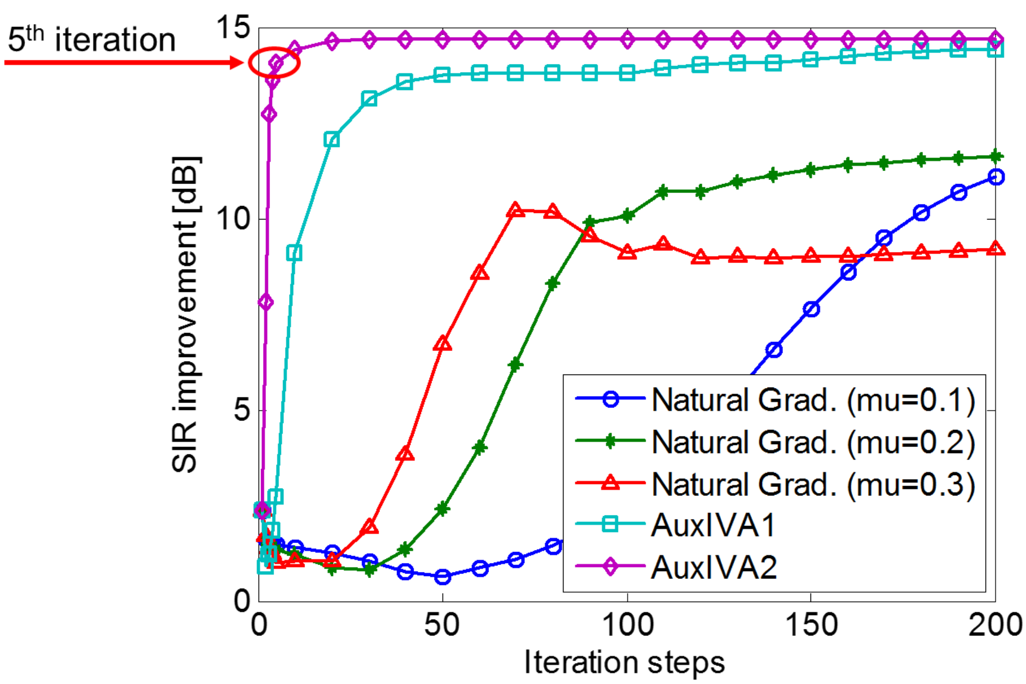


Figure 19. The experimental results for the separation performance of AuxIVA compared to the natural gradient-based IVA

Figure 20. A photo of an iPhone4 with a stereo microphone (TASCAM iM2 provided by TEAC Corporation)

REFERENCES

[1] Hyvärinen, A., Karhunen, J. and Oja, E., Independent component analysis, John Wiley & Sons, New York, NY, USA, (2001).

[2] Ono, N. and Miyabe, S., "Auxiliary-function-based independent component analysis for super-Gaussian sources," Proc. LVA/ICA, 165-172, (2010).

[3] Ono, N., "Stable and fast update rules for independent vector analysis based on auxiliary function technique," Proc. WASPAA, 189-192, (2011).

[4] Ono, N., "Fast stereo independent vector analysis and its implementation on mobile phone," Proc. IWAENC, (2012).

[5] Ono, T., Ono, N. and Sagayama, S., "User-guided Independent Vector Analysis with Source Activity Tuning," Proc. ICASSP, 2417-2420, (2012).

[6] Taniguchi, T., Ono, N., Kawamura, A. and Sagayama, S., "Online independent vector analysis based on auxiliary-function approach for real-time BSS," Proc. HSCMA, (2014).

[7] Kitamura, D., Ono, N., Sawada, H., Kameoka, H. and Saruwatari, H., "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," Proc. ICASSP, 276-280, (2015).

[8] Ramírez López, A., Ono, N., Remes, U., Palomäki, K. and Kurimo, M. "Designing multichannel source separation based on single-channel source separation," Proc. ICASSP, 469-473, (2015).

[9] Bell, A. J. and Sejnowski, T. J., "An information-maximization approach to blind separation and blind deconvolution," Neural Computation, vol. 7, no. 6, pp. 1129-1159, (1995).

[10] Amari, S., Cichocki, A. and Yang, H. H., "A new learning algorithm for blind signal separation," in Advances in Neural Information Processing Systems, Touretzky, D., Mozer, M. and Hasselmo, M., Eds., MIT Press, pp. 757-763, (1996).

[11] Hiroe, A., "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," Proc. ICA, pp. 601-608, (2006).

[12] Kim, T., Eltoft, T. and Lee, T.-W., "Independent vector analysis: An extension of ICA to multivariate components," Proc. ICA, pp. 165-172, (2006).

[13] http://www.onn.nii.ac.jp/index-e.html