Front. Electr. Electron. Eng. China 2008, 3(3): 343–346
DOI 10.1007/s11460-008-0064-9

RESEARCH ARTICLE

Hongyan LI, Jianfen MA, Deng'ao LI, Huakui WANG

Fixed-point blind source separation algorithm based on ICA

Abstract This paper introduces the fixed-point learning algorithm based on independent component analysis (ICA); the model and process of this algorithm and simulation results are presented. Kurtosis was adopted as the estimation rule of independence. The results of the experiment show that compared with the traditional ICA algorithm based on random grads, this algorithm has advantages such as fast convergence and no necessity for any dynamic parameter, etc. The algorithm is a highly efficient and reliable method in blind signal separation.

Keywords independent component analysis (ICA), blind source separation, fixed-point algorithm

1 Introduction

The crucial problem in blind source separation is to reconstruct source signals depending on proper transform methods such as principal component analysis (PCA) and factor analysis (FA) that have been applied widely. All these linear transform methods adopt almost only second-order statistics method rather than consider higher order statistics properties. Independent component analysis (ICA) [1–3] has been a new blind signal process method in the past several years, and the essence of the method is to decompose observational signals into several independent source components by an optimized algorithm depending on the statistics independent principle. The method has been applied in many areas because of good separation capability and few transcendental knowledge of source signals. In this paper, a fixed-point algorithm [2,3] based on ICA is introduced and the ICA model as well as the result are presented. Finally, a detailed convergence and stability is analyzed.

Translated from *Journal of Taiyuan University of Technology*, 2007, 38(1): 35–37 [译自: 太原理工大学学报]

Hongyan LI (✉), Jianfen MA, Deng'ao LI, Huakui WANG
College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China
E-mail: lhy6018@163.com

2 Basic principle of ICA

The ICA model is given in Fig. 1 [4]. The source signals, $\boldsymbol{S}(t) = [\boldsymbol{S}_1(t), \boldsymbol{S}_2(t), \ldots, \boldsymbol{S}_M(t)]^{\mathrm{T}}$ are supposed to be stationary and independent while it is unknown and mixed by linear system $\boldsymbol{A}$; we gain the observation signals $\boldsymbol{X}(t) = [\boldsymbol{X}_1(t), \boldsymbol{X}_2(t), \ldots, \boldsymbol{X}_N(t)]^{\mathrm{T}}$. The relationship of original signals and observation signals is given by the model [5]

$$\boldsymbol{X}(t) = \boldsymbol{A}\boldsymbol{S}(t). \tag{1}$$

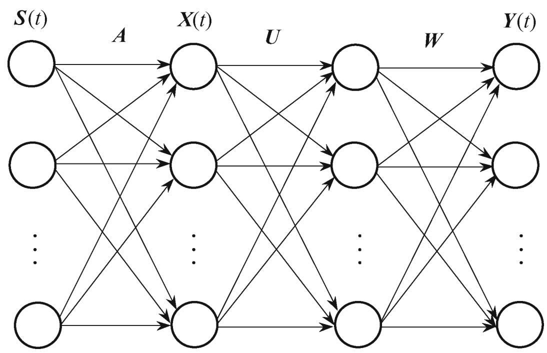

Fig. 1 ICA model

Pre-whitening the observation signals, we can get the observational signals $\hat{\boldsymbol{X}}(t) = \left[\hat{\boldsymbol{X}}_1(t), \hat{\boldsymbol{X}}_2(t), \ldots, \hat{\boldsymbol{X}}_N(t)\right]^{\mathrm{T}}$, and

$$\hat{\boldsymbol{X}}(t) = \boldsymbol{U}\boldsymbol{X}(t). \tag{2}$$

$\boldsymbol{U}$ is called whitening matrix that made the components of whitening signals uncorrelated and have unit-variance. The common method of whitening may be accomplished by eigenvalue decomposing of the covariance of the observation signals, that is

$$\boldsymbol{R}_X = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{T}}, \tag{3}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_m]$ is a diagonal matrix with the eigenvalues of the data covariance matrix $E[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$, and $\boldsymbol{V}$ is a matrix with corresponding eigenvectors as its column.

Make $\boldsymbol{U} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{V}^{\mathrm{T}}$, then the whitening signals is depicted in the form

$$\hat{\boldsymbol{X}}(t) = \boldsymbol{U}\boldsymbol{X}(t) = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}(t). \tag{4}$$

The goal of ICA is to estimate a separation matrix $\boldsymbol{W}$, making $\boldsymbol{Y}(t) = \boldsymbol{W}\hat{\boldsymbol{X}}(t)$ the estimation of original signals and its components independent of each other.

The separation matrix $\boldsymbol{W}$ is called a bogus inverse matrix of mixing matrix because the estimation of the separation matrix is not $\boldsymbol{A}^{-1}$, while it satisfies $\boldsymbol{WA} = \boldsymbol{PI} = \boldsymbol{C}$, where $\boldsymbol{P}$ is a permutation matrix, $\boldsymbol{I}$ is a unit matrix. This brings the nondetermination of the ICA solution since the order and the amplitude of original signals and separation signals may be different. In practice, the nondetermination may be accepted if the wave shape is not changed.

Assumptions are needed for the linear mixing ICA as follows:

1) The number of sensors is not less than that of original signals, that is, $M \geqslant N$.

2) The original signals are supposed to be independent of each other.

3) Only up to one source may be Gaussian.

4) The noise can be omitted.

3 Random grads algorithm

The key of ICA is how to measure the independence of the components of separated signals $\boldsymbol{Y}(t)$. Here, kurtosis is used as the judge ruler. Kurtosis is a classical measure method of non-Gaussianity, which can be defined as follows:

$$\mathrm{kurt}(\boldsymbol{S}_i) = E\left[\boldsymbol{S}_i^4\right] - 3\left(E\left[\boldsymbol{S}_i^2\right]\right)^2. \tag{5}$$

Kurtosis has characters depicted as follows:

$$\begin{aligned} &\mathrm{kurt}(x_1 + x_2) = \mathrm{kurt}(x_1) + \mathrm{kurt}(x_2), \\ &\mathrm{kurt}(ax_1) = a^4 \mathrm{kurt}(x_1). \end{aligned} \tag{6}$$

The objection function is defined as follows:

$$\mathrm{kurt}\left(\boldsymbol{w}^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right) = E\left[\left(\boldsymbol{w}^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^4\right] - 3\left[E\left\{\left(\boldsymbol{w}^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^2\right\}\right]^2. \tag{7}$$

Since the observation signals have been pre-whitened, Eq. (7) can be simplified as

$$\mathrm{kurt}\left(\boldsymbol{w}^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right) = E\left[\left(\boldsymbol{w}^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^4\right] - 3||\boldsymbol{w}||^4. \tag{8}$$

Seek the gradient for the objection function and obtain

$$\Delta w \alpha E\left[\hat{\boldsymbol{x}}_i\left(\boldsymbol{w}_i(k)^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^3\right] - 3||\boldsymbol{w}_i(k)||^2\boldsymbol{w}_i(k). \tag{9}$$

Consider the restriction condition $\|\boldsymbol{w}\| = 1$ and introduce a step gene μ in the iteration process to improve the convergence speed, the algorithm formula becomes

$$\boldsymbol{w}_i(k+1) = \boldsymbol{w}_i(k) + \mu\left\{E\left[\hat{\boldsymbol{x}}_i\left(\boldsymbol{w}_i(k)^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^3\right] - 3\boldsymbol{w}_i(k)\right\}, \tag{10}$$

where $\boldsymbol{w}_i(k)$ is a row vector corresponding to the ith original signal after iteration for k time. When the two adjacent $\boldsymbol{w}_i(k)$ have no change or vary little, the iteration finishes. The $\boldsymbol{w}_i(k)$ should be normalized in each iteration process, that is, make $\boldsymbol{w}_i(k) = \boldsymbol{w}_i(k)/\|\boldsymbol{w}_i(k)\|$ so that the separation signal has unit energy.

4 Fixed-point algorithm

The simulation and analysis adopt the random grads algorithm. The convergence speed of the algorithm is slow and dependent on μ, Therefore, the fixed-point algorithm is used here, which is a much faster and reliable algorithm. The fixed-point means that the variety of $\boldsymbol{w}$ is 0, that is,

$$E\left[\hat{\boldsymbol{x}}_i\left(\boldsymbol{w}_i(k)^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^3\right] - 3\boldsymbol{w}_i(k) = 0. \tag{11}$$

The iteration formula of the fixed-point algorithm becomes

$$\boldsymbol{w}_i(k+1) = c\left\{E\left[\hat{\boldsymbol{x}}_i\left(\boldsymbol{w}_i(k)^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^3\right] - 3\boldsymbol{w}_i(k)\right\}, \tag{12}$$

where c is a constant. Make $c = 1$, Eq. (12) becomes

$$\boldsymbol{w}_i(k+1) = E\left[\hat{\boldsymbol{x}}_i\left(\boldsymbol{w}_i(k)^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^3\right] - 3\boldsymbol{w}_i(k). \tag{13}$$

According to the iteration formula, the algorithm step can be given as follows:

1) Make $i = 1$.

2) Randomly select the normalized original vector $\boldsymbol{w}(0)$ and make $k = 1$.

3) Make $\boldsymbol{w}_i(k) = E\left[\hat{\boldsymbol{x}}_i\left(\boldsymbol{w}_i(k-1)^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right)^3\right] - 3\boldsymbol{w}_i(k-1)$.

4) Make $\boldsymbol{w}_i(k) = \boldsymbol{w}_i(k)/\|\boldsymbol{w}_i(k)\|$.

5) If $|\boldsymbol{w}_i(k)^{\mathrm{T}}\boldsymbol{w}_i(k-1)|$ converges to 1, the iteration finishes, and output vector $\boldsymbol{w}_i(k)$, or make $k = k+1$, return to step 3).

6) If i is less than the number of original signals, return to step 2) until all original signals have been separated.

A simple orthogonal should be added in cycle process to ensure that different components can be separated, that is, adding a transform before step 4):

$$\boldsymbol{w}_i(k) = \boldsymbol{w}_i(k) - \sum_{j=1}^{k-1} \boldsymbol{w}_i(k)^{\mathrm{T}}\boldsymbol{w}_i(j)\boldsymbol{w}_i(j).$$

5 Analysis of simulation and conclusions

We select three speech signals from ICALAB for Signal Processing as depicted in Fig. 2, of which the sample frequency is $f_{\mathrm{s}} = 8$ kHz, and the data length is $M = 5000$. After mixing linearly using a 3×3 randomly chosen matrix, $\boldsymbol{A}$ is obtained

$$\boldsymbol{A} = \begin{bmatrix} 0.4103 & 0.3529 & 0.1389 \\ 0.8936 & 0.8132 & 0.2028 \\ 0.0579 & 0.0099 & 0.1987 \end{bmatrix}.$$

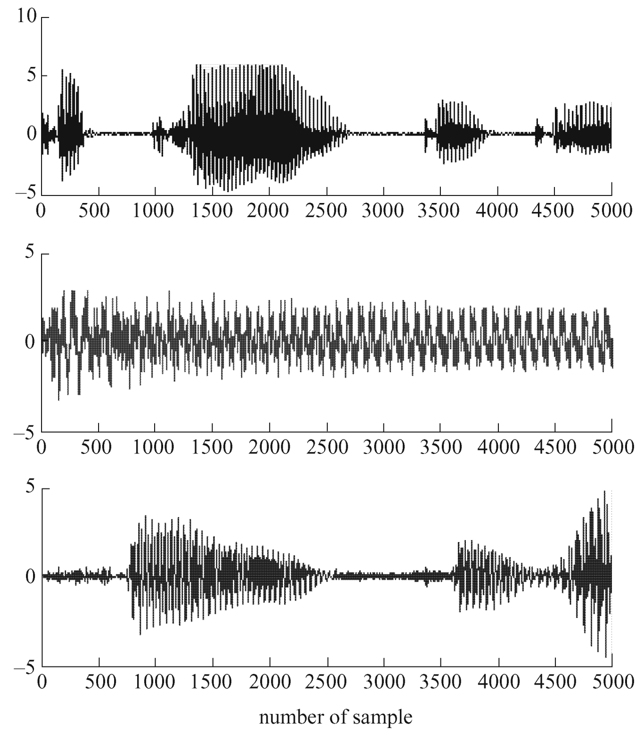

Fig. 2 Original speech signals

The three mixed speech signals are shown in Fig. 3 and the separated speech signals by fixed-point algorithm are depicted in Fig. 4.

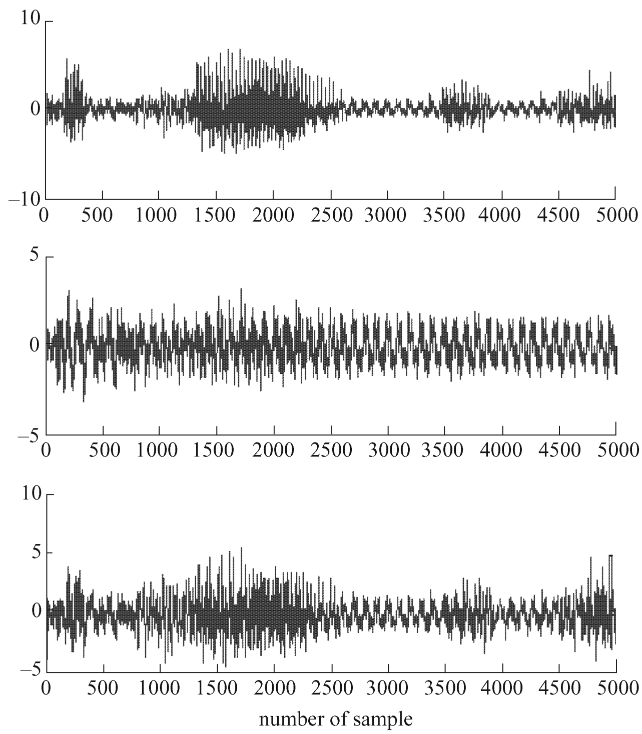

Fig. 3 Mixed speech signals

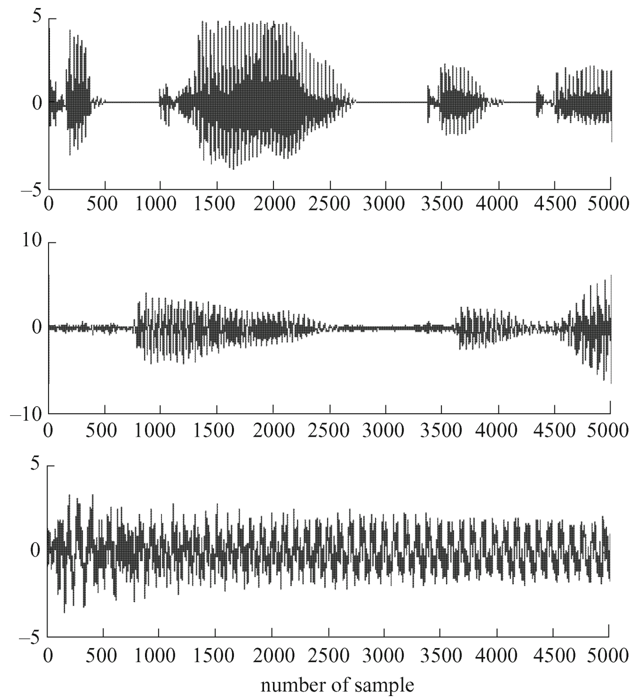

Fig. 4 Separated speech signals by fixed-point algorithm

Ten simulation comparisons for the two algorithms are conducted. Let $\mu = \lambda^n$ in the random grads algorithm, where $\lambda = 0.005$, $\lambda = 0.05$, $\lambda = 0.01$. When $\lambda = 0.005$, the convergence speed is the fastest, while the algorithm has obvious distortion, and when $\lambda = 0.01$, the simulation result is the best. However, the fixed-point algorithm demonstrates very stable convergence speed with ideal effect. When $\lambda = 0.05$, the convergence comparison between the fixed-point algorithm and the random grads algorithm shows the advantage of the former one.

Compared with the random grads algorithm (Fig. 5), the fixed-point algorithm is promising with many advantages

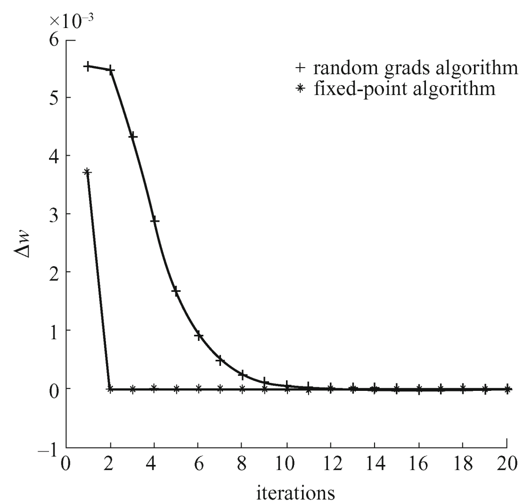

Fig. 5 Convergence performance comparison between random gradient algorithm and fixed-point algorithm

such as high convergence speed, high stability and independence of any dynamic parameter.

Acknowledgements This work was supported by the Shanxi Young Science Fund (No. 2008021022).

References

1. Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Networks, 2000, 13(4–5): 411–430
2. Hyvärinen A. Fast and robust fixed-point algorithm for independent component analysis. IEEE Transactions on Neural Networks, 1999, 10(3): 626–634
3. Shi Z, Tang H, Tang Y. A new fixed-point algorithm for independent component analysis. Neural Computing, 2004, 56: 467–473
4. Jutten C, Herault J. Blind Separation of sources, part 1: an adaptive algorithm based on neuromimetic architecture. Signal Processing, 1991, 24(1): 1–10
5. Karhunen J, Oja E, Wang L, et al. A class of neural networks for independent component analysis. IEEE Transactions on Neural Networks, 1997, 8(3): 486–504