Research Article
Improved Fast ICA Algorithm Using Eighth-Order Newton's Method

Tahir Ahmad, Norma Alias, Mahdi Ghanbari and Mohammad Askaripour
IbnuSina Institute for Fundamental Sciences, Universiti Teknologi Malaysia 81310 Skudai,
Johor, Malaysia

Abstract: Independent Component Analysis (ICA) is a computational method to solve Blind Source Separation (BSS) problem. In this study, an improved Fast ICA based on eighth-order Newton's method is proposed to solve BSS problems. Eight-order Newton's method for finding the solution of nonlinear equations is much faster than ordinary Newton's iterative method. The improved FastICA algorithm is applied to separate sound signals. The simulation results show the method has fewer iterations and faster convergence speed.

Keywords: FastICA, Independent Component Analysis (ICA), Newton's method

INTRODUCTION

Independent Component Analysis (ICA) or Blind Source Separation (BSS) is a signal processing method that extracts statistically independent components from a set of measured signals (Hyvarinen, 1998; Lee *et al.*, 1999; Ye *et al.*, 2006). There are various ICA approaches including maximum likelihood estimation (Shia *et al.*, 2004; Cardoso, 1998), mutual information minimization (Amari *et al.*, 1996; Bell and Sejnowski, 1995; Pham, 2004) and negentropy maximization (Hevarinen *et al.*, 2001). The FastICA algorithm was presented by Hyvarinen (1999). A kind of Newton-type algorithm was used in FastICA and the algorithm was based on fixed-point algorithm iteration to maximize nongaussianity as a measure of statistical independence. In FastICA algorithm, the Newton iteration method with two-order convergence was employed in a fixed-point algorithm for estimating the separation matrix.

In this study, we first briefly introduce ICA algorithms with Kurtosis and negentropy contrast function. Then, an improved FastICA based on an eighth-order convergence Newton's Method (Li *et al.*, 2011) is proposed. Using this new algorithm, the update equations inside the FastICA algorithm is developed for estimating separation matrix with less iterative. Then, the new algorithm is applied to separation sound signals. Finally, the performance of the newalgorithm is compared with FastICA algorithm in terms of evaluate performance rate and convergence rates.

INDEPENDENT COMPONENT ANALYSIS AND FASTICA

Independent Component Analysis (ICA) has become a powerful method for signal processing in last decade (Cichochi and Amari, 2004; Hevarinen *et al.*, 2001; Chio *et al.*, 2005). The basic linear mixture model (without noise) for ICA is as follows:

$$X = AS \tag{1}$$

where, $S = (S_1, \ldots, S_m)^T$ denotes the original independent sources and $X = (X_1, \ldots, X_n)^T$ denoted mixtures of original sources and A is the unknown $m \times n$ constant mixing matrix. ICA aims to recover or estimate the independent components (ICs) from their mixtures. An estimated IC is denoted by y and a separation vector is denoted by$w$, such that

$$y = w^T X = w^T AS \tag{2}$$

The signal mixtures tend to have Gaussian probability density functions (pdfs) and the source signals have nongaussianpdfs (Stone, 1999). In terms of central limit theorem (CLT), Gaussian signals often consist of a mixture of nongaussian signals (Stone, 1999). Given a set of such Gaussian mixtures signal $X = (X_1, \ldots, X_n)^T$. Therefore, the process of ICA is to find source signal by finding separation vector w that extract signal y, which has to be the most nongaussian. One of the classic measures for estimation of nongaussianity of random variable is Kurtosis. We denote the Kurtosis of y by $Kurt(y)$ and it is defined by:

$$Kurt(y) = E[y^4] - 3(E[y^2])^2 \tag{3}$$

where, y is a random variable with zero mean. If y is normalized, variance of y is equal to one i.e.$E[y^2] = 1$. Then, (3) is simplified to:

Corresponding Author: Tahir Ahmad, Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia 81310 Skudai, Johor, Malaysia

$$Kurt(y) = E[y^4] - 3 \qquad (4)$$

Negentropy for Nongaussianity: An important measure of nongaussianity is negentropy. The concept of negentropy strongly depended on entropy that it is another fundamental concept in information theory. Given X as a discrete value random variable, Entropy H for random variable X is defined by:

$$H(X) = -\sum_i P(X = x_i) \ln P(X = x_i) \qquad (5)$$

In Eq. (5), the x_i are the value of random variable X and P is density function of the random variable X (Cover and Thomas, 1991; Papoulis, 1991).

The definition of entropy for continues random variable is obtained from generalizing (5). Given X as a continue random variable with density function $P_X(.)$. The differential entropy H for continues random variable X is defined by:

$$H(X) = -\int P_X(\gamma) \ln(\gamma) d\gamma \qquad (6)$$

Let that X as a vector random variable. Negentropy of X is showed by $J(X)$ and is defined by:

$$J(X) = H(X_{gauss}) - H(X) \qquad (7)$$

where, X_{gauss} is gaussian random variable with same covariance of the X. Negentropy is another measure for nongaussianity. But, for the estimate of negentropy, we must first estimate the pdf of the vector random variable. Therefore, it is very hard to obtain it from the computation. One method to approximate negentropy by nonpolynomial function was proposed by Hyvarinen (1998). In related to this study, a method using expansion of pdf to approximate negentropy was given by Hevarinen *et al.* (2001). Let y be a whitened random variable. Then, approximate of $J(y)$ was given:

$$J(y) \propto (E[G(y)] - E[G(\vartheta)])^2 \qquad (8)$$

where, G is a nonquadratic function and variable ϑ is a gaussian variable with unit variance and zero mean. Note that "proportion to" is the means of symbol $\propto$. In order to achieve a very strongly estimator, choose G that grow slowly. Whereby:

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y \qquad (9)$$

$$G_2(y) = -\exp\left(-\frac{y^2}{2}\right) \qquad (10)$$

and $1 \leq a_1 \leq 2$ (Hevarinen *et al.*, 2001).

A fixed-point algorithm using negentropy: A FastICA algorithm based on maximizing negentropy was introduced by Hyvarinen (1999). In this algorithm, fixed-point iteration and Newton's method with two-order convergence is used to find separation w vector as follows:

$$w \leftarrow w - \frac{[E\{zg(w^Tz)\} - \beta w]}{[E\{g'(w^Tz)\} - \beta]} \qquad (11)$$

where, w is separation vector, g is function as Eq. (9) and (10), z is the whitening observation vector, $\beta = E\{w_{opt}^T zg(w_{opt}^T z)\}$ and w_{opt} corresponds to optimum w. Equation (11) can be simplified by multiplying both sides by $\beta - E\{g'(w^Tz)\}$ and:

$$w \leftarrow E\{g'(w^Tz)\} - E\{zg(w^Tz)\}w \qquad (12)$$

Let X is a vector random variable. Assume that z is given by the whitening of X. the fixed-point algorithm using negentropy is given as follow (Hevarinen *et al.*, 2001):

Step 1 : Let $k = 0$ and $w(0)$ is an initial random vector with unit norm

Step2 : Let $w(k) = E\{zg(w(k-1)^Tz)\} - E\{g'(w(k-1)^Tz)\}w(k-1) \qquad (13)$

Step 3 : Let $w(k) = \frac{w(k)}{\|w(k)\|} \qquad (14)$

Step 4 : If the algorithm is not converged, then $k = k + 1$ and go back to step 2.

From this algorithm, we obtain one component. Therefore, in order to estimate several independent components, the algorithm needs to be run for several times.

EIGHT-ORDER NEWTON'S METHOD

The classic Newton's method for single nonlinear equation can be expressed as follows:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (15)$$

This is an important and basic method, which converge quadratically. Pota and Petak (Potra and Ptak, 1984) proposed a modification Newton's method with third-order convergence defined by:

$$x_{n+1} = x_n - \frac{f(x_n) + f\left(x_n - \frac{f(x_n)}{f'(x_n)}\right)}{f'(x_n)} \qquad (16)$$

King (1973) developed a one- parameter family of forth-order methods, which is written as follows:

$$y_n = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (17)$$

$$x_{n+1} = y_n - \frac{f(x_n)+\beta f(y_n)}{f(x_n)+(\beta-2)f(y_n)} \frac{f(y_n)}{f'(x_n)} \qquad (18)$$

Such that $\in R$. Many methods have been proposed to improve the order of convergence such as (Kou, 2007; Chun, 2007; Parhi and Gupta, 2008; Bi *et al.*, 2008).

Li *et al.* (2011) proposed an eighth-order Newton's method convergence defined by:

$$u_{n+1} = x_n - \frac{f(x_n)+f\left(x_n-\frac{f(x_n)}{f'(x_n)}\right)}{f'(x_n)} \qquad (19)$$

$$v_{n+1} = u_{n+1} - \frac{f(u_{n+1})}{f'\left(x_n-\frac{f(x_n)}{f'(x_n)}\right)} \qquad (20)$$

$$x_{n+1} = v_{n+1} - \frac{f(v_{n+1})}{f'(v_{n+1})} \qquad (21)$$

It is observed that this method required less number of iterations then traditional Newton's method:

IMPROVED FASTICA USING EIGHTH-ORDER NEWTON'S METHOD

With regards to Eq. (11) can be derived new equations for estimating the separation matrix (un-mixing matrix) by using Eq. (19) until (21) as follows:

$$w \leftarrow w_v - \frac{E\{zg(w_v^T z)\}-\beta w_v}{E\{g'(w_u^T z)\}-\beta} \qquad (22)$$

Such that:

$$w_v \leftarrow w_u - \frac{E\{zg(w_u^T z)\}-\beta w_u}{E\{g'(w_y^T z)\}-\beta} \qquad (23)$$

$$w_u \leftarrow w - \frac{E\{zg(w^T z)\}-\beta w+E\{zg(w_y^T z)\}-\beta w_y}{E\{g'(w^T z)\}-\beta} \qquad (24)$$

$$w_y \leftarrow w - \frac{E\{zg(w^T z)\}-\beta w}{E\{g'(w^T z)\}-\beta} \qquad (25)$$

Equations (22) to (25) can be simplified to:

$$w \leftarrow E\{zg(w_v^T z)\} - E\{g'(w_u^T z)\}w_v \qquad (26)$$

where,

$$w_v \leftarrow E\{zg(w_u^T z)\} - E\{g'(w_y^T z)\}w_u \qquad (27)$$

$$w_u \leftarrow E\{zg(w^T z)\} + E\{zg(w_y^T z)\} - E\{g(w^T z)\}w + \beta w_y \qquad (28)$$

$$w_y \leftarrow E\{zg(w^T z)\} - E\{g'(w^T z)\}w \qquad (29)$$

and $\beta \in R$ in (28) will be updated in every iterations by $\beta = E\{w^T z g(w^T z)\}$. Hence, we propose following fixed-point algorithm for estimate the separation matrix:

Step1: Let $n = 0$ and given w_0 as an initial random vector with unit norm. Assume $\beta = E\{w_0^T z g(w_0^T z)\}$

Step 2: Let:

$$w_{n+1} = E\{zg(w_v^T z)\} - E\{g'(w_u^T z)\}w_v \qquad (30)$$

where,

$$w_v = E\{zg(w_u^T z)\} - E\{g'(w_y^T z)\}w_u \qquad (31)$$

$$w_u = E\{zg(w_n^T z)\} + E\{zg(w_y^T z)\} - E\{g(w_n^T z)\}w + \beta w_y \qquad (32)$$

$$w_y = E\{zg(w_n^T z)\} - E\{g'(w_n^T z)\}w_n \qquad (33)$$

Step 3: $w_{n+1} = \frac{w_{n+1}}{\|w_{n+1}\|}$

Step 4: If the algorithm is not converged then, $\beta = E\{w_{n+1}^T z g(w_{n+1}^T z)\}, n = n + 1$ *and* go back to step 2.

SIMULATION RESULTS

We used three sound signals from MATLAB database as shown in Fig. 1. Each size of simple signals is 1000. These source signals were randomly mixed. The mixed signals are shown in Fig. 2. After whitening the mixed signals, we ran the proposed improved FastICA algorithm to separate the mixed signals. The results are shown in Fig. 3.
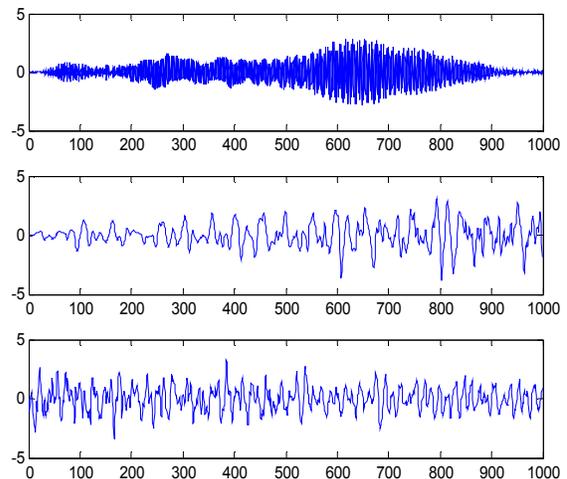
Fig. 1: Three source signals
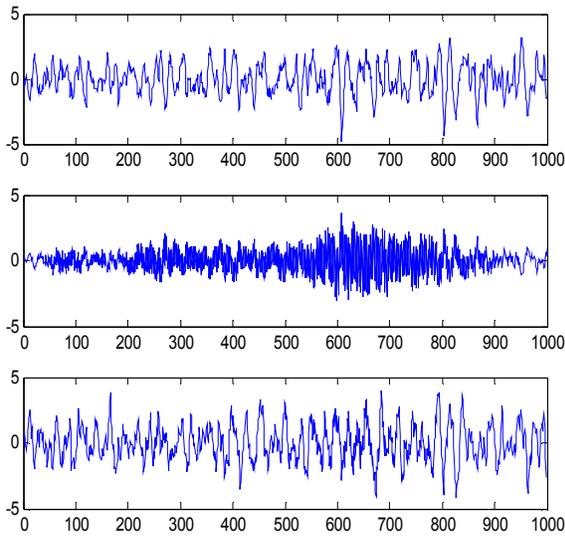
Fig. 2: Three mixed signals
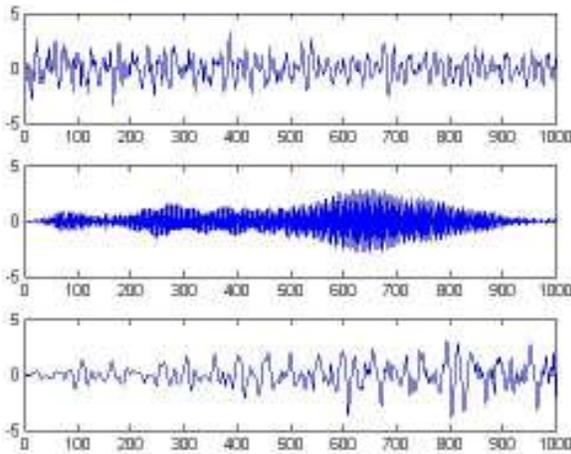
Fig. 3: Separation the signals

Table 1: Comparison of the separation performance

NO	1	2	3	4	5	6
FastICA	0.2156	19340	0.1873	0.1721	0.1964	0.2143
Improved algorithm	0.2155	19350	0.1871	0.1724	0.1961	0.2146

Table 2: Comparison of iteration numbers and CPU system

	Fast ICA	Improved algorithm
The 1th component	23.11	2.35
The 2th component	9.45	4.71
The 3th component	4	3.2N
Total iteration	6824.	10.26
CPU time	0.2412 sec.	0.1381 sec.

We can also get the same result when using the Fast ICA algorithm (Hevarinen *et al.*, 2001). In order to measure the accuracy of separation, we calculated the performance index (Amari *et al.*, 1996; Shi and Zhang, 2006).

$$PI = \frac{1}{n^2}\left\{\sum_{i=1}^{n} rPI_i + \sum_{j=1}^{n} cPI_j\right\} =$$
$$\frac{1}{n^2}\left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|}{\max_k|p_{ik}|} - 1\right) +\right.$$
$$j=1ni=1npij\max kpkj-1 \tag{34}$$

where, $rPI_i = \sum_{j=1}^{n}\frac{|p_{ij}|}{\max_k|p_{ik}|} - 1$ and $cPI_j = \left(\sum_{i=1}^{n}\frac{|p_{ij}|}{\max_k|p_{kj}|} - 1\right)$ in which p_{ij} is the ijth element of $n \times n$ matrix $P = wVA$, where w is a separation matrix, V is a whitening matrix and A isa mixing matrix. The large value PI is the poor statistical performance of the BSS algorithm (Amari *et al.*, 1996; Shi and Zhang, 2006). These algorithms were run six times to evaluate their performances. The results are presented in Table 1. Whereby the separation performance of proposed algorithm is almost the same as the FastICA algorithm. Furthermore, we compared the convergence speed of these two algorithms. We randomly ran each of these two methods for 15 times executively. We then calculated the average iterative numbers and presented in Table 2.

As a result, the improved Fast ICA algorithm performances with less iteration than Fast ICA, while maintaining comparable separation performance. This is due to the fact that convergence of eighth-order Newton's method is much faster than the traditional Newton's method with second-order convergence to find the solution for non-linear equation. To calculate the CPU time, two algorithms 15 times was run by using the MATLAB version R2010 on the Intel(R) Core(TM) 2 DUE CPU T6400 2.00 GHZ processor with 2.00 GB RAM. Table 2 describes iteration numbers and the CPU times required for sound signal separation from three mixed sound signal, whereby the average CPU times are 0.2412 s and 0.1381s, respectively. This demonstrated that the proposed method exhibited faster convergence with less iteration than Fast ICA algorithm, with comparable separation performance.

CONCLUSION

In this study, improved Fast ICA using eight-order Newton's method was proposed for BSS. The separation performance of the proposed algorithm was almost the same as the Fast ICA algorithm. Besides that, the mixed sound signal was separated by new algorithm with fewer iteration and faster convergence than the Fast ICA.

REFERENCES

Amari, S., A. Cichocki and H. Yang, 1996. A new learning algorithm for blind separation of sources. Adv. Neural Inform. Process. Syst., 8: 757-763.

Bell, A. and T. Sejnowski, 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comp., 7: 1129-1159.

Bi, W., H. Ren and Q. Wu, 2008. New family of seventh-order methods for nonlinear equations. Appl. Math. Comp., 203: 408-412.

Cardoso, J.F., 1998. Blind signal separation: Statistical principles. Proceeding of the IEEE, 86: 2009-2025.

Chun, C., 2007. Some improvements of Jarratt's method with sixth-order convergence. Appl. Math. Comp., 190: 1432-1437.

Cover, T.M. and J.A. Thomas, 1991. Elements of Information Theory. Publisher John Wiley & Sons, Hoboken, pp: 640, ISBN: 0471748811.

Cichochi, A. and S. Amari, 2004. Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. Wiley, New York.

Chio, S., A. Cichocki and H.M. Park, 2005. Blind source separation and independent component analysis: A review. Neural Inform. Process. Lett. Rev., 6(1): 1-57.

Hyvarinen, A., 1998. New approximations of differential entropy for independent component analysis and projection pursuit. Adv. Neural Inform. Proc. Syst., 10: 273-279.

Hyvarinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE T. Neural Networ., 10(3): 626-634.

Hevarinen, A., J. Karhunen and E. Oja, 2001. Independent Component Analysis. John Wiley & Sons, New York.

Stone, J.V., 1999. Independent Component Analysis a Tutorial Introduction. The MIT Press, Cambridge, Mass, pp: 193, ISBN: 0262693151.

King, R., 1973. A family of fourth order methods for nonlinear equations, SIAM J. Numer. Anal., 10: 876-879.

Kou, J., 2007. The improvements of modified Newton's method. Appl. Math. Comp., 189: 602-609.

Lee, T.W., M. Girolami and T.J. Sejnowski, 1999. Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super- Gaussian source. Neural Comp., 11: 417-441.

Li, Z., C. Peng, T. Zhou and J. Gao, 2011. A new newton-type method for solving nonlinear equations with any integer order of convergence. J. Comp. Inform. Syst., 7: 2371-2378.

Parhi, S.K. and D.K. Gupta, 2008. A sixth order method for nonlinear equations. Appl. Math. Comput., 203: 50-55.

Pham, D.T., 2004. Fast algorithms for mutual information based independent component analysis. IEEE T. Signal Proc., 52(10).

Potra, F.A. and V. Ptak, 1984. Nondiscrete Induction and Iterative Processes. Pitman, Boston, pp: 207, ISBN: 0273086278.

Papoulis, A., 1991. Probability, Random Variables and Stochastic Processes. 3rd Edn., McGraw-Hill, New York.

Shia, Z., H. Tanga and Y. Tang, 2004. A new fixed-point algorithm for independent component analysis. Neurocomputing, 56: 467-473.

Shi, Z. and C. Zhang, 2006. Energy predictability to blind source separation. Elec. Lett., 42(17): 1006-1008.

Ye, Y., Z.L. Zhang, S. Wu and X. Zhou, 2006. Improved multiplicative orthogonal-group based ICA algorithm for separating mixed sub-Gaussian and super-Gaussian sources. Proceeding of International Conference on Communications, Circuits and Systems, Guilin, China (ICCCAS), pp: 340-343.