

## Model of independent component analysis with reference curve and its application in removing artifact from electroencephalograph

Lizhi Cui<sup>1,2,a</sup>, Zhihao Ling<sup>1,b</sup>, Josiah Poon<sup>2</sup>, Simon K. Poon<sup>2</sup>

<sup>1</sup>Key Laboratory of Advanced Control and Optimization for Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai, 200237, China

<sup>2</sup>School of information technologies, University of Sydney, Sydney, 2006, Australia

<sup>a</sup>email: andyclzh@yahoo.com, <sup>b</sup>email:zhhliling@mail.ecust.edu.cn

**Keywords:** Artifacts removal; Electroencephalogram; Independent Component Analysis with Reference Curve

**Abstract.** This paper proposes a model named Independent Component Analysis with Reference Curve (ICARC) to extract and remove artifact signal from Electroencephalogram (EEG). Firstly, an additional requirement and a priori information are introduced directly into the contrast function of the traditional ICA model. Then, an augmented Lagrangian function is formed based on this new model. Finally, the iterative solution is calculated by using the Newton iterative method. The simulations and experiments are implemented to indicate the performance of our model comparing with other method. The results show that: 1) more stable results are given by our model; 2) higher precision is obtained in the results by the ICARC model.

### Introduction

The electroencephalogram (EEG) is a useful diagnostic signal for the clinical conditions [1], but with problem in reading and analyzing due to the artefacts such as eye blinking, eye movement and heartbeat [2]. It is difficult to reduce these artifacts by the traditional filtering methods due to their high amplitude and overlapping frequency band with the EEG signal [3]. The regression-based methods [4] and adaptive filtering methods [5] can reduce artifacts, but the measurement of reference channels is not always available. The Principal component Analysis (PCA) method decomposes the raw EEG signal to components, and reconstructs the clean EEG by eliminating the artifact components [6]. But the assumption that the components are algebraically orthogonal is difficult to be fulfilled. The Independent Component Analysis (ICA) is applied in the EEG decomposition, but with two major insufficiencies: 1) the noise and artifacts bring too many useless components; 2) the noise or artifacts components need to be identified by other methods. By introducing a priori information to the ICA model, a method named constrained independent component analysis (cICA) model is proposed in literature [8]. Later, the paper [9] proposes an improved cICA method to obtain more accurate results. However, the results of the cICA model vary among different implementations. This paper proposes a method named ICA constrained by Reference Curve (ICARC) to overcome the drawbacks of the cICA model.

### The principle of the ICARC model

The classical ICA model can be expressed as the following linear relationship:

$$\mathbf{X}_{cxt} = \mathbf{A} \times \mathbf{S} \quad (1)$$

where,  $\mathbf{X}_{cxt} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_c]^T$  is the EEG signal;  $c$  represents the detected channels;  $t$  represents the sample points;  $\mathbf{A}$  is an mixing matrix;  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m]^T$  is the independent components.  $\mathbf{X}$  is assumed with zero average row vectors, which is always not fulfilled. So a preprocessing [10] is applied as

$$\begin{cases} \mathbf{X} = \bar{\mathbf{X}} + \hat{\mathbf{X}} = [\bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \cdots & \bar{\mathbf{x}}_c]^T + [\hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_2 & \cdots & \hat{\mathbf{x}}_c]^T \\ \tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_2 & \cdots & \tilde{\mathbf{x}}_t] = \mathbf{D}_w \times \bar{\mathbf{X}} \end{cases} \quad (2)$$

where,  $\bar{\mathbf{x}}_i, (i = 1, 2, \dots, c)$  are the vectors filled with the mean of  $\mathbf{x}_i, (i = 1, 2, \dots, c)$  in equation (1);  $\hat{\mathbf{x}}_i, (i = 1, 2, \dots, c)$  are vectors subtracting  $\bar{\mathbf{x}}_i$  from  $\mathbf{x}_i$ ;  $\tilde{\mathbf{X}}$  is the matrix after preprocessing;  $\tilde{\mathbf{x}}_i, (i = 1, 2, \dots, t)$  satisfy  $E\{\tilde{\mathbf{x}}_i \times \tilde{\mathbf{x}}_i^T\} = \mathbf{I}$ ;  $\mathbf{D}_w$  is called whiten matrix [10]. Based on the principle of the ICA model, the ICARC model is proposed as

$$\begin{cases} \max E\{G(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} - \|\bar{\mathbf{b}}^T \times \tilde{\mathbf{X}}^1 - \mathbf{r}\|_2^2; \quad \text{subject to } E\{(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)^2\} = \|\mathbf{b}\|_2^2 = 1 \\ \bar{\mathbf{b}}^T = [\mathbf{b}^T, d_i]; \quad \tilde{\mathbf{x}}^0 = [\tilde{\mathbf{x}}; 0]; \quad \tilde{\mathbf{X}}^1 = [\tilde{\mathbf{x}}^1 & \tilde{\mathbf{x}}^2 & \cdots & \tilde{\mathbf{x}}^t]; \quad \tilde{\mathbf{x}}_i^1 = [\tilde{\mathbf{x}}_i; 1], i = 1, 2, \dots, t \end{cases} \quad (3)$$

where,  $\bar{\mathbf{b}}$  is the vector wanted;  $\mathbf{r}$  is the reference curve for  $\mathbf{s}_i$  in equation (1);  $\tilde{\mathbf{x}}$  represents all the column vector in  $\tilde{\mathbf{X}}$ ;  $G(\square)$  is a nonlinear function [10]. According to the Karush-Kuhn-Tucher (KKT) condition [11], the Lagrangian function for equation (3) is given as

$$L(\bar{\mathbf{b}}, \beta) = E\{G(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} - \|\bar{\mathbf{b}}^T \times \tilde{\mathbf{X}}^1 - \mathbf{r}\|_2^2 + \beta \times (\|\mathbf{b}\|_2^2 - 1) \quad (4)$$

where,  $\beta$  is the Lagrange multiplier. The solution of  $\bar{\mathbf{b}}$  should satisfy following requirement.

$$F = \frac{\partial L}{\partial \bar{\mathbf{b}}} = E\{\tilde{\mathbf{x}}^0 \times g(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} - \sum_{j=1}^t 2 \times \tilde{\mathbf{x}}_j^1 \times (\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}_j^1 - \mathbf{r}_j) + \beta \times \mathbf{b} = 0 \quad (5)$$

where,  $g(\square)$  is the derivative of  $G(\square)$ ;  $t$  is the maximum sample point;  $\mathbf{r}_j$  represents the  $j$ th element of  $\mathbf{r}$ . In order to solve equation (5), its Jacobian matrix is calculated firstly as

$$JF = E\{\tilde{\mathbf{x}}^0 \times \tilde{\mathbf{x}}^{0T} \times g'(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} - \sum_{j=1}^t 2 \times \tilde{\mathbf{x}}_j^1 \times \tilde{\mathbf{x}}_j^{1T} + \beta \times \mathbf{I} = E\{g'(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} \times \mathbf{I} - 2 \times t \times \mathbf{I} + \beta \times \mathbf{I} \quad (6)$$

where,  $E\{\tilde{\mathbf{x}}^0 \times \tilde{\mathbf{x}}^{0T}\} = E\{\tilde{\mathbf{x}} \times \tilde{\mathbf{x}}^T\} = \mathbf{I}$ ;  $\sum_{j=1}^t 2 \times \tilde{\mathbf{x}}_j^1 \times \tilde{\mathbf{x}}_j^{1T} = 2 \times t \times \mathbf{I}$ ;  $g'(\square)$  is the derivative of  $g(\square)$ . The iterative solution can be obtained by equation (7) by using Newton iteration method [12].

$$\bar{\mathbf{b}}^+ = \bar{\mathbf{b}} - \frac{F}{JF} = \frac{\bar{\mathbf{b}} \times E\{g'(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} - \tilde{\mathbf{x}}^0 \times g(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0) + 2 \times \bar{\mathbf{b}} + \sum_{j=1}^t 2 \times \tilde{\mathbf{x}}_j^1 \times (\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}_j^1 - \mathbf{r}_j)}{E\{g'(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} - 2 \times t + \beta} \quad (7)$$

Because the denominator is a scalar, equation (7) can be replaced by

$$\begin{cases} \bar{\mathbf{b}}^+ = \bar{\mathbf{b}} \times E\{g'(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0)\} - \tilde{\mathbf{x}}^0 \times g(\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}^0) + 2 \times \bar{\mathbf{b}} + \sum_{j=1}^t 2 \times \tilde{\mathbf{x}}_j^1 \times (\bar{\mathbf{b}}^T \times \tilde{\mathbf{x}}_j^1 - \mathbf{r}_j) \\ \bar{\mathbf{b}}^* = \bar{\mathbf{b}}^+ / \|\bar{\mathbf{b}}^+\| \end{cases} \quad (8)$$

With the optimal value  $\bar{\mathbf{b}}^*$ , a signal according to the reference curve  $\mathbf{r}$  will be given by

$$\mathbf{y} = \bar{\mathbf{b}}^* \times \tilde{\mathbf{X}}^1 \quad (9)$$

where,  $\tilde{\mathbf{X}}^1$  is shown in equation (3). Because  $\mathbf{y}$  is independent to other signals. It can be removed from each observed channel  $\mathbf{x}_i, (i = 1, 2, \dots, c)$  as shown in equation (10).

$$\begin{cases} \mathbf{X}_{clc} = \mathbf{X} - [a_1 & a_2 & \cdots & a_c]^T \times \mathbf{y} \\ a_i = \rho(\mathbf{x}_i^T, \mathbf{y}), i = 1, 2, \dots, c \end{cases} \quad (10)$$

where,  $\mathbf{X}_{clc}$  represents the clean EEG data set;  $\rho(\mathbf{x}_i^T, \mathbf{y})$  is the correlation coefficient operator.

## Simulations, experiments and discussions

Figure 1 (a) gives the four synthetic signals. Figure 1 (a) gives the mixed data set. Figure 1 (c)

gives the reference signals, which are obtained from the sign of the four original sources. Four signals are extracted by the ICARC and cICA model for 10 times respectively, the results are list in table 1. From the results we can see:

1) The results given by the ICARC model are consistent among the 10 times, but the results given by the cICA model are not. All the errors given by the ICARC model are much smaller.

2) The time used by the ICARC model is much less than that by the cICA model. The average time used by the ICARC model is 0.0025 Seconds, while the cICA model uses 0.5 seconds.

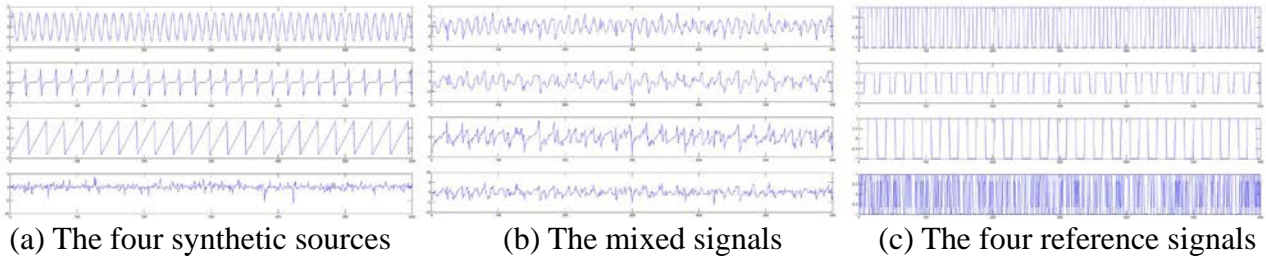


Fig.1. the simulation data set

Table 1. The results of the simulations. The ‘error<sub>*i*</sub>’ is the error between the *i*<sup>th</sup> synthetic signal and the calculated signal. The ‘time’ are the time used. The ‘Ave’ is the mean value of the ten simulations.

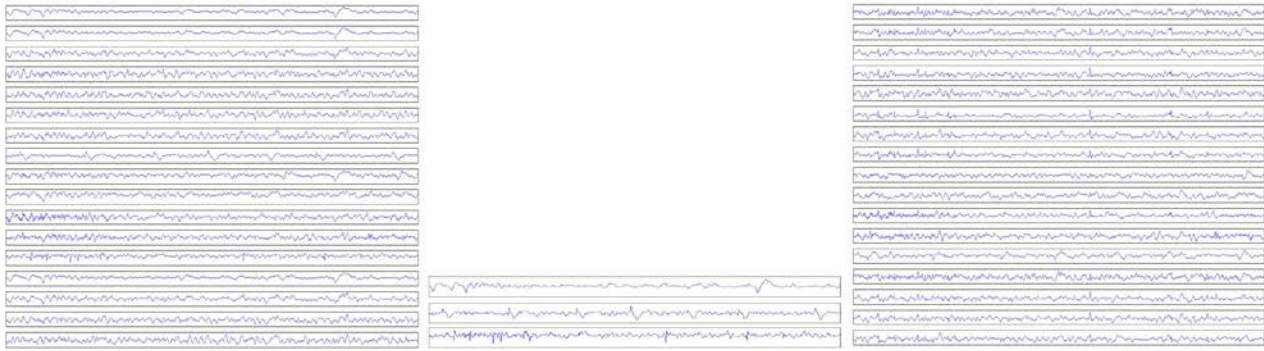
No.	1	2	3	4	5	6	7	8	9	10	Ave	
cICA	error <sub>1</sub>	5.4	5.0	4.6	19.6	5.0	4.9	23.6	4.9	5.8	23.4	10.2
	error <sub>2</sub>	4.2	0.8	0.6	4.7	1.5	0.8	0.9	4.0	3.6	0.8	2.2
	error <sub>3</sub>	22.4	22.6	16.3	16.5	4.0	5.6	1.8	9.8	5.7	2.8	10.7
	error <sub>4</sub>	1.5	5.5	2.0	4.7	5.6	4.9	4.8	1.7	1.7	1.5	3.4
	time(s)	0.3	0.4	0.7	0.5	0.2	0.4	0.9	0.5	0.4	0.7	0.5
ICARC	error <sub>1</sub>	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	error <sub>2</sub>	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	error <sub>3</sub>	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	error <sub>4</sub>	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64
	time(s)	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.001	0.001	0.0025

Figure 2 (a) shows the 17-channel EEG with the artifacts. Let the seriously contaminated channels (channel 1, 8 and 13) be adopted as the reference signals for the eye blinking artifact, the heart artifact and the eye movement artifact respectively. Figure 2 (b) shows the three calculated artifact signals. Figure 2 (c) shows the clean EEG signals. Ten experiments are implemented respectively. The results are list in Table 2. From the results we can see:

1) The time used by the ICARC model for every artifact extraction is much less.

2) The time and step used by the ICARC model is more stable than that by the cICA model.

3) The errors between the individual curve and the mean curve of the ten experiments are small.



(a) The EEG signal with artefacts (b) Three artifact components (c) The clean EEG signals  
Fig.2. The experiment data set

## Conclusion

This paper proposed the ICARC model to extract the artifact signals from the EEG data set by introducing a priori knowledge into the contrast function of the ICA model directly. Through the simulations and experiments, following conclusions are drawn: 1) the results given by the ICARC model were consistent; 2) the errors given by the ICARC model were much smaller; 3) the time used by the ICARC model was much less; 4) the ICARC model are practical for EEG data set.

## Acknowledgement

Lizhi Cui thanks Chinese Scholarship Council to provide financial support during his study in the University of Sydney (The students number is 201206740061).

Table 2. The results of the experiments. The ‘time<sub>*i*</sub>’ and ‘step<sub>*i*</sub>’ are the time and step used for the *i*<sup>th</sup> artefact extraction. The ‘error<sub>*i*</sub>’ is the error between *i*<sup>th</sup> calculated artifact signal and the mean cure of the ten results. The ‘Ave’ is the mean value of the ten experiments. The ‘Std’ is the standard derivation of the ten experiments. The column of ‘Factor’ is the order of the magnitude.

No.	1	2	3	4	5	6	7	8	9	10	Ave	Std	Factor	
cICA	time <sub>1</sub> (s)	1.9	0.9	1.1	0.06	1.3	1.8	2.6	0.05	0.06	0.06	0.9	0.9	$\times 10^{-1}$
	step <sub>1</sub>	289	187	218	11	254	339	503	11	12	11	183.5	170.7	$\times 10^0$
	error <sub>1</sub>	17.1	16.8	16.8	35.8	16.8	16.6	16.5	34.2	39.9	32.5	24.3	9.9	$\times 10^0$
	time <sub>2</sub> (s)	1.1	1.6	0.08	0.1	0.08	1.7	3.3	0.9	0.09	0.09	0.9	1	$\times 10^{-1}$
	step <sub>2</sub>	120	198	11	11	11	224	390	107	13	12	109.7	128	$\times 10^0$
	error <sub>2</sub>	38.4	35.9	39.7	23.5	31.2	24.4	40.0	24.5	23.7	39.5	32.1	7.4	$\times 10^0$
	time <sub>3</sub> (s)	5.6	4.2	0.05	1.5	0.06	0.98	0.06	4.4	4.1	0.06	2.1	2.2	$\times 10^{-1}$
	step <sub>3</sub>	877	808	11	291	11	186	11	843	803	11	385.2	396.3	$\times 10^0$
error <sub>3</sub>	27.2	27.2	33.1	27.8	44.8	44.1	36.6	27.4	27.8	43.9	34.0	7.7	$\times 10^0$	
ICARC	time <sub>1</sub> (s)	10	9.5	10	9.6	9.3	6.5	6.2	11	10	11	9.3	1.7	$\times 10^{-4}$
	step <sub>1</sub>	4	4	4	4	4	3	3	4	4	4	3.8	0.4	$\times 10^0$
	error <sub>1</sub>	0.75	0.75	0.75	0.75	0.75	3.1	3.0	0.75	0.75	0.75	1.22	0.98	$\times 10^{-7}$
	time <sub>2</sub> (s)	11	11	12	8.4	6.7	6.9	11	11	11	6.9	9.6	2.1	$\times 10^{-4}$
	step <sub>2</sub>	4	4	4	3	3	3	4	4	4	3	3.6	0.52	$\times 10^0$
	error <sub>2</sub>	2.1	2.1	2.1	3.1	3.0	3.2	2.1	2.1	2.1	3.5	2.5	0.57	$\times 10^{-7}$
	time <sub>3</sub> (s)	7.9	6.2	7.3	7.3	7.6	9.5	6.7	5.9	5.9	8.9	7.3	1.2	$\times 10^{-4}$
	step <sub>3</sub>	3	3	3	3	3	4	3	3	3	4	3.2	0.4	$\times 10^0$
error <sub>3</sub>	0.8	1.9	0.4	0.9	0.9	1.2	1.8	0.6	0.7	1.2	1.0	0.5	$\times 10^{-7}$	

## References

- [1] Saeid Sanei, J.A. Chambers. EEG signal processing. Wiley, New York, 2007.

- [2] S. Boudet, L. Peyrodie, G. Forzy, A. Pinti, H. Toumi, P. Gallois. Improvement of Adaptive Filtering by Optimal Projection to filter different artifact types on long duration EEG recordings. *Computer Methods and Programs in Biomedicine*, 2012 108 234-249.
- [3] J. Mateo, A.M. Torres, M.A. Garcia. Dynamic Fuzzy Neural Network Based Learning Algorithms for Ocular Artefact Reduction in EEG Recordings, *Neural Process Lett*, 2014 39 45-67.
- [4] Rodney J. Croft, Robert J. Barry. EOG correction: which regression should we use? *Psychophysiology*, 2000 37 123-125.
- [5] Mehrdad Fatourehchi, Ali Bashashati, Rabab K. Ward, Gary E. Birch. EMC and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 2007 118 480-494.
- [6] P. He, G. Wilson, C. Russell. Removal of ocular artefacts from electroencephalogram by adaptive filtering. *Medical and Biological Engineering and Computing*, 2004 42 407-412.
- [7] Shunichi Amari, Andrzej Cichocki. Adaptive Blind Signal Processing-Neural Network Approaches. *Proceedings of the IEEE*, 1998 86 2026-2048.
- [8] Wei Lu, Jagath C. Rajapakse. Approach and Applications of Constrained ICA. *IEEE Transactions on Neural Networks*, 2005 16 203-212.
- [9] Deshuang Huang, Jianxun Mi. A new Constrained Independent Component Analysis Method. *IEEE Transactions on Neural networks*, 2007 18 1532-1535.
- [10] Aapo Hyvarinen, Erkki Oja. *Independent Component Analysis: Algorithms and Applications*. Neural networks, 2000 13 411-430.
- [11] Stephen Boyd, Lieven Vandenberghe. *Convex Optimizaiton*. Cambridge University Press, Cambridge, 2009.
- [12] David G. Luenberger. *Optimization by vector space methods*. John Wiley and Sons, New York, 1969.