Online Recursive Independent Component Analysis for Real-time Source Separation of High-density EEG

Sheng-Hsiou Hsu, *Student Member, IEEE,* Tim Mullen, *Member, IEEE,* Tzyy-Ping Jung, *Senior Member, IEEE,* and Gert Cauwenberghs, *Fellow, IEEE*

***Abstract*— Online Independent Component Analysis (ICA) algorithms have recently seen increasing development and application across a range of fields, including communications, biosignal processing, and brain-computer interfaces. However, prior work in this domain has primarily focused on algorithmic proofs of convergence, with application limited to small 'toy' examples or to relatively low channel density EEG datasets. Furthermore, there is limited availability of computationally efficient online ICA implementations, suitable for real-time application. This study describes an optimized online recursive ICA algorithm (ORICA), with online recursive least squares (RLS) whitening, for blind source separation of high-density EEG data. It is implemented as an online-capable plugin within the open-source BCILAB (EEGLAB) framework. We further derive and evaluate a block-update modification to the ORICA learning rule. We demonstrate the algorithm's suitability for accurate and efficient source identification in high density (64-channel) realistically-simulated EEG data, as well as real 61-channel EEG data recorded by a dry and wearable EEG system in a cognitive experiment.**

I. INTRODUCTION

Independent Component Analysis (ICA), as a means for blind source separation, has enjoyed great success in biosignal processing and communications [1]. In biomedical applications, such as electroencephalography (EEG), the application of ICA is justified by the reasonable assumption that multi-channel scalp EEG signals arise as a mixture of weakly dependent non-Gaussian sources [2]. In particular, offline ICA methods have been widely used for separating artifacts such as eye blinks and muscle activity [3], as well as used to extract and study activity generated within the brain [4]. However, for many real-world applications, including real-time functional neuroimaging and brain-computer interfaces (BCI) [5], online source separation methods are needed. Desirable properties include fast convergence and real-time computational performance.

Several online ICA algorithms have been proposed. Amongst the most promising candidates are recursive-least-squares (RLS) type algorithms, extended from iterative natural gradient optimization of independence-maximizing objective functions [6][7]. Akhtar et al [8] proposed an RLS-type Online Recursive ICA algorithm (ORICA), derived as a fixed-point solution to the widely-used natural gradient Infomax ICA learning rule. Infomax ICA has been shown to outperform most alternative ICA algorithms, in terms of maximizing independence and biological plausibility of EEG sources [2]. ORICA builds on an iterative inversion formula, yielding faster convergence and lower computational load than the alternatives. However, as with other RLS-type algorithms, stability, convergence speed, and computational load are important practical factors to consider.

This study utilizes two approaches to improve performance. We combine an optimized implementation of ORICA with the online RLS whitening filter of [7]. We also derive a multiple measurement vector (MMV) block-update rule to increase processing speed without sacrificing performance. The proposed online ICA pipeline is implemented in MATLAB as a BCILAB plugin [9]. Real-time performance capability, accuracy, convergence speed, and scalability of the pipeline is analyzed on a realistic simulation of 64-channel EEG data. Finally, we demonstrate real-world applicability of the pipeline for online source separation, with quantitative and qualitative comparison to EEGLAB's "gold standard" implementation of Extended Infomax ICA [10], using 61-channel dry, wearable EEG data recorded from a subject performing an Eriksen Flanker task.

II. METHODS

We assume the standard ICA generative model $x = As$, where x are scalp EEG observations, s are unknown sources, and A is an unknown N-by-N mixing matrix. The objective is to learn an unmixing (weight) matrix $W = A^{-1}$ such that the sources are recovered by $y = Wx$.

A. *Online recursive-least-squares (RLS) whitening*

Whitening (decorrelating) the data reduces the number of independent parameters ICA must learn, and can improve convergence [1]. In order to fit in the online pipeline with online RLS-type ICA, we use the similar online RLS whitening algorithm proposed by [7]:

$$M_{n+1} = \frac{1}{1-\lambda_n}[I - \frac{v_n v_n^T}{\frac{1-\lambda_n}{\lambda_n} + v_n^T v_n}]M_n \tag{1}$$

where M_n is the whitening matrix, $v_n = M_n x_n$ is the whitened data, and λ_n is the forgetting factor.

As shown in [7], the RLS-type filter converges faster than a least-mean-squares type filter, e.g. running average

This work was in part by a gift by the Swartz Foundation (Old Field, NY), by the Army Research Laboratory under Cooperative Agreement Number W911NF-10-2-0022, by NIH grant 1R01MH084819-03 and NSF EFRI-M3C 1137279.

S.-H. Hsu (shh078@ucsd.edu) is with Dept. of Bioengineering (BIOE), Swartz Center for Computational Neuroscience (SCCN), and Institute for Neural Computation (INC) of University of California, San Diego (UCSD).

T. Mullen (tmullen@ucsd.edu) is with Dept. of Cognitive Science, SCCN and INC of UCSD.

T.-P. Jung (tpjung@ucsd.edu) is with BIOE, SCCN and INC of UCSD.

G. Cauwenberghs (gert@ucsd.edu) is with BIOE and INC of UCSD.

of covariance matrix. Also, since online RLS whitening and ORICA have a similar recursive form and adaptation property, e.g. forgetting rate, they can be easily combined.

B. Online recursive ICA (ORICA)

The ORICA algorithm derives from the general incremental update form of the well-known natural gradient learning rule for Infomax ICA:

$$W_{n+1} = W_n + \eta[I - f(y_n) \cdot y_n^T]W_n$$

where $y_n = W_n x_n$, η is the learning rate, and $f(\cdot)$ is a nonlinear projection function. In the limit of a small η and assuming a fixed f, the convergence criterion $\langle f(y) \cdot y^T \rangle = I$ leads to a fixed-point solution in an iterative inversion form:

$$A_{n+1} = (1 - \lambda_n)A_n + \lambda_n x_n \cdot f_n^T \qquad (2)$$

where A_n is the pseudo-inverse of W_n and λ_n is the forgetting factor for an exponentially weighted series of updates (note that λ_n differs from η, which is the step size for stochastic gradient optimization).

Applying the Sherman-Morrison matrix inversion formula to Eq. 2, the final online recursive learning rule becomes [8]:

$$W_{n+1} = \frac{1}{1 - \lambda_n}\Big[I - \frac{y_n \cdot f^T(y_n)}{\frac{1-\lambda_n}{\lambda_n} + f^T(y_n) \cdot y_n}\Big]W_n \qquad (3)$$

Eq. 3 of ORICA is similar to Eq. 1 of online RLS whitening, albeit with nonlinear projection $f(\cdot)$ ensuring independence of sources. ORICA can thus be understood as a nonlinear form of online RLS whitening.

Following [8], component-wise nonlinear functions are $f(y) = -2\tanh(y)$ for super-Gaussian sources and $f(y) = \tanh(y) - y$ for sub-Gaussian sources. Also, the heuristic time-varying forgetting factor is used:

$$\lambda_n = \frac{\lambda_0}{n^\gamma}$$

where λ_0 is an initial forgetting factor and γ determines the exponential decay rate of λ.

1) Number of sub- and super-Gaussian sources: While approaches for adaptively selecting f within ORICA have been proposed [8], these are heuristic and presently lack convergence proofs. In practice, we find that both convergence and run-time performance are improved by preassuming a fixed number of sub- and super-Gaussian sources. A more extensive characterization of the performance of ORICA with heterogeneous source distributions is beyond the scope of this report, and will be the subject of a forthcoming paper.

2) Block-update rule: Performing updates for each sample can be costly. To reduce the computational load and ensure consistent real-time performance, we update the weight matrix for a short block of samples at once. To achieve this without loss of accuracy, we solve Eq. 3 for time index $l = n$ to $l = n + L - 1$, assuming y_l is approximated as $W_n x_l$ and λ_l is small. This leads to a block-update rule:

$$W_{n+L} \approx \Big(\prod_{l=n}^{n+L-1} \frac{1}{1 - \lambda_l}\Big) \cdot \Big[I - \sum_{l=n}^{n+L-1} \frac{y_l \cdot f^T(y_l)}{\frac{1-\lambda_l}{\lambda_l} + f^T(y_l) \cdot y_l}\Big]W_n \qquad (4)$$

TABLE I

LIST OF PARAMETERS FOR THE ONLINE PIPELINE: (A) IIR HIGH-PASS FILTER, (B) ONLINE RLS WHITENING FILTER, AND (C) ONLINE RECURSIVE ICA FILTER.

Filters	Parameters	Values	Description
A	BW	0.2–2 Hz	Transition band
B,C	λ_0	0.995	Initial forgetting factor
	γ	0.60	Decay rate of forgetting factor
	L	16	Block-update size
C	n_{sub}	0 (Sim. EEG) 1 (Real EEG)	Number of subgaussian sources

In this form, the sequence of updates can be vectorized for fast MATLAB computation. Note that Eq. 4 appropriately accounts for the decaying forgetting factor at each time point. This keeps the approximation error to a minimum.

III. MATERIALS

A. Data collection

1) Simulated EEG data: We used the SIFT EEG simulation module, with an approach similar to [11]. We generated 64 super-Gaussian independent source time-series from stationary and random-coefficient order-3 autoregressive models (300Hz sampling rate, 10-min), assigned each source a random cortical dipole location, and projected these through a zero-noise 3-layer BEM forward model (MNI "Colin27"), yielding 64-channel EEG data.

2) Real EEG data: Two sessions of high-density EEG data were collected from a 24 year-old right-handed male subject using a 64-channel wearable wireless dry EEG headset (Cognionics, Inc). The first session was a 10-min resting session. In the second session, the subject performed a modified Eriksen Flanker task [12] with a 133 ms delay between flanker and target presentation for 20 minutes. Flanker tasks are known to produce robust error-related negativity (ERN, Ne) at frontal-central electrode sites. Our goal was to extract these ERP components from high-density EEG data in a real-world setting using the proposed online ICA pipeline.

B. Online ICA pipeline

Simulated and real EEG data were streamed into MATLAB and analyzed in a simulated online environment using BCILAB, an open source MATLAB toolbox designed for B-CI research [9][13]. The pipeline for simulated data consisted of three filters: a Butterworth IIR high-pass filter, an online RLS whitening filter, and an ORICA filter. The high-pass filter removes trend and low-frequency drift, ensuring the zero-mean criterion for ICA is satisfied. For both datasets, we chose block size $L = 16$ to demonstrate the accuracy of block-update. We set the number of sub-Gaussian sources to zero for simulated EEG data and one for real EEG data, allowing for 60Hz line noise. Table I summarizes the parameters of the three filters.

C. Processing of real EEG data

For Flanker task EEG data, we applied additional processing steps and techniques. Firstly, an automatic removal of bad (e.g. flatlined or abnormally correlated) channels was

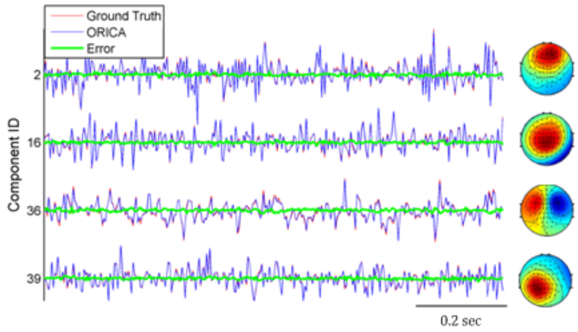

Fig. 1. Source dynamics and the corresponding component maps of four randomly selected components reconstructed by ORICA at the end of the time series (blue) superimposed on ground truth (red) with error, i.e. difference, (green) on simulated EEG data.

applied prior to the online pipeline, using BCILAB routines, removing 3 channels. Secondly, we warm-started (initialized) all filters using the first 3 minutes EEG data in the resting session. This step (offline) costs little computation time and used no data from the separate Flanker task session, while accelerating ORICA convergence. Following application of the pipeline, response-locked event-related potentials (ERPs) were analyzed offline in EEGLAB [14]. IC time-series (20-minute session) were epoched around responses in a -400 to 600 ms window, yielding 693 epochs (104 error trials, 589 correct). Error trials were then averaged to produce ERPs.

IV. Results

A. Simulated 64-ch stationary EEG data

1) Evaluation of the decomposed components: Fig. 1 shows a 1-sec segment of reconstructed ORICA IC time-series (at convergence) superimposed on ground truth. The error is close to zero.

Fig. 2(a) shows the correlation between spatial filters learned by ORICA and their ground-truth counterparts, with component matching via the Hungarian method. All correlations approach 1 by the end of the 10-minute session, albeit with systematic variation in convergence speed. A common empirical heuristic for the number of training samples required for separating N stable (Infomax) ICA sources is kN^2, where $k > 25$ [15]. For 64 channels, learning requires $64 \times 64 \times 25 = 102400$ samples $= 5.7$ mins. We observe that 77% (91%) of ICs reach a correlation of 0.95 (0.8) within 5.7 minutes (vertical dotted line).

Fig. 2(b) shows the evolution of the spatial filter of a randomly selected IC (#29), and its correlation with ground truth. This IC converged to a steady-state correlation of 0.95 in 4 minutes. A global performance index is superimposed in green.

For a known N-by-N mixing matrix A, the performance index E characterizes the ICA global convergence property, defined in [16]:

$$E = \frac{1}{N-1}\left[N - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\max\limits_{1\le j\le m}|C_{ij}(n)|^2}{\sum_{j=1}^{N}|C_{ij}(n)|^2} + \frac{\max\limits_{1\le j\le m}|C_{ji}(n)|^2}{\sum_{j=1}^{N}|C_{ji}(n)|^2}\right)\right] \quad (5)$$

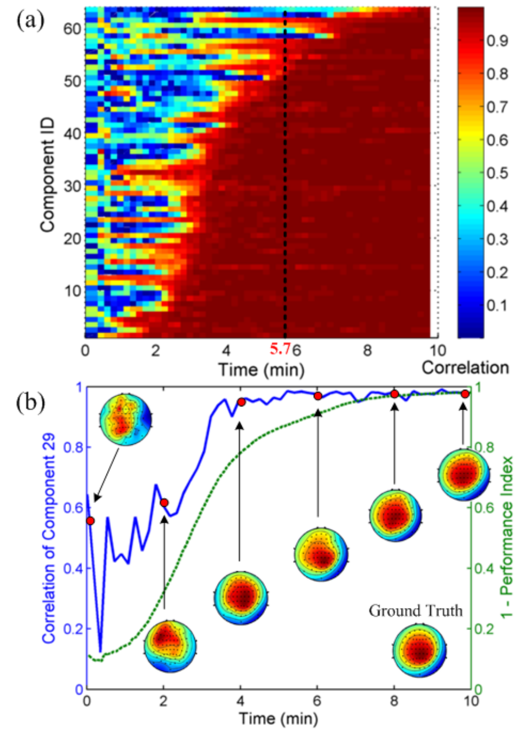

Fig. 2. (a) Evolution of component-wise correlation between ORICA-decomposed ICs and ground truth on simulated 64-ch EEG data. ICs sorted with respect to time required to reach a correlation of 0.95. The dotted line (5.7 min) is the heuristic time for separating 64 stable ICs. (b) Evolution of correlation (blue) and spatial filter of a randomly selected IC (#29). One minus the performance index (green) is superimposed.

TABLE II

Averaged execution time for 1-second (300 samples) 64-ch data using (A) IIR high-pass filter, (B) online RLS whitening filter , and (C) ORICA filter with different block sizes.

Online filters	A	B		C				
Block size	–	1	2	1	2	4	8	16
Averaged execution time (ms)	1.1	2.4	1.3	460	220	110	54	42

Run in MATLAB 2012a on a dual-core 2.50GHz Intel Core i5-3210M CPU with 8GB RAM.

where $C(n) = W(n)M(n)A$ and M is the whitening matrix. At convergence, C results in a permuted and scaled identity matrix, and the performance index is close to zero. The evolution of 1 minus the performance index provided further proof of convergence. We note the overall trend similarity to both individual correlation profiles (e.g. IC29) and to the percentage of converged ICs in Fig. 2(a).

2) Computational load: Table. II shows the average execution time (MATLAB Profiler) required to apply the entire pipeline to 1-second of data. Runtime was uniformly less than one second, demonstrating real-time capability. Note the execution time of the ORICA filter is nearly halved as block size doubles when $L \leq 8$, with diminishing returns for $L \geq 16$. We confirmed on simulated data that the approximation error in Eq. 4 was negligible for all examined block sizes.

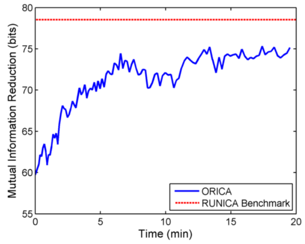

Fig. 3. Evolution of mutual information reduction of sources reconstructed by ORICA (blue) and the benchmark provided by RUNICA on real 61-ch EEG data from the Flanker Task.

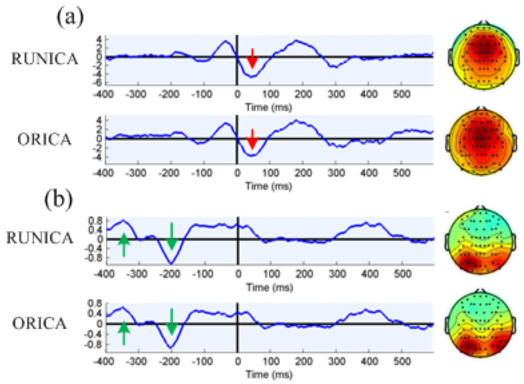

Fig. 4. Component maps and the corresponding ERPs of (a) frontal-central and (b) occipital components reconstructed by ORICA and RUNICA on 61-ch flanker task EEG data. Red arrows denote ERN while green arrows denote VEPs

B. Real 61-ch EEG data from the Flanker task

Since ground truth is unknown for real data, we adopted as a "gold standard" the offline Extended Infomax ICA algorithm [10], as implemented in the EEGLAB [14] function RUNICA. The robustness and stability of this algorithm on high-density EEG data has been shown to outperform most blind source separation algorithms [2].

Fig. 3 compares ORICA and RUNICA in terms of the mutual information reduction (MIR) in decomposing channel data into ICs [2]. At convergence, ORICA achieved an MIR near that of RUNICA (75 vs. 79 bits). RUNICA's runtime was 28.6 minutes with 512 passes through the data, versus a single, < 1 minute run for ORICA. Note ORICA MIR does not start at 0 bits, due to warm-starting (Sec. III).

As a further qualitative benchmark, we examined a set of ORICA and RUNICA ICs with stereotypical fronto-central and occipital spatial topographies (regions involved in stimulus detection and error processing). Fig. 4 (a) demonstrates that the ORICA- and RUNICA-decomposed fronto-central IC topographies and classic ERN ERPs are comparable and consistent with previous studies [13][17]. Fig. 4 (b) shows that averaged occipital visual evoked potentials (VEPs) elicited by flanker (left green arrows) and target (right green arrows) presentation are clearly observed using both methods. This demonstrates ORICA's efficacy in extracting informative components and corresponding ERPs in high-density, real-world EEG data.

V. Conclusions

This study proposed two procedures to achieve fast convergence and real-time application of online ICA: (1) combining an optimized implementation of ORICA with online RLS whitening, and (2) an MMV block-update. Application to simulated 64-ch and real 61-ch EEG data characterized the convergence speed, steady state performance, and computational load of the algorithm. A subsequent paper will examine the impact of non-stationarity, source kurtosis, and forgetting factor on ORICA performance. The described pipeline is integrated in the BCILAB toolbox [9] with utility for future applications in high-density source separation, artifact rejection, and BCI [13].

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2004, vol. 46.

[2] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent eeg sources are dipolar," *PloS one*, vol. 7, no. 2, p. e30135, 2012.

[3] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. Mckeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[4] S. Makeig, M. Westerfield, T.-P. Jung, S. Enghoff, J. Townsend, E. Courchesne, and T. Sejnowski, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, 2002.

[5] Y. Wang and T.-P. Jung, "Improving brain–computer interfaces using independent component analysis," in *Towards Practical Brain-Computer Interfaces*. Springer, 2013, pp. 67–83.

[6] X. Giannakopoulos, J. Karhunen, and E. Oja, "An experimental comparison of neural algorithms for independent component analysis and blind separation," *International Journal of Neural Systems*, vol. 9, no. 02, pp. 99–114, 1999.

[7] X. Zhu, X. Zhang, and J. Ye, "Natural gradient-based recursive least-squares algorithm for adaptive blind source separation," *Science in China Series F: Information Sciences*, vol. 47, no. 1, pp. 55–65, 2004.

[8] M. T. Akhtar, T.-P. Jung, S. Makeig, and G. Cauwenberghs, "Recursive independent component analysis for online blind source separation," in *Circuits and Systems (ISCAS)*. IEEE, 2012, pp. 2813–2816.

[9] C. A. Kothe and S. Makeig, "Bcilab: a platform for brain–computer interface development," *Journal of neural engineering*, vol. 10, no. 5, p. 056014, 2013.

[10] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural computation*, vol. 11, no. 2, pp. 417–441, 1999.

[11] S. Haufe, R. Tomioka, G. Nolte, K.-R. Muller, and M. Kawanabe, "Modeling sparse connectivity between underlying brain sources for eeg/meg," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 8, pp. 1954–1963, Aug 2010.

[12] G. McLoughlin, B. Albrecht, T. Banaschewski, A. Rothenberger, D. Brandeis, P. Asherson, and J. Kuntsi, "Performance monitoring is altered in adult adhd: a familial event-related potential investigation," *Neuropsychologia*, vol. 47, no. 14, pp. 3134–3142, 2009.

[13] T. Mullen, C. Kothe, Y. M. Chi, A. Ojeda, T. Kerth, S. Makeig, G. Cauwenberghs, and T.-P. Jung, "Real-time modeling and 3d visualization of source dynamics and connectivity using wearable eeg." in *IEEE EMBS.*, vol. 2013, 2013, pp. 2184–2187.

[14] A. Delorme, T. Mullen, C. Kothe, Z. A. Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "Eeglab, sift, nft, bcilab, and erica: new tools for advanced eeg processing," *Computational intelligence and neuroscience*, vol. 2011, p. 10, 2011.

[15] J. Onton and S. Makeig, "Information-based modeling of event-related brain dynamics," *PROG BRAIN RES*, vol. 159, pp. 99–120, 2006.

[16] S. C. Douglas, "Blind signal separation and blind deconvolution," *Handbook of neural network signal processing*, vol. 6, pp. 1–28, 2001.

[17] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "Erp components on reaction errors and their functional significance: a tutorial," *Biological psychology*, vol. 51, no. 2, pp. 87–107, 2000.