Evolutionary Based ICA With Reference for EEG μ Rhythm Extraction

SWATHI SRI KAVURI[1], KALYANA C. VELUVOLU[2], (Senior Member, IEEE), AND QUEK HIOK CHAI[1], (Senior Member, IEEE)
[1]School of Computer Engineering, Nanyang Technological University, Singapore 639798
[2]School of Electronics Engineering, Kyungpook National University, Daegu 702-701, South Korea

Corresponding author: Kalyana C. Veluvolu (veluvolu@ee.knu.ac.kr)

ABSTRACT Independent component analysis with reference (ICA-R), a paradigm of constrained ICA (cICA), incorporates textita priori information about the desired sources as reference signals into the contrast function of ICA. Reference signals direct the search toward the separation of desired sources more efficiently and accurately than the ICA. The penalized contrast function of ICA-R is non-smooth everywhere and the ICA-R algorithm does not always reach the global optimum due to the Newton-like learning used. In this paper, we propose a constrained differential evolutionary algorithm with an improved initialization strategy to solve the constrained optimization problem of ICA-R that can asymptotically converge to the optimum. It completely avoids the formulation of a penalized contrast function and scaling (due to the Lagrangian multipliers) by incorporating the ICA contrast function and the violation of the closeness constraint into the selection process of the evolution. Experiments with synthetic data and isolation of μ rhythmic activity from EEG showed improved source extraction performance over ICA-R and its recent enhancements.

INDEX TERMS Constrained ICA (cICA), differential evolution (DE), brain computer interface (BCI), electroencephalography (EEG), independent component analysis (ICA), ICA with reference (ICA-R).

I. INTRODUCTION

Independent component analysis is widely used for separation of independent sources from their mixtures [1]. An ICA algorithm aims to uncover components of input signals by optimizing a contrast function, which maximizes the independence among the separated components [1]. If one observes n time varying signals $x(t) = (x_i(t))_{i=1}^n$ at time t, which are the mixtures of m independent source signals $s(t) = (s_i(t))_{i=1}^m$ that are zero-mean, the noise-free linear model for ICA is given by

$$x(t) = As(t) \qquad (1)$$

where A is an $n \times m$ mixing matrix of full column rank.

The ICA attempts to find an $m \times n$ demixing matrix W to recover the signal sources assuming the observed signals are linear mixtures of independent sources. The signal sources separated using ICA are given by

$$y(t) = Wx(t) \qquad (2)$$

where the vector $y(t) = (y_i(t))_{i=1}^m$ denotes the separated signal sources and each component $y_i(t)$ denotes an estimate of an independent source signal $s_i(t)$ to an indeterminacy of scale and order [2]. It is presumed that (1) and (2) hold for all time t of observations. In this work, we assume $m = n$, which is referred to as the complete ICA, and a noise free model.

Real world source separation problems often wish to extract a desired subset of sources from observed signals. Conventional ICA approaches separate all the components of data so in order to extract a subset of sources, a posteriori selection of the desired components [3] or extraction of components one by one is required [4]. Because of completely blind nature, such methods may result in over-splitting or -clumping, and therefore errors of extracted components compared to semi-blind approaches [5]. The constrained ICA (cICA) framework was introduced for separation of interesting subset of independent components (ICs) by incorporating prior information as constraints to the ICA contrast function [6].

Lu and Rajapakse proposed ICA with reference (ICA-R) [5] in the cICA framework to incorporate reference signals (that is, rough templates of the desired signals) as constraints. The closeness between the reference signals

and desired components act as penalty to the ICA contrast function. The penalized contrast function of the ICA-R has been able to guide the separation of desired sources more accurately than the ICA. Therefore, the ICA-R algorithms have found increasing applications recently, for example in electroencephalography (EEG) [7], [8], electrocardiography (ECG) [9], functional MRI [10], speech aggregation [11], content based image retrieval [12], face reconstruction [13], object tracking [11], H NMR analysis [14] etc. Reference signals have been introduced to the penalized contrast function as temporal [9], spectral [15], spatial [10], and spatiotemporal constraints.

The penalized contrast function of ICA-R is generally non-smooth everywhere and convergence to the optimum is largely affected by the reference signals, Newton-like learning algorithm, the closeness function, and the upper-bound (a threshold) of the closeness [2], [5]. Several attempts have been made to improve the convergence and stability of ICA-R. Huang *et al.* have investigated different closeness measures to improve the convergence of the ICA-R [16]. Lin *et al.* [17] proposed the weight normalization scheme to avoid the equality constraint. Improved cost function based fast One-Unit Independent Component Analysis that is suitable to extract the FECG was proposed in [18]. Zhang [15] proposed a second order statistics based method for designing suitable reference signals for reliably extracting weak periodic or quasi-periodic desired source signals. A reference based initialization of unmixing matrix that can guide the learning towards the global optimum was proposed, to improve the convergence of ICA-R [19]. Li *et al.* [20] removed the upper bound of the distance function of ICA-R by transforming the constrained optimization by alternatively optimizing the negentropy and the closeness function. In [21], an analytical expression of the expected signal of interest was incorporated as a priori information to improve the signal separation quality. However, these attempts have not completely resolved inherent problems of ICA-R in converging to the optimal solution and are therefore limited to specific applications.

In this paper, we propose to use a population-based (or a genetic) algorithm - constrained differential evolution (cDE) [22] - for the optimization of ICA-R contrast function - constrained DE for ICA-R (cDE-ICAR).

Differential Evolution (DE) is one of the most powerful population-based stochastic algorithm, which keeps a population of solutions that are successively updated by addition, subtraction, and component swapping of the members. The population evolves asymptotically towards the optimum more likely than gradient based techniques [23] and requires less number of tuning parameters [24]. With the popularity of DE in applications, more and more researchers are paying attention to the theoretical studies on global convergence of DE [25], [26].

The novelty of this work lies in the application of differential evolution algorithm to the optimization problem in ICA. Moreover, an improved initialization strategy was also proposed to speed up the convergence of the DE. The proposed cDE-ICAR algorithm explicitly computes and incorporates the violation of closeness constraint (if it is not zero or near zero) to the selection process of evolution. This reduces the penalty on violation of the closeness constraint and solutions that are closer to the reference are transferred to the next generations till the stopping condition is met.

A preliminary version of this work has been presented as a conference paper [27]. In this work, we propose an improved initialization strategy for cDE-ICAR and introduce a specific termination criteria using similarity measure. The efficacy and validity of the proposed constrained DE for ICA-R (cDE-ICAR), are quantitatively evaluated by separating desired sources from synthetic data. And its utility is demonstrated by separating μ rhythms from real EEG datasets. The results are compared with the ICA-R [5], improved cICA [16], and improved ICA-R [20] algorithms. The manuscript is organized as follows. Section II summarizes one-unit ICA-R and Newton-like learning rules and discusses its convergence and stability. Section III describes the cDE-ICAR algorithm. Section IV presents experiments with both synthetic and real EEG data for extracting the desired brain rhythms. Discussion and conclusions are provided in Section V.

II. ICA WITH REFERENCE (ICA-R)

The ICA-R separates a desired subset of independent sources from input signals when reference signals of the sources are available [2], [5]. In this section, we first present the one-unit ICA-R algorithm that extracts a single component closest to the reference signal. The one-unit ICA-R algorithm is then used to separate multiple sources by using a deflationary scheme separating one source by one at a time [2]. In the sequel, the time index t is omitted in the sequel in order to simplify the notations.

A. ONE-UNIT ICA-R

The one-unit ICA-R separates only a single source at a time and therefore determines a single weight vector of the demixing matrix. If the separated source is y and the corresponding weight vector is w, then $y = wx$. The estimate y correspond to one of the sources in $(s_i)_{i=1}^{m}$.

The negentropy is used as the contrast function of the ICA as non-Gaussianity implies the independence [1]. The negentropy of signal y is defined as

$$J_G(y) = H(y_G) - H(y) \qquad (3)$$

where y_G is a Gaussian random variable having the same variance as signal y and H denotes the entropy. Maximizing negentropy finds the independent sources when the sources are uncorrelated [28]. Hyvarinen introduced a flexible and reliable approximation of negentropy [28]:

$$J_G(y) = \rho[E\{f(y)\} - E\{f(\nu)\}]^2 \qquad (4)$$

where ρ is an irrelevant positive constant, ν is a Gaussian variable having zero mean and unit variance, and $E\{\cdot\}$ denotes

the expectation. The function f represents the distribution of the sources and a non-quadratic function is used as the sources are mostly super-Gaussian in practice.

The contrast function in (4) is augmented in the ICA-R by introducing a closeness measure $\varepsilon(y, r)$ between the extracted source y and the reference r as an *a priori* constraint. Several distance measures including the mean square error and the correlation have been considered for closeness measure [16].

The penalized contrast function of one-unit ICA-R is formulated from the following constrained minimization problem:

$$\text{minimize} \quad -J_G(y) \tag{5}$$

$$\text{subject to} \quad g(y) = \varepsilon(y, r) - \xi \le 0 \tag{6}$$

$$\text{subject to} \quad h(y) = E\{y^2\} - 1 = 0 \tag{7}$$

where ξ denotes the upper-bound of the closeness to ensure separation of the desired source. The equality constraint ensures that the weight vector w is bounded [17].

Lu and Rajapakse introduced slack variable z to transform the inequality constraint to an equality constraint: $g(y) + z^2 = 0$. By explicitly manipulating the optimal value z^*, the optimization of ICA-R in (5) becomes the penalized contrast function [2], [5]:

$$J(y) = J_G(y) + \lambda_1 h(y) + \frac{1}{2}\gamma |h(y)|^2$$
$$+ \frac{1}{2\gamma}\left((\max\{\lambda_2 + \gamma g(y, r), 0\})^2 - \lambda_2^2\right) \tag{8}$$

where λ_1 and λ_2 are Lagrange multipliers for the constraints of $g(y, r)$ and $h(y)$, respectively, γ is the scalar penalty parameter, and the term $\frac{1}{2}\gamma |h(y)|^2$ ensures local quadratic approximation.

The ICA-R algorithm uses a Newton-like learning algorithm to minimize the penalized contrast function in (8). The change of weight Δw in each iteration is given by

$$\Delta w = -\eta(\nabla_w^2 J)^{-1}\nabla_w J \tag{9}$$

where $\nabla_w J$ and $\nabla_w^2 J$ are the gradient vector and Hessian matrix of the penalized contrast function J in (8) with respect to the weight vector w. The learning rate η is set using the backtracking line search [29], which starts with one and gradually decreases to zero. In this paper, we use correlation $\varepsilon(y, r) = 1/(E\{yr\})^2$ as the closeness measure as it is a widely used in ICA to determine the desired component and converges faster than the other measures [16]. In order to simplify the matrix inversion, the Hessian matrix is approximated as

$$\nabla^2 J \approx \delta(w)R_{xx} \tag{10}$$

where $\delta(w) = \overline{\rho}E\{f''(y)\} + 2\lambda_1 + 6\lambda_2 \frac{E\{r\}^2}{E\{yr\}^4}$, $R_{xx} = E\{xx^T\}$ is a non-singular covariance matrix and $\overline{\rho} = 2\rho(E\{f(y)\} - E\{f(v)\})$ [2]. The weight change in (9) now becomes

$$\Delta w = -\frac{\eta}{\delta(w)}R_{xx}^{-1}\nabla_w J \tag{11}$$

where $\nabla_w J = \overline{\rho}E\{xf'(y)\} + 2\lambda_1 E\{xy\} - 2\lambda_2 \frac{E\{r\}}{E\{yr\}^3}$ complete this equation.

The Lagrange multipliers λ_1 and λ_2 are updated using following gradient-ascent rules:

$$\Delta\lambda_1 = \eta_1 h(y)$$
$$\Delta\lambda_2 = \max\{-\lambda_2, \eta_2 g(y)\} \tag{12}$$

where η_1 and η_2 are the learning rates for the respective Lagrange multipliers. The weight vector w is initialized with the pseudo-inverse of the input vector and λ_1 and λ_2 are to zeros. In each iteration, w, λ_1, and λ_2 then updated according to (11) and (12) until the convergence is reached.

B. FACTORS AFFECTING THE CONVERGENCE OF ICA-R
The penalized contrast function $J(y)$ is generally non-smooth and may contain several local minima depending on the number of the desired sources. The convergence of the ICA-R depends on several factors: learning algorithm, closeness measure, reference signals, etc.

The ICA-R algorithm uses a Newton-like learning algorithm [5]. The Newton-like learning converges to the weight vector w defined by Kuhn-Tucker (KT) triple $(w, \lambda_1, \lambda_2)$ that satisfies the first-order conditions: $\nabla_w J = 0$, $h(w) = 0$, $g(w) \le 0$, $\lambda_1 > 0$, $\lambda_2 \ge 0$ and $\lambda_2 g(w) = 0$ and second-order condition of positive-definiteness of the Hessian matrix. In order to ensure that the ICA-R algorithm achieves its optimum, the upperbound ξ of the distance measure is needed such that

$$\varepsilon(w_i x, r) - \xi = \begin{cases} \le & 0, \quad \text{if } w_i = w \\ > & 0, \quad \text{if } w_i \ne w, \forall i = 1, 2, \dots m \end{cases} \tag{13}$$

where the weight vector corresponding to the desired signal $w = \arg\min_{i=1,\dots m}\{\varepsilon(w_i x, r)\}$.

Convergence and stability of the cICA and ICA-R have received the attention of many researchers [19], [20], [30]. One way to avoid the problem of getting trapped in to a local optima is to have a good initialization. An improved cICA (I-cICA) method that uses a demixing matrix initialized by the Moore-Penrose generalized inverse of input based on the prior information has been proposed [20].

The ICA-R can recover the correct source only with appropriate selection of the upper bound ξ of the contrast function. In order to avoid the upper bound, an Improved ICA-R method (I-ICAR) is proposed by formulating the constrained optimization in ICA-R as a dual-optimization problem. The optimization is carried sub-optimally by using a dual gradient-decent strategy that optimizes the contrast function (4) and the closeness $\varepsilon(y, r)$, followed by weight normalization at each iteration [20]. The above attempts increase the likelihood but cannot guarantee the convergence of the solution to the desired signal.

III. CONSTRAINED DE FOR ICA-R
We propose a differential evolutionary (DE) algorithm to solve the constrained optimization problem (5) of ICA-R. Evolutionary algorithms are a class of heuristic optimization techniques that search from a "population" of solutions

by iteratively applying genetic operators such as crossover and/or mutation to generate new candidates that are biased towards the optimal solution. The DE algorithm is evolved in a manner that good candidates appear more frequently in the populations, asymptotically converging to a population containing the optimal solution. The DE is one such quasi-random population-based method that uses efficient stochastic heuristic strategies to create new candidates [31]. Compared to other heuristic search methods, it is simple, efficient, fast and more likely to find the optimal solution when real variables are involved [23]. Therefore, it is more likely to find the global minima in the penalized contrast function of the ICA-R than Newton-like learning.

The DE uses a population of K solutions $(w^k)_{k=1}^K$ where $w^k = (w_j^k)_{j=1}^n$ represents the k-th solution or weight and w_j^k denotes the j-th element. In each generation of evolution, crossover and mutation operations are performed in order to find new solutions from the existing population to asymptotically reach a population containing the optimal solution w [23], [31]. Below, we demonstrate the various steps involved in the constrained DE algorithm for the ICA-R (cDE-ICAR).

A. INITIALIZATION

The efficacy of the search using DE depends on the initial population [32]. The initial unmixing vector w_0 is obtained by assuming that the reference signal r is equal to the estimated signal, such that $w_0 x = r$. The vector w_0 is then approximated as $w_0 = r x^\dagger$, where $x^\dagger$ is the Moore-Penrose generalized inverse of x: that is, $x^\dagger = x^{\mathrm{T}}(xx^{\mathrm{T}})^{-1}$. A good initial population was obtained by introducing a small random Gaussian perturbation to the unmixing vector w_0. This provides a good seed population closer to the desired optimal solution and with sufficient diversity.

B. MUTATION

Mutation is performed to produce a *mutant* vector stochastically by linear combination of differences of randomly chosen members [23]. The following strategy is used to generate the mutant vector $v^k = (v_j^k)_{j=1}^n$ is a vector of size w and $k \in [1, K]$ from the *parent* vectors $(w^k)_{k=1}^K$ in the population:

$$v^k = w^{\alpha_1} + \kappa(w^{\alpha_2} - w^{\alpha_3}) + \kappa(w^{\alpha_4} - w^{\alpha_5})$$

where $\alpha_1, \alpha_2, \alpha_3, \alpha_4,$ and α_5 are mutually independent random integers generated in the range $[1, K]$ and the real scaling factor $\kappa (\geq 0)$ which scales the differences of the candidate vectors is a uniform random number in $[0, 1]$.

C. CROSSOVER

To improve the diversity of the population, one or more elements of the mutant v^k are crossed uniformly with the parent w^k vector to yield a *trial* vector $u^k = (u_j^k)_{j=1}^n$:

$$u_j^k = \begin{cases} v_j^k & \text{if } q \leq q_c \text{ OR } q' = j \\ w_j^k & \text{otherwise} \end{cases} \qquad (14)$$

where $q \in [0, 1]$ is the j-th evolution of a uniform random number and $q' \in \{0, n\}$ is a randomly chosen index for the member in the population. The crossover rate $q_c \in [0, 1]$ is a user-specified parameter that controls the fraction of the mutant that are copied to the offspring.

D. SELECTION

The selection between the trial vector or the parent vector is made based on their fitness and feasibility. The feasibility of a member in the population is measured by its constraint violations. The constraint violation $V(w^k)$ of a member w^k is measured by

$$V(w^k) = \frac{1}{2}(G(w^k) + H(w^k)) \qquad (15)$$

where

$$G(w) = \begin{cases} 0, & \text{if } \varepsilon(y, r) - \xi \leq 0; \\ \varepsilon(y, r) - \xi, & \text{otherwise.} \end{cases} \qquad (16)$$

$$H(w) = \begin{cases} 0, & \text{if } |h(y)| \leq 0 \\ |h(y)|, & \text{otherwise.} \end{cases} \qquad (17)$$

A solution is said to be a feasible solution if no constraint is violated or the constraint violation is zero. It takes a positive value when at least one constraint is violated.

The trial vector is selected if it has a lower constraint violation than the parent, or it has the same constraint violation but has a better fitness value. That is, the offspring is selected such that

$$w^k = \begin{cases} u^k, & \text{if } (V(u^k) < V(w^k)) \text{ or} \\ & (V(u^k) = V(w^k) \text{ and } J(u^k) \leq J(w^k)); \\ w^k, & \text{otherwise.} \end{cases} \qquad (18)$$

The constraint violation in (18) guides the search towards the feasible areas of the search space and helps to find the optimal weight asymptotically when the fitness reaches a plateau [33]. The optimal solution is assured only when the feasible solution is achieved with a best fitness.

E. TERMINATION

As cDE-ICAR algorithm usually converges when there are no changes in the optimal solution vector and when the members of the population becomes closer to one another [33]. In order to detect the convergence of the algorithm, a similarity among the members of the population $\{w^k\}_{k=1}^K$ is defined as follows:

$$\text{similarity} = \sum_{k=1}^K \sum_{k'>k}^K \left\| w^k - w^{k'} \right\| \qquad (19)$$

Over the generations of cDE-ICAR, the similarity measure decreases and asymptotically reaches a zero (or to a minimum). However, in some instances, the algorithm may take a large number of generations to converge or exhibit fluctuations of similarity at the convergence. Therefore, in addition to the similarity measure, a maximum number of generations $N_{\max}$ is used as the stopping criterion.

The cDE-ICAR evolves with genetic operators of mutation, crossover and selection to all the members and converges to the optimal solution when a stopping criterion is met. The cDE-ICAR is given in Algorithm (1).

Algorithm 1 cDE-ICAR Algorithm

Generate an initial population $\{w^k\}_{k=1}^K$ of size K

Set κ and q_c

repeat

   for each individual k in the population **do**

      Generate integers, $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5 \in [1, K]$ randomly s.t. $\alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \alpha_4 \neq \alpha_5 \neq k$

$$v^k = w^{\alpha_1} + \kappa(w^{\alpha_2} - w^{\alpha_3}) + \kappa(w^{\alpha_4} - w^{\alpha_5})$$

      for every element j **do**

         Generate a uniform random number $q \in [0, 1]$ and random integer $q' \in \{0, n\}$

$$u_j^k = \begin{cases} v_j^k & \text{if } q \leq q_c \text{ OR } q' = j \\ w_j^k & \text{otherwise} \end{cases}$$

      end for

$$w^k = \begin{cases} u^k & \text{if } (V(u^k) < V(w^k)) \text{ or} \\ & (V(u^k) = V(w^k) \text{ and } J(u^k) \leq J(w^k)), \\ w^k & \text{otherwise} \end{cases}$$

   end for

until a stopping criterion is met

F. PARAMETER SETTING

There are two key parameters in the cDE-ICAR algorithm: scaling factor κ and the crossover rate q_c. The scaling factor should be large enough to escape the local optima and must be above a certain critical value to avoid premature convergence. A large crossover rate q_c speeds up convergence but from a certain value upwards, the convergence rate may decrease or the population may converge prematurely. When $q_c = 1.0$, the number of trial solutions will be reduced dramatically which may lead to stagnation [31]. The other two parameters in the evolutionary algorithm are the population size K and the maximum number of generations $N_{\max}$.

Although there are no definite rules or criteria to determine the above parameters, possible ranges that works for many applications have been determined [34]: $K = 5 \cdot n$, $\kappa \in [0.4, 0.95]$ and $q_c \in [0.3, 0.9]$. In our experiments, the algorithms were run for different values of κ and q_c within the recommended ranges and their values were empirically fixed. The maximum number of generations $N_{\max}$ were determined by looking at the performance of the algorithm.

IV. EXPERIMENTS AND RESULTS

Experiments on simulated signals and real EEG data were performed in order to demonstrate the performance of cDE-ICAR method in extracting the desired sources by using correlation as closeness measure. The results were compared with those obtained with ICA-R [5], I-cICA [20] and I-ICAR [19] algorithms and ICA with post selection. In this paper, correlation $\varepsilon(y, r) = 1/(E\{yr\})^2$ was used as the closeness measure.

A. SIMULATED DATA

Five signals were simulated: two deterministic periodic signals s_1 and s_2; a non-periodic super-Gaussian source signal s_3, a sub-Gaussian signal s_4, and a noise signal s_5 whose power spectrum matches the power spectrum of human EEG.

$$s_1 = \sin(2\pi 12t) + \sin(2\pi 10t)$$
$$s_2 = \cos(2\pi 20t) + \cos(2\pi 15t) + \cos(2\pi 16t)$$
$$s_3 = \cos(2\pi 25t) + 2\cos(2\pi 26t) + 0.03(\exp(1.5t)$$
$$\quad - \exp(2t))$$
$$s_4 = \cos(2\pi 0.9t)\sin(2\pi 2t)$$

All the source signals were normalized to zero mean and unit variance. Each signal had 1000 time samples with a sampling interval of 4 ms. The excessive kurtosis of the super-Gaussian and sub-Gaussian signals were 5.38 and -0.76, respectively.

The signals were linearly mixed using a weight matrix in which the elements were randomly drawn from a uniform distribution on the unit interval to obtain input signals. For example, synthetic sources and mixed signals obtained with the following mixing matrix A:

$$A = \begin{bmatrix} 0.98 & 0.59 & 0.12 & 0.09 & 0.73 \\ 0.44 & 0.26 & 0.30 & 0.26 & 0.49 \\ 0.11 & 0.60 & 0.32 & 0.80 & 0.58 \\ 0.26 & 0.71 & 0.42 & 0.03 & 0.24 \\ 0.41 & 0.22 & 0.51 & 0.93 & 0.46 \end{bmatrix}$$

The signals recovered by cDE-ICAR algorithm from these synthetic mixtures are shown in Fig. 1. The cDE-ICAR algorithm was used to extract source signals s_1, s_2, s_3 and s_4 from the five input signals to demonstrate cDE-ICAR algorithm. The reference signal for extracting each of the desired sources was obtained by band-pass filtering of the observed signals that has the highest power in the frequency band of the desired source [7]. The parameters $K = 30$, $\kappa = 0.7$, $q_c = 0.7$, and $N_{\max} = 3000$ were set as explained earlier. The signals extracted by ICA, followed by post-selection are also given.

1) PERFORMANCE EVALUATION

The performance index (PI) was used to evaluate the goodness of separation of desired sources from input mixtures [38]. For a source extracted using one-unit ICA-R,

$$\text{PI} = \frac{\sum_{j=1}^n |p_j|}{\max_{j'} |p_{j'}|} - 1 \tag{20}$$

where p_j is the j-th element of $p = w^{\text{T}}A$. PI is zero when the desired IC is perfectly separated. Distortion of the recovered signal is measured using the signal-to-noise ratio (SNR). To compensate for scale ambiguity of ICA, a scalar factor
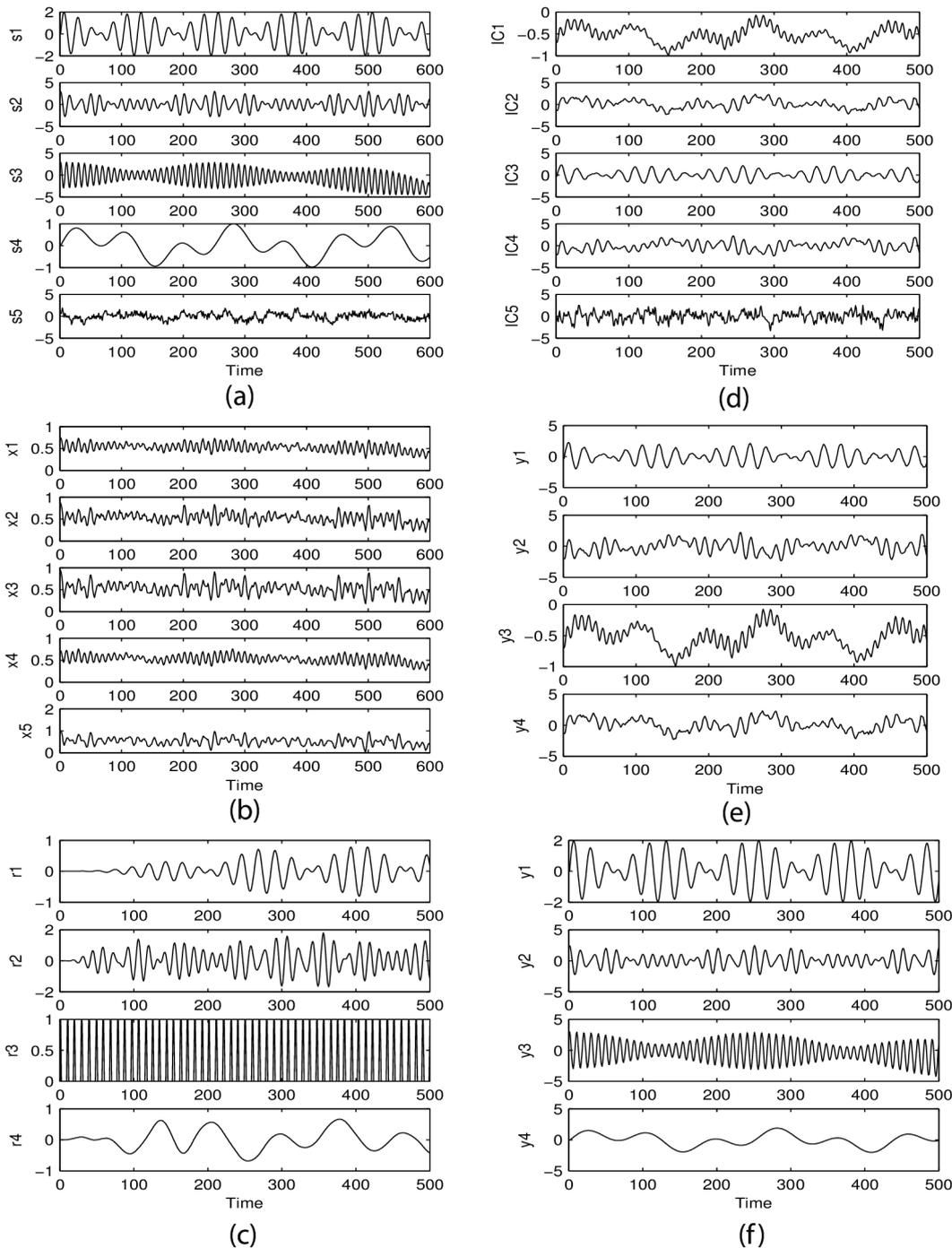
FIGURE 1. Illustration of experiments with simulated data: (a) source signals, (b) input signals after linear mixing of sources with mixing matrix *A*, (c) reference signals used for extraction of s_1, s_2, s_3 and s_4, (d) signals extracted by using ICA, (e) signals extracted by ICA and post-selection, and (f) signals recovered using the cDE-ICAR algorithm. One time unit = 4 *ms* in the time axis.

$\beta = \arg\min_{\beta'} E\{(\beta'y - s)^2\}$ is obtained such that $\beta = E\{y^2\}^{-1}E\{ys\}$; and the SNR is then given by

$$\text{SNR(dB)} = 10 \log_{10}\left(\frac{\sigma^2}{E\{(\beta y - s)^2\}}\right) \quad (21)$$

where $E\{(\beta y - s)^2\}$ is the mean square error between the source and the recovered signal, and $\sigma^2 = E\{s^2\}$ is the variance of the desired source. A higher SNR means that the separated signals are closer to original source signals.

The extraction of signal sources s_1, s_2, s_3 and s_4 was performed 30 times with input signals simulated by mixing with different random mixing matrices *A*. The mean and variance of the performance measures SNR and PI that were obtained from ICA, ICA-R, I-cICA, I-ICAR, and cDE-ICAR

TABLE 1. SNR and PI values of extracting synthetic sources.

		ICA	ICA-R	I-cICA	I-ICAR	cDE-ICAR
s_1	SNR	14.92 ± 2.43	30.97 ± 2.78	22.66 ± 2.14	29.56 ± 1.53	37.36 ± 1.82
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	
	PI	0.61 ± 0.24	0.22 ± 0.13	0.25 ± 0.05	0.23 ± 0.03	0.06 ± 0.01
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	
s_2	SNR	4.24 ± 3.25	22.75 ± 2.75	12.12 ± 2.39	30.43 ± 1.32	43.34 ± 1.56
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	
	PI	0.37 ± 0.01	0.42 ± 0.02	0.34 ± 0.02	0.52 ± 0.04	0.03 ± 0.02
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	
s_3	SNR	4.75 ± 4.23	3.41 ± 1.85	3.96 ± 2.25	4.73 ± 1.32	20.02 ± 1.66
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	
	PI	0.38 ± 0.01	0.68 ± 0.05	0.60 ± 0.04	0.56 ± 0.04	0.03 ± 0.02
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	
s_4	SNR	3.56 ± 2.49	15.38 ± 2.52	14.16 ± 2.25	20.16 ± 1.32	20.98 ± 1.58
	p-value	0.11	< 0.001	< 0.001	0.45	
	PI	0.55 ± 0.01	0.54 ± 0.03	0.57 ± 0.04	0.03 ± 0.04	0.04 ± 0.02
	p-value	< 0.001	< 0.001	< 0.001	0.62	

algorithms are provided in Table 1. As seen, cDE-ICAR algorithm showed better performance in extracting source signals overall. In order to see whether the performances are indeed different, significant tests of the performance measures of cDE-ICAR against those of other methods were performed. The PI and SNR values were significantly better ($p < 0.001$) in extracting s_1, s_2 and s_3 compared to all other methods. For signal s_4, except the SNR ($p = 0.45$) and PI ($p = 0.62$) values over I-ICAR, improvement over all the other techniques was statistically significant ($p < 0.001$).

2) CONVERGENCE AND STABILITY
The plots in figure 2 illustrate the convergence of cDE-ICAR in extracting source signals: s_1, s_2, s_3 and s_4 in terms of distance measure of populations containing the solutions and constraint violations against generations, and fitness values against function evaluations. As the algorithms converge, distance measure and constraint violations reach zero while fitness values reach its optima. That is, the solutions are feasible and have the best fitnesses, indicating that the convergence to an optimal solution or extraction of desired sources. Table 2 shows the constraint violations in extracting the two sources with ICA-R, I-cICA, I-ICAR, and cDE-ICAR algorithms. Unlike cDE-ICAR, the other algorithms do not explicitly compute and incorporate constraint violations into the optimization. The cDE-ICAR algorithm achieves zero constraint violations, ensuring feasibility at convergence. The other ICA-R methods do not (satisfy the constraints) necessarily ensure the feasibility of the solution upon convergence, which makes it difficult determine if the extraction of the desired source is at global optima.

B. EEG DATA
To demonstrate the validity of the proposed method, experiments with real EEG data were performed and results were compared with other methods.

1) EXPERIMENTAL SETUP
EEG data is taken from BCI Competition 2003, which was provided by the Wadsworth Center, NYC Department of Health [35]. Data was collected from three subjects A, B and C. The subject sat in a reclining chair facing a video screen and was asked to remain motionless during the experiment. Scalp electrodes recorded 64 channels of EEG [36], each with reference to an electrode on the right ear (amplification 20,000; bandpass of 0.1-60 Hz). All 64 channels were digitized at 160 Hz and stored. Data was collected from each subject for 10 sessions of 30 min each. Each session consisted of 6 runs, separated by 1 min break, and each run consisted of approximately 32 individual trials. Each trial began with a 1-s period during which the screen was blank. Then, the target appeared at one of four possible positions on the right edge. One second later, a cursor appeared at the middle of the left edge of the screen which used subject's μ or β rhythm (i.e., frequencies between 8-12 Hz or 18-24 Hz, respectively) to control the vertical cursor movement towards the target that appears at one of the four possible positions on right edge of the video screen. The target on the upper conner of the right edge is labeled as "top", whereas the lower conner target is labeled as "bottom". When the cursor reached the right edge, the screen went blank. This event signaled the end of the trial.

The cDE-ICAR and other algorithms were used to extract μ rhythms from the trials corresponding to target position: "top" and "bottom" of subjects A, B and C. Electrodes from the motor cortex part of the brain - FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6 - were chosen for source analysis as the interest was to see the sensorimotor activity of the brain. The parameters of cDE-ICAR were $K = 40$, $\kappa = 0.7$, $q_c = 0.7$ and $N_{\max} = 3500$.

2) REFERENCE SIGNALS
The reference signal for the experiments was obtained using wavelet packet decomposition (WPD) of observed EEG data as described in [7]. EEG signals from channels C3 and C4, closer to the sensorimotor cortex from where μ rhythms originate were decomposed using Daubechies mother wavelets. The decomposition was carried out for each signal at a certain scale ζ, which splits the signal into 2^ζ subbands resulting in a
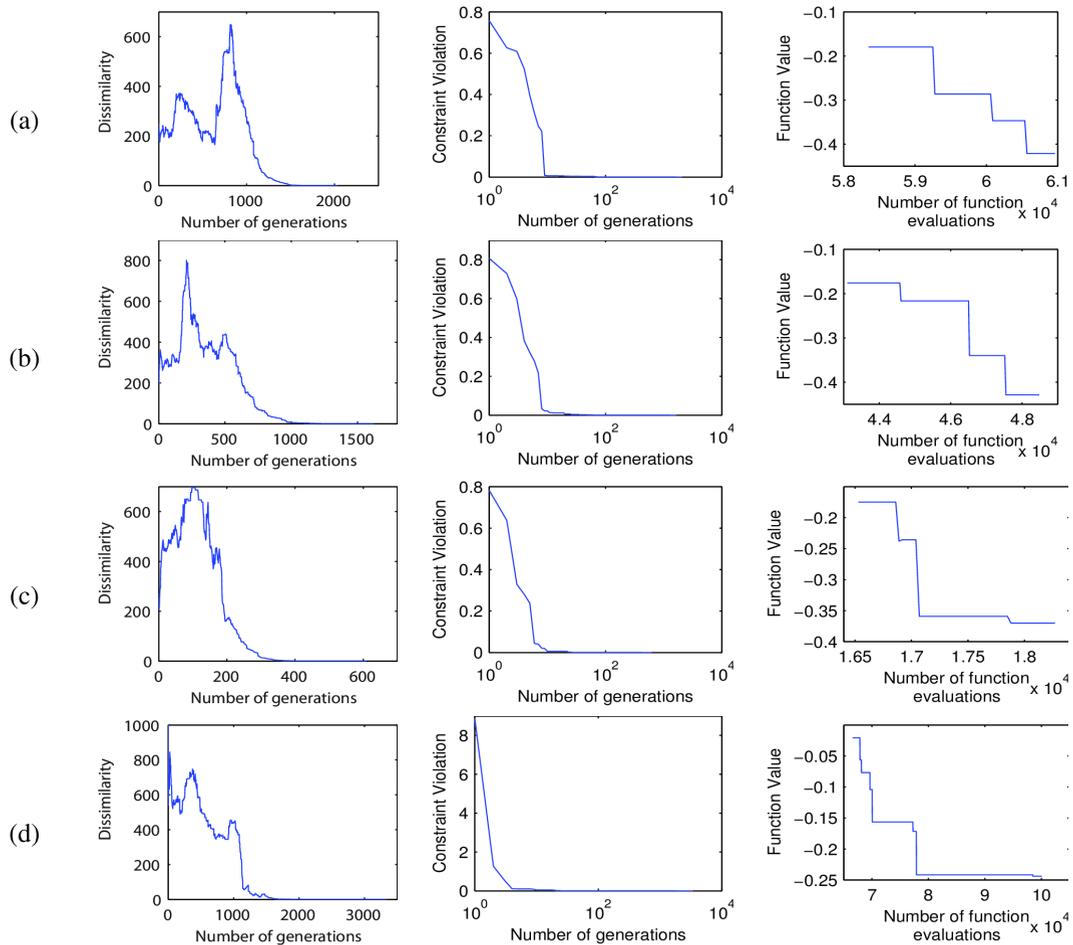
FIGURE 2. Illustration of convergence of cDE-ICAR algorithm extracting sources: (a) s_1, (b) s_2, (c) s_3, and (d) s_4, in terms of distance between the members of the population, constraint violations, and best fitness.

TABLE 2. Constraint violations in extracting simulated source signals at various generations.

Sources		2	10	20	30	100	500	3000
s_1	ICA-R	12.86	2.86	2.86	2.86	2.86	2.86	2.86
	I-cICA	8.17	3.40	3.39	3.39	3.39	3.39	3.39
	I-ICAR	1.80	1.08	1.08	1.08	1.08	1.08	1.08
	cDE-ICAR	0.99	0.11	0	0	0	0	0
s_2	ICA-R	24.22	1.56	1.56	1.56	1.56	1.56	1.56
	I-cICA	13.88	7.35	7.35	7.35	7.35	7.35	7.35
	I-ICAR	1.08	1.02	1.02	1.02	1.02	1.02	1.02
	cDE-ICAR	1.77	0.20	0.07	0.02	0	0	0
s_3	ICA-R	3.30	0.65	0.65	0.65	0.65	0.65	0.65
	I-cICA	5.29	0.42	0.42	0.42	0.42	0.42	0.42
	I-ICAR	0.73	0.37	0.37	0.37	0.37	0.37	0.37
	cDE-ICAR	0.09	0.02	0.01	0	0	0	0
s_4	ICA-R	42.86	9.69	9.66	9.66	9.66	9.66	9.66
	I-cICA	1.36	0.56	0.52	0.52	0.52	0.52	0.52
	I-ICAR	3.93	3.93	3.93	3.93	3.93	3.93	3.93
	cDE-ICAR	0.89	0.04	0.03	0	0	0	0

balanced binary tree structure. Scalp electrodes recorded over 64 channels of EEG were each band-pass filtered in the range 0.1-60 Hz. Therefore choosing $\zeta = 4$, the signal is split it in to 16 sub-bands with a frequency resolution of approximately $3 \sim 4$ Hz. The wavelet packet coefficients of the subband whose frequency corresponds to the frequency of the

desired μ rhythms (8-12 Hz) were used to reconstruct the reference signal.

3) PERFORMANCE EVALUATION

To evaluate the performance of extracting the sensorimotor activity (μ rhythm), a reactivity index R_μ is defined as the
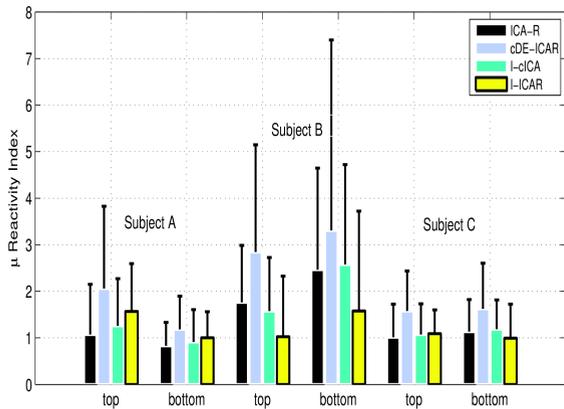
FIGURE 3. The mean μ reactivity index of the extracted sources across different subjects performing "top" and "bottom" targets, using ICA-R and cDE-ICAR methods.

relative normalized variation of power spectral density between the stimulus ON and OFF states relative to the power in the OFF state:

$$R_\mu = \frac{\sum_\mu P_{\text{ON}} - \sum_\mu P_{\text{OFF}}}{\sum_\mu P_{\text{OFF}}} \qquad (22)$$

where ON state corresponds from 3 s to 4.5 s of each trial after stimulus onset where the subject uses μ rhythm (8-13 Hz) to control the cursor and OFF corresponds from 0 s to 3 s of each trial where the subject shows no μ-band activity.

Figure 3 shows bars representing mean$+1.96\times$standard error of μ reactivity indices over all the trials across different subjects for each method. The μ reactivity indices R_μ of sources across subjects clearly depict that the component extracted using cDE-ICAR shows a significant μ activity during the stimulus period in both "top" and "bottom" targets, compared to those extracted by other methods. The performance of I-cICA and I-ICAR methods over ICA-R was not consistently better for all the subjects.

In addition to ICA based methods, we compare the performance of extracting μ rhythms by using Common Spatial Patterns (CSP) and Linear Discriminant Analysis (LDA) which has been used in many applications of motor imagery based BCI [35] for classification of sources/features extracted using different methods as desired or undesired. The training data for classification is generated using all the trials that correspond to target positions: "top" and "bottom" of subjects A, B and C. For both the targets, the trails that hit the target are labelled as desired and the trails that missed the target are labelled as undesired.

CSP based spatial filtering essentially projects multi-channel EEG data in to low-dimensional spatial subspace with a projection matrix, each row of which corresponds to the weights of each channel called the spatial pattern. The algorithm is based on the simultaneous diagonalization of covariance matrices of EEG data under two conditions [37]. CSP filters maximize the variance of the spatially filtered signal under one condition while minimizing it for the other condition.

TABLE 3. Average classification accuracies (mean ± std %) for different methods.

Subject	A	B	C
CSP	61.75 ± 2.2	68.16 ± 3.1	64.14 ± 2.8
ICA	50.18 ± 3.5	59.47 ± 3.1	55.28 ± 2.5
ICA-R	65.25 ± 3.5	69.86 ± 4.2	64.34 ± 3.8
I-cICA	70.98 ± 2.5	82.47 ± 3.3	68.78 ± 2.4
I-ICAR	76.60 ± 4.7	70.82 ± 5.5	71.847 ± 3.2
cDE-ICAR	85.41 ± 3.1	92.53 ± 3.5	87.91 ± 3.7

After having band-pass filtered EEG signals to the desired μ rhythms of interest, spatial patterns are extracted to discriminate the strong rhythm (desired) and a weak rhythm (undesired). The feature vector is computed as the normalized log power (variance) of these signals. The log-transformation serves to approximate normal distribution of the data.

For the component based methods of ICA and ICA-R methods, the sum of power spectral density in the μ-band of the component extracted during the stimulus ON condition is used as a feature for classification of sources. LDA used training data to learn the parameters of classification and feature information to classify each trial in a run of testing data as desired or not. Table. 3 lists average classification accuracies that are achieved for a given subject.

The estimated extracted time courses of activation correspond to one of the ICA components. The inverse of estimated weight vector gives relative projection strengths of the extracted components at each of scalp sensors. The strengths of scalp projections provide evidence for physiological origins of the component. The scalp maps are obtained by back-projecting μ rhythms extracted by ICA-R and cDE-ICAR techniques on the subjects, as shown in Fig. 4 for different experiments. For the entire stimulus ON period (3 s to 4.5 s), the scalp maps show similar topological pattern. For sake of illustration, scalp maps at time instant t = 3.19 s are shown. Most scalp maps show neuronal sources from the sensory motor cortex part of the brain. The activation patterns obtained by cDE-ICAR method were more focal and highlighted in the expected regions of the brain compared to those obtained with ICA-R.

V. DISCUSSION
The DE algorithm is simple, efficient, and deliver better results in optimization problems involving real variables. It creates diverse search directions by mutating the differences of randomly selected parent vectors; this promotes the exploration of search space and thereby converges more efficiently and accurately towards the global optima. Compared to other population-based algorithms, DE algorithm requires tuning of less parameters whose ranges are known. The selection of the upper-bound of the distance function *a priori* or learning rates of Newton-like algorithm is not needed in the cDE-ICAR algorithm.

To demonstrate the utility in cDE-ICAR, μ-rhythms from EEG data collected in a BCI application were extracted.
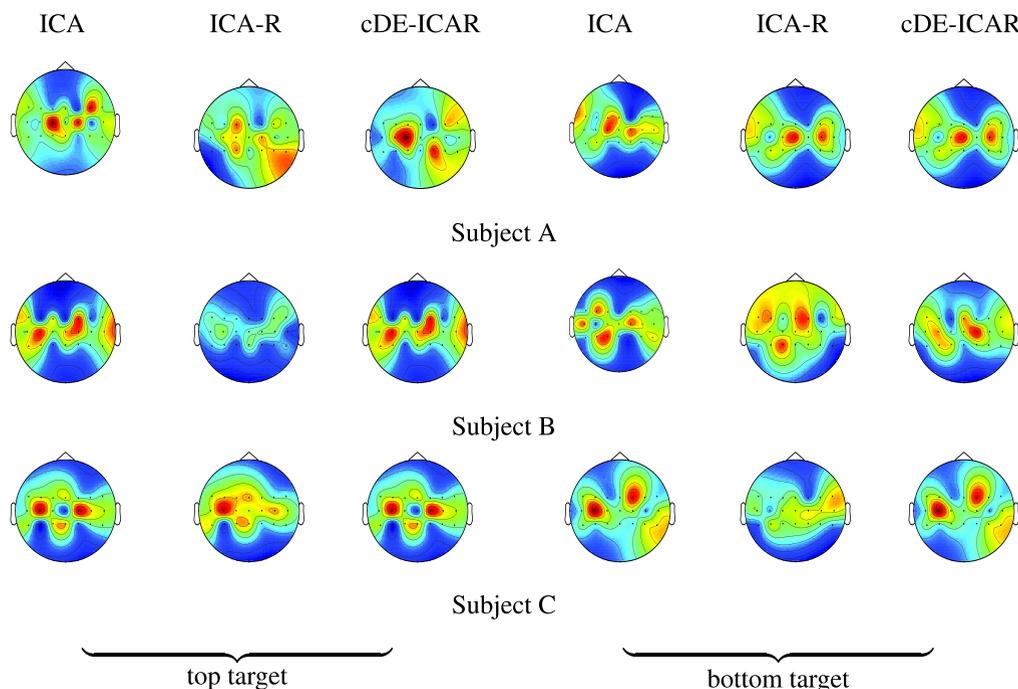
FIGURE 4. The scalp maps obtained by back-projectingthe extracted μ rhythms by ICA with post selection, ICA-R and cDE-ICAR methods on different subjects. The maps are obtained by plotting components at $t = 3.19$ s for different subjects with both top and bottom targets .

Experiments showed that cDE-ICAR was able to recover more rhythmic activity in μ-band and locate the source signals better; the scalp maps show more focal and higher source signals extracted by the cDE-ICAR than those obtained from other methods. The color difference in the scalp projection obtained from ICA-R of subject B for top target can be contributed to the smaller amplitude, if the difference was caused by the on/off control strategy, as both ICA and cDE-ICAR employ the same data, the ICA and cDE-ICAR will show similar pattern.

Because of the nature of the evolutionary algorithms, the computational complexity of cDE-ICAR was higher than the earlier approaches. Though our experiments indicate that cDE-ICAR algorithm is likely to have better convergence and higher stability in applications compared to other algorithms such as I-ICAR and I-cICA, there is no guarantee that the cDE-ICAR will converge to global optima of solutions and produce the desired signal without errors and artifacts. One way to avoid this problem is to perform source extraction with cDE-ICAR with different initial populations and select the best solution out of all attempts. But in our experiments, with the initialization seeded by the pseudo-inverse of the input signals, the convergence was always found the best solution.

Experiments with synthetic data show that the search continued towards the solutions with lower constraint violations as opposed to the Newton-like learning method of ICA-R where the constraints are violated. The evolutionary nature of cDE-ICAR brings the search out of local saddle points, thereby converging more efficiently and accurately

towards the global optimal solution. Though our experiments empirically demonstrated that cDE-ICAR is likely to have better convergence and higher stability as compared against CSP and other ICA based algorithms, there is no theoretical proof that guarantees the convergence of cDE-ICAR to global optima.

VI. CONCLUSIONS

In order to overcome the problems due to the penalized contrast function of ICA-R, a framework based on constrained DE learning algorithm for ICA-R (cDE-ICAR) is developed. The contrast function and the constraints were incorporated into the selection process of evolution. By setting the upper-bound of the distance function it was shown that a major bottleneck for convergence and stability in ICA-R is not required for cDE-ICAR as the distance between the reference and extracted signals is employed as a measure of contrast violation. By keeping a population of solutions, the cDE-ICAR is able to reach the global optima and extracts the desired signal. Results show increased stability of the algorithm for different mixing matrices and improved convergence while recovering the desired signal accurately. Performance evaluation of synthetic data showed statistically significant improvement in results as compared to ICA, ICA-R, I-cICA and I-ICAR methods.

REFERENCES

[1] P. Comon, "Independent component analysis: A new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[2] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ICA," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 203–212, Jan. 2005.

[3] Y. Wang, Y.-T. Wang, and T.-P. Jung, "Translation of EEG spatial filters from resting to motor imagery using independent component analysis," *PLoS ONE*, vol. 7, no. 5, p. e37665, 2012.

[4] S. A. Cruces-Alvarez, A. Cichocki, and S.-I. Amari, "From blind signal extraction to blind instantaneous signal separation: Criteria, algorithms, and stability," *IEEE Trans. Neural Netw.*, vol. 15, no. 4, pp. 859–873, Jul. 2004.

[5] W. Lu and J. C. Rajapakse, "ICA with reference," *Neurocomputing*, vol. 69, nos. 16–18, pp. 2244–2257, 2006.

[6] X. Ma, H. Zhang, X. Zhao, L. Yao, and Z. Long, "Semi-blind independent component analysis of fMRI based on real-time fMRI system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 3, pp. 416–426, May 2013.

[7] K. S. Sri and J. C. Rajapakse, "Extracting EEG rhythms using ICA-R," in *Proc. IJCNN*, Jun. 2008, pp. 2133–2138.

[8] M. De Vos, L. De Lathauwer, and S. Van Huffel, "Spatially constrained ICA algorithm with an application in EEG processing," *Signal Process.*, vol. 91, no. 8, pp. 1963–1972, 2011.

[9] J. Lee, K. L. Park, and K.-J. Lee, "Temporally constrained ICA-based foetal ECG separation," *Electron. Lett.*, vol. 41, no. 21, pp. 1158–1160, Oct. 2005.

[10] Q.-H. Lin, J. Liu, Y.-R. Zheng, H. Liang, and V. D. Calhoun, "Semiblind spatial ICA of fMRI using spatial constraints," *Human Brain Mapping*, vol. 31, no. 7, pp. 1076–1088, 2010.

[11] Q.-H. Lin, Y.-R. Zheng, F. L. Yin, and H.-L. Liang, "Speech segregation using constrained ICA," in *Advances in Neural Networks—ISNN* (Lecture Notes in Computer Science), vol. 3173, F. Yin, J. Wang, and C. Guo, Eds. Berlin, Germany: Springer, 2004, pp. 755–760.

[12] N. D. Thang, T. Rasheed, Y.-K. Lee, S. Lee, and T.-S. Kim, "Content-based facial image retrieval using constrained independent component analysis," *Inf. Sci.*, vol. 181, no. 15, pp. 3162–3174, 2011.

[13] Z.-L. Sun and K.-M. Lam, "Depth estimation of face images based on the constrained ICA model," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 2, pp. 360–370, Jun. 2011.

[14] Y. B. Monakhova, A. M. Tsikin, T. Kuballa, D. W. Lachenmeier, and S. P. Mushtakova, "Independent component analysis (ICA) algorithms for improved spectral deconvolution of overlapped signals in ^{1}H NMR analysis: Application to foods and related products," *Magn. Reson. Chem.*, vol. 52, no. 5, pp. 231–240, 2014.

[15] Z.-L. Zhang, "Morphologically constrained ica for extracting weak temporally correlated signals," *Neurocomputing*, vol. 71, no. 7, pp. 1669–1679, 2008.

[16] D.-S. Huang and J.-X. Mi, "A new constrained independent component analysis," *IEEE Trans. Neural Netw.*, vol. 18, no. 5, pp. 1532–1535, Sep. 2007.

[17] Q.-H. Lin, Y.-R. Zheng, F.-L. Yin, H. Liang, and V. D. Calhoun, "A fast algorithm for one-unit ICA-R," *Inf. Sci.*, vol. 177, no. 5, pp. 1265–1275, 2007.

[18] Y. Jia and X. Yang, "A fetal electrocardiogram signal extraction algorithm based on fast one-unit independent component analysis with reference," *Comput. Math. Methods Med.*, vol. 2016, Aug. 2016, Art. no. 5127978.

[19] Z.-L. Sun and L. Shang, "An improved constrained ICA with reference based unmixing matrix initialization," *Neurocomputing*, vol. 73, nos. 4–6, pp. 1013–1017, 2010.

[20] C. Li, G. Liao, and Y. Shen, "An improved method for independent component analysis with reference," *Digit. Signal Process.*, vol. 20, no. 2, pp. 575–580, 2010.

[21] L. Breuer, M. Axer, and J. Dammers, "A new constrained ICA approach for optimal signal decomposition in polarized light imaging," *J. Neurosci. Methods*, vol. 220, no. 1, pp. 30–38, 2013.

[22] X. Zhang and H. Duan, "An improved constrained differential evolution algorithm for unmanned aerial vehicle global route planning," *Appl. Soft Comput.*, vol. 26, pp. 270–284, Jan. 2015.

[23] J. Ronkkonen, S. Kukkonen, and K. V. Price, "Real-parameter optimization with differential evolution," in *Proc. IEEE Congr. Evol. Comput.*, Sep. 2005, pp. 506–513.

[24] S. M. Elsayed, R. A. Sarker, and D. Essam, "On an evolutionary approach for constrained optimization problem solving," *Appl. Soft Comput.*, vol. 12, no. 10, pp. 3208–3227, 2012.

[25] Z. Hu, S. Xiong, Q. Su, and X. Zhang, "Sufficient conditions for global convergence of differential evolution algorithm," *J. Appl. Math.*, vol. 2013, Aug. 2013, Art. no. 193196.

[26] Z. Zhan and J. Zhang, "Enhance differential evolution with random walk," in *Proc. 14th Int. Conf. Genetic Evol. Comput. Conf. Companion (GECCO)*, 2012, pp. 1513–1514.

[27] S. S. Kavuri, J. M. Zurada, and J. C. Rajapakse, "Evolutionary approach to ICA-R," in *Proc. IEEE World Congr. Comput. Intell.*, Barcelona, Spain, Jul. 2010, pp. 1–6.

[28] A. Hyvärinen, "New approximations of differential entropy for independent component analysis and projection pursuit," in *Proc. Adv. Neural Inform. Process. Syst. (NIPS)*, 1998, pp. 273–279.

[29] H. Hindi, "A tutorial on convex optimization II: Duality and interior point methods," in *Proc. Amer. Control Conf.*, Minneapolis, MN, USA, Jun. 2006, p. 11.

[30] J.-X. Mi, "A novel algorithm for independent component analysis with reference and methods for its applications," *PLoS ONE*, vol. 9, no. 5, p. e93984, 2014.

[31] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, pp. 341–359, 1997.

[32] F. Wu, *A Framework for Memetic Algorithms*. Auckland, New Zealand: Univ. Auckland, 2001.

[33] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evol. Comput.*, vol. 8, no. 2, pp. 173–195, 2000.

[34] R. Gämperle, S. D. Müller, and P. Koumoutsakos, "A parameter study for differential evolution," *Adv. Intell. Syst., Fuzzy Syst., Evol. Comput.*, vol. 10, no. 10, pp. 293–298, 2002.

[35] B. Blankertz *et al.*, "The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1044–1051, Jun. 2004.

[36] F. Sharbrough, "American electroencephalographic society guidelines for standard electrode position nomenclature," *J. Clin. Neurophysiol.*, vol. 8, no. 2, pp. 200–202, 1991.

[37] G. Blanchard and B. Blankertz, "BCI competition 2003-data set IIA: Spatial patterns of self-controlled brain rhythm modulations," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1062–1066, Jun. 2004.

[38] W. Lu and J. C. Rajapakse, "ICA with reference," in *Proc. 3rd Int. Conf. Independent Component Anal. Blind Signal Separation (ICA)*, San Diego, CA, USA, 2001.

SWATHI SRI KAVURI received the bachelor's degree in biomedical engineering from Jawaharlal Nehru Technological University, India, in 2005, and the M.Sc. and Ph.D. degrees from Nanyang Technological University, Singapore, in 2006 and 2016, respectively. From 2012 to 2014, she was a Research Associate with the Artificial Brian Research lab, Kyungpook National University, Daegu, South Korea, where she is currently with the School of Electronics Engineering. Her current areas of research include biomedical signal processing, component analysis, EEG-based emotion recognition, convolution neural networks, and sentiment analysis.

KALYANA C. VELUVOLU (S'03–M'06–SM'13) was born in 1981. He received the B.Tech. degree in electrical and electronic engineering from Acharya Nagarjuna University, Guntur, India, in 2002, and the Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 2006. From 2006 to 2009, he was a Research Fellow with the Biorobotics Group, Robotics Research Center, Nanyang Technological University. Since 2009, he has been with the School of Electronics Engineering, Kyungpook National University, Daegu, South Korea, where he is currently an Associate Professor. He was with the School of Mechanical and Aerospace Engineering, Nanyang Technological University, as a Visiting Professor from 2016 to 2017. He has been a Principal Investigator or a Co-Investigator on a number of research grants funded by the National Research Foundation of Korea and other agencies. He has authored or co-authored over 120 journal and conference proceedings articles. His current research interests include nonlinear estimation and filtering, sliding mode-control, brain–computer interface, autonomous vehicles, biomedical signal processing, and surgical robotics.

QUEK HIOK CHAI (M'96–SM'10) received the bachelor's degree in science and the Ph.D. degree from the Heriot-Watt University, Edinburgh. He has been with the School of Computer Engineering, Nanyang Technological University, since 1990. His research interests include neurocognitive informatic, biomedical engineering, and computational finance. He has done significant research work in these research areas and published over 200 top quality international conference and journal papers. He is a member of the IEEE Technical Committee on Computational Finance and Economics. He has constantly and successfully groomed several high calibre research manpower who are awarded prestigious Singapore Millenium Foundation Scholar and Fellowships, Lee Kuan Yew Fellowship, and A*Star Scholarhips. He has been often invited as a program committee member or referee and reviewer for a number of premier conferences and journals, including the IEEE TNN and TEvC.

• • •