Available online at www.sciencedirect.com
SciVerse ScienceDirect
Magnetic Resonance Imaging 29 (2011) 1288 – 1303
MAGNETIC RESONANCE IMAGING

Fixed-point algorithms for constrained ICA and their applications in fMRI data analysis

Ze Wang*

Department of Psychiatry, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 14 March 2011; revised 18 May 2011; accepted 6 July 2011

Abstract

Constrained independent component analysis (CICA) eliminates the order ambiguity of standard ICA by incorporating prior information into the learning process to sort the components intrinsically. However, the original CICA (OCICA) and its variants depend on a learning rate, which is not easy to be tuned for various applications. To solve this problem, two learning-rate-free CICA algorithms were derived in this paper using the fixed-point learning concept. A complete stability analysis was provided for the proposed methods, which also made a correction to the stability analysis given to OCICA. Variations for adding constraints either to the components or to the associated time courses were derived too. Using synthetic data, the proposed methods yielded a better stability and a better source separation quality in terms of higher signal-to-noise-ratio and smaller performance index than OCICA. For the artificially generated brain activations, the new CICAs demonstrated a better sensitivity/specificity performance than standard univariate general linear model (GLM) and standard ICA. Original CICA showed a similar sensitivity/specificity gain but failed to converge for several times. Using functional magnetic resonance imaging (fMRI) data acquired with a well-characterized sensorimotor task, the proposed CICAs yielded better sensitivity than OCICA, standard ICA and GLM in all the target functional regions in terms of either higher *t* values or larger suprathreshold cluster extensions using the same significance threshold. In addition, they were more stable than OCICA and standard ICA for analyzing the sensorimotor fMRI data.

Keywords: Independent component analysis; Constrained independent component analysis; Augmented Lagrangian; Fixed-point iteration; Newton-like gradient; General linear model; fMRI

1. Introduction

Independent component analysis (ICA) [1] is an advanced signal processing method designed to separate mutually independent components (ICs) from their observed mixtures without using any prior information of them. As a multivariate and data-driven method, ICA has gained popularity in many signal processing fields [1], including functional magnetic resonance imaging (fMRI) data analysis [2–4]. Without need of prior modeling for brain response, ICA bestows a large freedom to analyzing fMRI data with various kinds of tasks or even the null hypothesis data [5,6]. However, standard ICA does not extract ICs with a fixed order [7], resulting in a problem of IC comparisons across different runs or across different subjects. Constrained ICA (CICA) [8,9] provides a solution for this problem. In this approach, prior information contained in constraints is integrated into the source separation process, so that the decomposed ICs are sorted intrinsically. This idea was first proposed by Luo et al. [10] but was explicitly defined by Lu and Rajapakse who also provided a CICA algorithm based on a more advanced ICA method [8,11]. For this reason, the conceptual framework of CICA proposed by Lu and Rajapakse will be employed in this paper, though a similar approach, the semi-blind ICA, has later been proposed by Calhoun et al. [12].

* Department of Psychiatry, Treatment Research Center, University of Pennsylvania, Philadelphia, PA 19104, USA. Tel.: +1 215 222 3200; fax: +1 215 386 6770.
E-mail addresses: zewang@mail.med.upenn.edu, redhatw@gmail.com.

0730-725X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.mri.2011.07.017

One problem of the original CICA (OCICA) algorithm proposed by Lu and Rajapakse [8,11] is that it uses a learning rate to control the updating gradient at each iteration. This learning rate is not easy to tune for various applications, and a bad choice of it could completely destroy the algorithm convergence [1,13]. For fMRI data analysis, the OCICA was directly adopted to extract temporally independent components using temporal constraints. As extracting spatially independent components represents another major interest of ICA-based fMRI data analysis [2,14], CICA should be adapted to have these two capabilities, as well as applying either spatial or temporal constraints. Moreover, the stability analysis for CICA given in Ref. [11] was incomplete and was based on a nonestablished condition, so a complete stability analysis is still required to prove the convergence of CICA.

We have recently proposed a fixed-point CICA in a conference abstract [15], which does not need a learning rate. The purpose of this paper is to provide full derivations, comprehensive evaluations and detailed convergence analysis for the fixed-point CICA algorithms. Method evaluations were conducted using synthetic one-dimensional (1D) signal and two-dimensional (2D) fMRI data and sensorimotor fMRI data.

2. Theory

2.1. ICA with constraints and the OCICA algorithm

Assuming the observed signal $\mathbf{x}=(x_1, x_2,\dots, x_n)^T$ to be a linear mixture of some unknown mutually independent sources $\mathbf{s}=(s_1, s_2,\dots, s_m)^T$: $\mathbf{x}=\mathbf{A}\mathbf{s}$, where $\mathbf{A}$ is an unknown $n\times m$ matrix ($m\le n$), standard ICA [7] is to seek an $m\times n$ unmixing matrix $\mathbf{W}=(\mathbf{w}_1, \mathbf{w}_2,\dots, \mathbf{w}_m)^T$, such that s can be approximately recovered by $\mathbf{y}=\mathbf{W}\mathbf{x}$, with $\mathbf{y}=(y_1, y_2,\dots, y_m)^T$. In general, ICA can be formed and solved as an optimization problem to minimize an object function $\mathcal{O}(.)$ that measures the independence of the output components. One such function used in FastICA [1,16] and OCICA is:

$$\mathcal{O}(\mathbf{y}) = -\sum_i \mathcal{J}(y_i), \tag{1}$$

where $\mathcal{J}$ (.) is the negentropy and can be approximated by Ref. [17]:

$$\mathcal{J}(\mathbf{y}_i) \approx c[E\{G(\mathbf{y}_i)\} - E\{G(v)\}]^2, \tag{2}$$

where c is a positive constant, $G(.)$ is a nonquadratic function and v is a Gaussian variable with a zero mean and a unit variance. This approximation was used in the OCICA and will be used in our proposed methods as well.

Due to the blindness to $\mathbf{s}$ and $\mathbf{A}$, ICA is up to two ambiguities [1]: the scale ambiguity and the order ambiguity. Since signal shape is usually more informative than the amplitude in practice, the scale ambiguity is generally less problematic than the order ambiguity. To eliminate the order ambiguity, prior information was incorporated to constrain the directions of IC learning in CICA [8,11]. Denote the constraint function by $(\mathbf{y})=(q_1(y_1),\dots q_M(y_M))^T$ ($M\le m$) and convert them into equality constraints: $q_i(y_i)+z_i^2=0$ ($i=1,\dots M$, and z_i is a slack variable). OCICA was solved using an augmented Lagrangian function [13]:

$$\begin{aligned}\mathfrak{L}(\mathbf{W},\mu,z,\lambda) = \mathcal{O}(y) + \mu^T\left(q(y)+z^2\right) \\ + \frac{1}{2}\gamma||q(y)+z^2||^2 \\ + \lambda^T h(y) + \frac{1}{2}\gamma||h(y)||^2,\end{aligned} \tag{3}$$

where $\mu=(\mu_1,\dots, \mu_M)^T$ and $\lambda=(1_1,\dots, \lambda_N)^T$ are two sets of Lagrange multipliers, γ is the scalar penalty parameter, $\mathbf{z}=(z_1,\dots z_i,\dots, z_M)^T$ and $\|\cdot\|$ means Euclidean norm. $\mathbf{h}(\mathbf{y})=(h_1(\mathbf{y}),\dots, h_N(\mathbf{y}))^T$ are the N equality constraints: $E\{(\mathbf{Wx})^2\}=\mathbf{I}$ used to facilitate an easy ICA solution [1]. The last item, $1/2\gamma\|\cdot\|^2$, is the penalty term to ensure that the local convexity assumption holds at a solution of the optimization problem [18]. Since the optimal value of $\mathbf{z}$ can be obtained as $(z^*)^2 = max\left\{0, -\frac{1}{\gamma}[\mu+\gamma q(y)]\right\}$ via maximizing $\mathfrak{L}_1(\mathbf{W},\mu,\mathbf{z}) = \mu^T(\mathbf{q}(\mathbf{y})+\mathbf{z}^2) + \frac{1}{2}\gamma||\mathbf{q}(\mathbf{y})+\mathbf{z}^2||^2$ with respect to (wrt) $\mathbf{z}^2$, the augmented Lagrangian function in Eq. (3) can be rewritten as:

$$\mathfrak{L}_{\mathbf{z}^*}(\mathbf{W},\mu,\lambda) = \mathfrak{L}(\mathbf{y}) + \mathfrak{L}_{1\mathbf{z}^*}(\mathbf{W},\mu) + \lambda^T\mathbf{h}(\mathbf{y}) + \frac{1}{2}\gamma||\mathbf{h}(\mathbf{y})||^2, \tag{4}$$

where $\mathfrak{L}_{1\mathbf{z}^*} = \mathfrak{L}_1(\mathbf{W},\mu,\mathbf{z}^*)$ (see Eq. A.4). Without introducing any ambiguity, $\mathfrak{L}_z{*}$ and $\mathfrak{L}$ will be used interchangeably hereafter in this paper, and so will be $\mathfrak{L}_{1z}{*}$ and $\mathfrak{L}_l$. A Newton-like gradient method can be used to solve the optimization problem of CICA [8]:

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \eta\nabla_{\mathbf{w}}\mathfrak{L} / \nabla_{\mathbf{W}}\nabla_{\mathbf{W}}\mathfrak{L}, \tag{5}$$

where ∇_{W} and $\nabla_{\mathrm{W}}\nabla_{\mathrm{W}}$ mean the first and second derivative, respectively, and η is the learning rate. This algorithm involves an inversion of the Hessian matrix, which is usually computationally intensive even when an approximation was used [8,11]. Nevertheless, this matrix inversion can be avoided [19] upon using data prewhitening that is now a general preprocessing step in ICA [1].

We can then get $\nabla_W\mathfrak{L}_1 = [\nabla_{w1}\mathfrak{L}_1,\dots,\nabla_{w_m}\mathfrak{L}_1]$ and $\nabla_W\nabla_W\mathfrak{L}_1 = [\nabla_{W_1}\nabla_{W_1}\mathfrak{L}_1,\dots,\nabla_{W_m}\nabla_{W_m}\mathfrak{L}_1]$ with:

$$\nabla_{\mathbf{w}_p}\mathfrak{L}_1 = \mu_p\nabla_{\mathbf{w}_p}q_p(y_p) = \mu_p E\{\mathbf{x}q'_p(y_p)\} \tag{6a}$$

$$\begin{aligned}\nabla_{\mathbf{w}_p}\nabla_{\mathbf{w}_p}\mathfrak{L}_1 = \mu_p E\{\mathbf{x}^T\mathbf{x}q''_p(y_p)\} \approx \mu_p E\{\mathbf{x}^T\mathbf{x}\}E\{q''_p(y_p)\} \\ = \mu_p E\{q''_p(y_p)\}\end{aligned} \tag{6b}$$

Since the maxima of $\mathfrak{J}(y_i)$ is approached at certain optima of $E\{G(\mathbf{y}_i)\}$, the second item in Eq. (2) can be ignored during the process of pursuing the optimal $\mathbf{W}$ [16].