Accurate Estimation of ICA Weight Matrix by Implicit Constraint Imposition Using Lie Group

IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 20, NO. 10, OCTOBER 2009

S. Easter Selvan, Alexandru Mustățea, C. Cecil Xavier, and Jean Sequeira

February 10, 2012

Outline

Introduction I

GOAL: "To propose an algorithm to optimize the independence criterion among multivariate data using local, global, and hybrid optimizers, in conjunction with techniques involving a Lie group and its corresponding Lie algebra, for implicit imposition of the orthonormality constraint among the estimated sources"

Introduction I

- ICA Constraint: $WW^T = I$
- To preserve the orthogonality constraint at all time can be handled by techniques which will “lock” the weight matrix onto the constraint surface.
- The IC estimation approaches using a parameterization method such as the Lie group for implicit imposition of the orthonormality constraint, which have been reported in the literature, were designed to yield only a local optimal solution.
- Because the optimization landscape of a contrast function used for estimating the ICs of a multivariate data is nonconvex, the methods we have developed here are to look for near-global-optimum solutions.

Contributions I

1. The use of global and hybrid optimizers in conjunction with a Lie group, to **maximize the independence criterion**, has never been explicitly studied before.
2. The issue of computational complexity is addressed, by performing the **spectral screening of the image data** prior to the IC estimation.
3. Recommended here is **a local optimizer**—the quasi- Newton method—employing Lie group techniques, which can yield more accurate IC estimates, compared to a conventional technique.

Outline

Lie Group of IC Weight Matrices I

The possible IC weight matrices W belong to the set of nxn orthogonal matrices such that $WW^T = I$ form a **Lie group** under matrix multiplication because the following group axioms are satisfied:

- closed under matrix multiplication, associativity, identity element, and existence of an inverse;
- the group operations are differentiable;

Lie Group of IC Weight Matrices I

- The Lie group formed by the set of *nxn* orthogonal matrices with determinant +1, called the "special" orthogonal matrices $SO(n)$, is applied in the estimation of the ICs.
- For every Θ lying in the space of skew-symmetric matrices $so(n)$, there exists a corresponding weight matrix W, which lies in the space of orthogonal matrices $SO(n)$

$$\mathbf{W} = \exp \mathbf{\Theta}.$$

$$\mathbf{\Theta} = \begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix}$$

$$\widehat{\mathbf{W}} = (a, b, c, d, e, f)$$

Lie Group of IC Weight Matrices II

- The orthonormality constraint among the estimated ICs will never be violated during the course of optimization, in the proposed approach using the Lie group.

Outline

MI as Contrast Function I

The mutual information (MI) was chosen to be the suitable contrast function, to separate the sources with mutually exclusive specific information from the multivariate data:

- The MI, expressed as the negative sum of **negentropy approximations**, is the natural measure of dependence.
- MI:

$$I(z_1, z_2, \ldots, z_n) \propto -\sum_i J(z_i)$$

where $J(z_i) \propto [E\{G(z_i)\} - E\{G(\gamma)\}]^2$ are the negentropy approximations based on the maximum-entropy principle computed for the linear combinations $z_i = \mathbf{w}_i^T \mathbf{x}, i = 1, 2, \ldots, n$

MI as Contrast Function I

The nonquadratic function is selected from the following choices G_1 and G_2, to compute the negentropy approximations

$$G_1(z_i) = \frac{1}{a_1}\log\cosh(a_1 z_i)$$
$$G_2(z_i) = -\frac{1}{a_2}\exp\left(-a_2\frac{z_i^2}{2}\right)$$

The overall optimization problem is:

$$\min I(z_1, z_2, \ldots, z_n) \approx -\sum_i \left[E\{G(z_i)\} - E\{G(\gamma)\}\right]^2$$

subject to the constraint $WW^T = I$

Outline

Local Minimization of MI with Lie Group Techniques I

MI minimization using the quasi-Newton method and the Lie group, reasons:

1. It uses both gradient and curvature informations about the optimization landscape during the search process,
2. Hessian update can be obtained using only gradient information, the computational overhead is relatively small, the method converges superlinearly, and it performs well for cases where the steepest ascent methods suffer from convergence difficulty

Implementation

- The weight matrix W_1 corresponding to a randomly chosen representative vector $\widehat{W_1}$ of dimension $n(n-1)/2$,

$$\mathbf{W}_{m+1} = \exp\left(-\alpha_m \Theta_{\boldsymbol{\delta}_m}\right) \mathbf{W}_m$$

m the iteration number.

- Optimum is tertermined by the line-search iterative technique employing the cubic/quadratic polynomial method

$$f(\mathbf{W}_{m+1}) < f(\mathbf{W}_m).$$

- The skew-symmetric matrix $\Theta_{\boldsymbol{\delta}_m}$ is constructed from the search direction $\boldsymbol{\delta}_m$

Local Minimization of MI with Lie Group Techniques II

$$\boldsymbol{\delta}_m = \mathbf{H}_m^{-1}\mathbf{g}_m.$$

- $\mathbf{H}_m$ is the Hessian update of the contrast function. g_m is the representative vector of the gradient of in the Lie algebra space that is:

$$\frac{\partial I}{\partial \boldsymbol{\Theta}_m} = \left(\frac{\partial I}{\partial \mathbf{W}_m}\right)\mathbf{W}_m^T - \mathbf{W}_m\left(\frac{\partial I}{\partial \mathbf{W}_m}\right)^T$$

- Hessian update:

$$\mathbf{H}_{m+1} = \mathbf{H}_m + \frac{\boldsymbol{\eta}_m\boldsymbol{\eta}_m^T}{\boldsymbol{\eta}_m^T\mathbf{s}_m} - \frac{\mathbf{g}_m\mathbf{g}_m^T}{\boldsymbol{\delta}_m^T\mathbf{g}_m}$$

$$\mathbf{s}_m = \widehat{\mathbf{W}}_{m+1} - \widehat{\mathbf{W}}_m \text{ and } \boldsymbol{\eta}_m = \mathbf{g}_{m+1} - \mathbf{g}_m$$

Local Minimization of MI with Lie Group Techniques III

- The MI is computed as the negative sum of the negentropy approximations of the data values in each dimension.

Outline

Near-global Minimization of MI with Lie Group Techniques I

- It is imperative to design the global optimizers in conjunction with the Lie group, to produce more accurate IC estimates, which are insensitive to the initial choice of the random input vectors.
  - Data processing using spectral screening
  - Simulated Annealing With Lie Group Techniques
  - Cross-Entropy Method With Lie Group Techniques

Data processing using spectral screening I

- To make the estimation of the ICs in very large data computationally feasible with the global optimizers, instead of taking the entire set of data vectors into account, a **small subset of data vectors** can be used such that a few vectors represent each group of "similar" vectors.
- The similarity measure between the data vectors and is the **spectral angle**:

$$\alpha(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}||_2 ||\mathbf{y}||_2}\right).$$

- **Algorithm:**
  - S: The set of all pixel vectors present in the multispectral image.
  - One of the pixel vectors is placed in the subset S_1.
  - Each pixel vector from is compared with all the pixel vectors already placed in S_1.

Data processing using spectral screening II

- For all $\mathbf{x} \in S_1$, if the similarity measure $\alpha(\mathbf{x}, \mathbf{y})$ in (11) is greater than the angle threshold β, then $\mathbf{y}$ is included in S_1; otherwise, it is discarded. At the end of the process, S_1 will contain the representative pixel vectors from S that are quite dissimilar.

Outline

Minimization of MI using Hybrid Optimizers and Lie Group Techniques I

- There is an improvement in the quality of the solution while using the Lie group with the global optimization techniques. Problem: they are being trapped in the local minima.
- To avoid this: Hybrid optimizers.
  - PSO (global optimizer) with the quasi-Newton local optimizer during every iteration (**PSO-QN**) or at a periodic intervals (**PSO-periodic QN**).
  - FastICA (local optimizer) to minimize the MI produced by the PSO in all the iterative steps (**PSO-fastICA**) or periodically (**PSO-periodic fastICA**)

Minimization of MI using Hybrid Optimizers and Lie Group Techniques

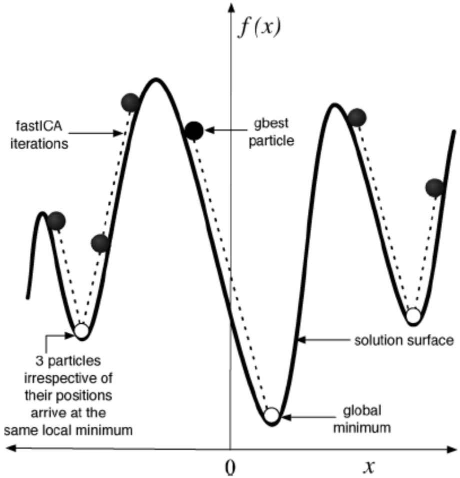


Fig. 2. Implementation of the hybrid optimizer—a combination of the PSO and fastICA.

Minimization of MI using Hybrid Optimizers and Lie Group Techniques I

- In the PSO-QN algorithm, the quasi-Newton method minimizes the MI by treating the weight vectors $(\widehat{W}_i(1))$ as the set of inputs; the resulting weight vectors related to the local minima are the updated particles' positions in the first iteration.
- In the PSO-fastICA algorithm, the weight matrices $W_i(1)$ corresponding to $\widehat{W}_i(1)$ serve as the inputs for the fastICA algorithm employing symmetric orthogonalization to minimize the MI.
- Instead of updating the particles' positions based on the fastICA results, the local minimum MI values are assigned to the particles' initial positions itself.
- In the periodic versions, at iteration k, the search positions $x_i(k)$ of the particles:

Minimization of MI using Hybrid Optimizers and Lie Group Techniques II

$$\mathbf{v}_i(k+1) = \chi\,[\mathbf{v}_i(k) + \varphi_1 * \mathbf{rand1}_i * (\mathbf{pbest}_i(k) - \mathbf{x}_i(k)) + \varphi_2 * \mathbf{rand2}_i * (\mathbf{gbest}(k) - \mathbf{x}_i(k))] \quad (18)$$

$$\mathbf{x}_i(k+1) = \mathbf{x}_i(k) + \mathbf{v}_i(k+1) \quad (19)$$

$$\chi = \frac{2\kappa}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|}$$

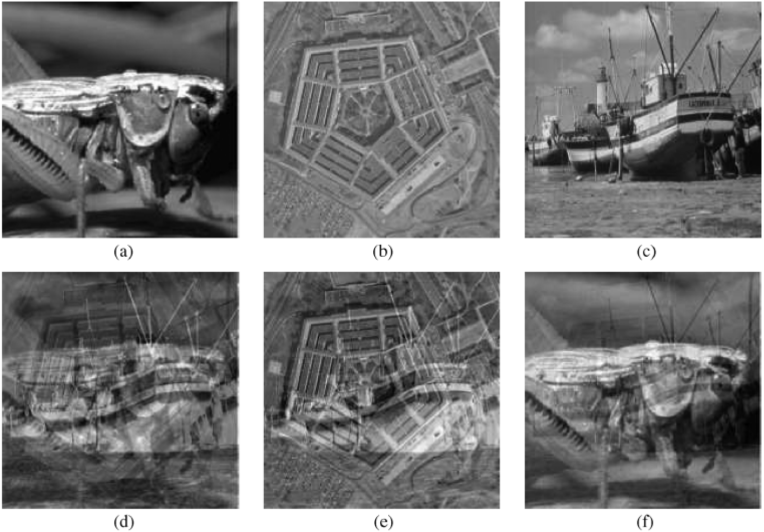
$$\varphi = \varphi_1 + \varphi_2, \qquad \varphi > 4$$
$$\kappa \in [0, 1].$$

Minimization of MI using Hybrid Optimizers and Lie Group Techniques

- While implementing the PSO-fastICA algorithm, the weight matrices $W_i(k+1)$ corresponding to $\widehat{W}_i(k+1)$ serve as the inputs for the fastICA algorithm to assign the local minimum MI values to $x_i(k+1)$.

Outline

Experimental Results and Discussion I



Experimental Results and Discussion

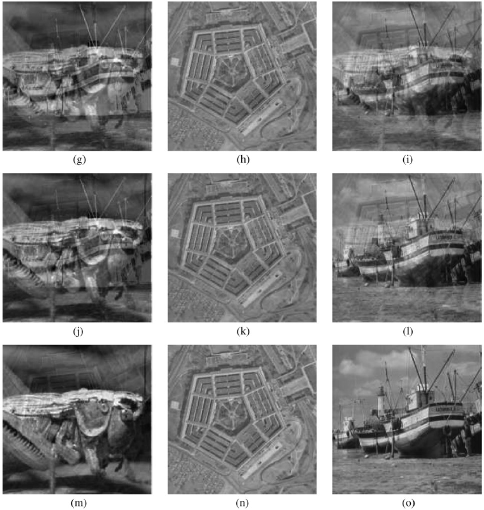


Fig. 3. (a)–(c) Original grayscale images of size 200 × 200. (d)–(f) Mixed images generated with an orthonormal mixing matrix given in Section VII-B. (g)–(i) Source images estimated with the fastICA method, resulting in an MI value of -0.000348. (j)–(l) Source images estimated with the PSO-fastICA algorithm, that yielded the least MI value of -0.000496. (m)–(o) Source images estimated with the PSO-fastICA algorithm, using the spectral screened input data of size 344 × 3.

Experimental Results and Discussion

TABLE I
PERFORMANCE EVALUATION OF FASTICA AND QUASI-NEWTON ALGORITHM WITH LIE GROUP

Local optimizer minimizing MI	G_1			G_2		
	MI mean	MI std. dev.	least MI %	MI mean	MI std. dev.	least MI %
FastICA	-0.016217	0.016045	0	-0.018909	0.014836	0
QN numerical gradient	**-0.020256**	0.016127	44	-0.022559	0.015535	42
QN analytical gradient	-0.020204	0.015890	56	**-0.022680**	0.015479	58

Experimental Results and Discussion

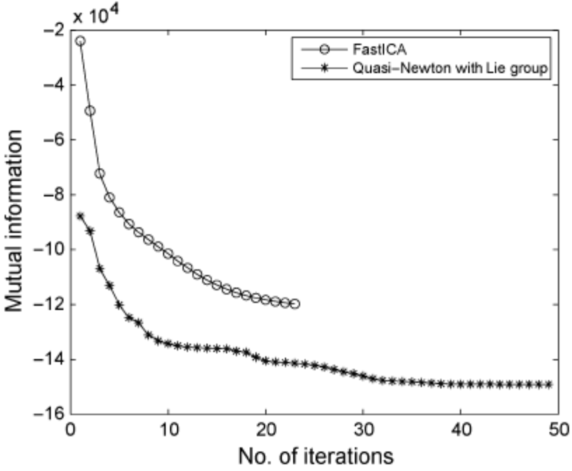


Fig. 4. Convergence plot of the local optimizers showing faster convergence while employing the fastICA, compared to the quasi-Newton approach with the Lie group.

Experimental Results and Discussion

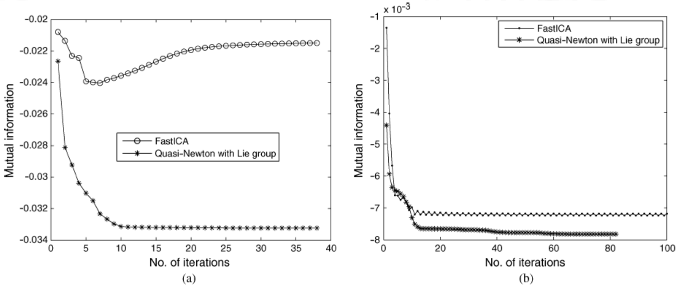


Fig. 5. (a) Poor convergence characteristics of the fastICA, as compared to the quasi-Newton approach. (b) MI minimization is oscillatory in the fastICA, whereas the quasi-Newton approach converges monotonically in a few iterations.

Experimental Results and Discussion

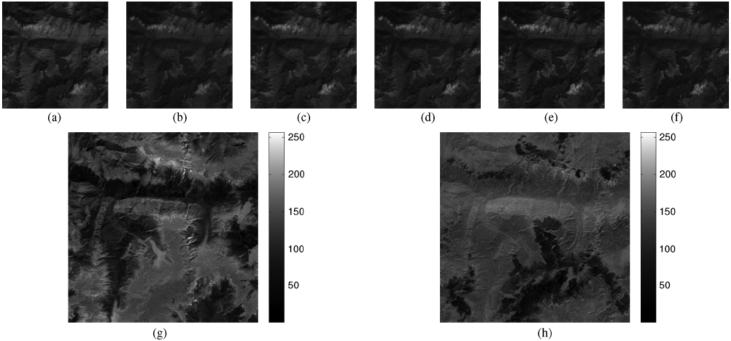


Fig. 7. (a)–(f) The 6-D multispectral ASTER image sections of size 500×500, acquired at different ranges of wavelength, as specified in Section VII. (g) and (h) IC source images generated using the weight matrix estimate from the PSO-QN, while inputting the image sections shown in (a)–(f), which were selected from the variance scree plot.

Experimental Results and Discussion

TABLE II
PERFORMANCE EVALUATION OF GLOBAL AND HYBRID ALGORITHMS WITH LIE GROUP

Global/hybrid optimizer minimizing MI	G_1				G_2			
	MI mean	MI std. dev.	least MI %	MI $<$ multi-start fastICA (%)	MI mean	MI std. dev.	least MI %	MI $<$ multi-start fastICA (%)
PSO	-0.021776	0.018016	0	56	-0.023526	0.016994	0	32
EP	-0.017241	0.014606	0	0	-0.018378	0.013757	0	0
SA	-0.020617	0.017332	0	30	-0.022665	0.016355	0	38
CE	-0.020771	0.018054	0	22	-0.021672	0.014990	0	20
PSO-fastICA	-0.023162	0.019688	0	96	-0.025207	0.017950	2	94
PSO-periodic fastICA	-0.023069	0.019719	0	86	-0.024956	0.017619	0	86
PSO-QN	**-0.023431**	0.019707	80	100	**-0.025499**	0.017960	76	100
PSO-periodic QN	-0.023419	0.019709	20	100	-0.025465	0.017829	22	100
Multi-start fastICA	-0.022253	0.019620	0	—	-0.024207	0.017891	0	—

Experimental Results and Discussion

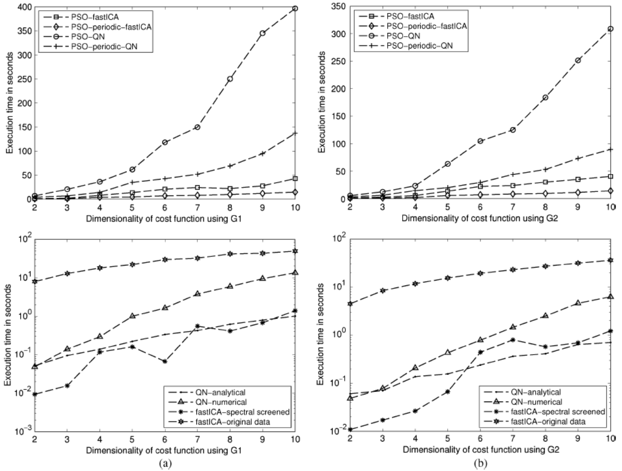


Fig. 8. Relationship between the average CPU time taken by the various algorithms, implemented with the same stopping criterion, and the data dimension: (a) using G_1; (b) using G_2. The proposed approaches were supplied with the spectral screened input, to illustrate how the execution time of the PSO-fastICA and the PSO-periodic fastICA could be brought lower than that of the fastICA implemented with the original image input of size 200 × 200. Implementing the quasi-Newton method with the analytical gradient makes it comparable with the fastICA, and the CPU time noticeably reduces. Notice that the y-axis is logarithmic for the bottom row.

Outline

Conclusion I

- Source separation algorithms are proposed and investigated for accurately estimating the weight matrix in the ICA model, with the help of Lie group techniques.
- We have proposed an approach to use a local optimizer, the quasi-Newton method, in conjunction with the Lie group, to impose the orthonormality constraint implicitly.
- This approach produces more accurate IC estimates in comparison with the fastICA, provided both the approaches are supplied with the same initial random input vector.
- We have attempted an approach wherein the global optimizers, the PSO, EP, SA, and CE method, are implemented with the Lie group.
- We have demonstrated how the variations of the hybrid optimizers, preserve a reasonable estimation accuracy by periodically integrating the local optimizers with the global one.