# GENERALISED FASTICA FOR INDEPENDENT SUBSPACE ANALYSIS

*Hao Shen and Knut Hüper*

National ICT Australia, Canberra Research Laboratory
Locked Bag 8001, Canberra ACT 2601, Australia
*and*
Department of Information Engineering, RSISE,
The Australian National University, Canberra ACT 0200, Australia
Hao.Shen@rsise.anu.edu.au, Knut.Hueper@nicta.com.au

## ABSTRACT

Independent Subspace Analysis (ISA) was developed as an extension of Independent Component Analysis (ICA) when statistical independences are assumed to exist between groups of components rather than between individual components. Due to the superiority of FastICA against other linear ICA algorithms, an intuitive analogy, the so-called FastISA algorithm, has been proposed to solve the problem of ISA. Experimental evidences so far have shown the capability of FastISA, regardless of any independence criterion. Since standard FastICA can be viewed as a special case of an approximate Newton ICA method and moreover can be generalised as a scalar shifted fixed point algorithm, in this work, we propose two new classes of ISA algorithms, an approximate Newton-like ISA method and a matrix shifted fixed point ISA algorithm on the Graßmann manifold. As an aside, FastISA is a special case in the class of matrix shifted fixed point ISA algorithms. Performances of the proposed algorithms are investigated by numerical experiments.

***Index Terms***— Independent Subspace Analysis (ISA), Graßmann manifold, Newton-like method on manifolds, FastICA, FastISA.

## 1. INTRODUCTION

In the past two decades, Independent Component Analysis (ICA) [1] has attracted enormous attention from various communities. Many efficient ICA algorithms have been proposed and used in various application areas. However, the application of many ICA algorithms is often limited since the standard ICA model requires mutual statistical independence between all components. In large scale ICA problems, generally, there are often groups of independent components which have statistical dependences within one group, and are statistically independent from any component in other groups. Such problems can be tackled by a technique now referred to as Independent Subspace Analysis (ISA) [2]. The standard ICA problem can just be considered as a special case of ISA, which extracts one-dimensional independent subspaces.

The FastICA algorithm is one prominent ICA algorithm proposed by the Finnish school [3]. It computes one independent component at a time. An intuitive analogy, the so-called FastISA algorithm, has been proposed to solve the problem of ISA [4]. Regardless of the choice of independence criterion, FastISA has shown its capacity of extracting independent subspaces. Recent work by the present authors show that standard FastICA can essentially be regarded as an algorithm on real projective space, which is a special case of a Graßmann manifold, see [5] and references therein for more details. Simultaneously, it has been shown that FastICA can be viewed as a special case of an approximate Newton ICA method, which eliminates the sign flipping behavior of the original FastICA, and furthermore has been generalised as a scalar shifted fixed point algorithm.

In this work, we study the ISA problem to compute only one independent subspace. Two classes of ISA algorithms living on the Graßmann manifold are proposed: (i) an approximate Newton-like ISA method, and (ii) a matrix shifted fixed point ISA algorithm. It turns out that FastISA is just a special case in the class of matrix shifted fixed point ISA algorithms. The performance of the proposed algorithms is investigated by numerical experiments.

## 2. APPROXIMATE NEWTON-LIKE ISA METHODS

Recall the demixing linear ISA model, formulated by the relation $Y = X^\top W$, where $W = [w_1, \dots, w_n] \in \mathbb{R}^{m\times n}$ is the whitened observation, the matrix $X \in \mathbb{R}^{m\times p}$ is the demixing matrix, and $Y \in \mathbb{R}^{p\times n}$ represents a group of $p$ signals, see [2]. If a demixing matrix $X^* \in \mathbb{R}^{m\times p}$ with $\operatorname{rk} X^* = p$ extracts a statistical independent group of $p$ signals, then a different ma-

National ICT Australia Limited is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through *Backing Australia's Ability* and the ICT Research Centre of Excellence programs.

trix $X \in \mathbb{R}^{m\times p}$ with $\operatorname{span} X = \operatorname{span} X^*$ represents the same independent group. The set of all $p$-dimensional linear subspaces of $\mathbb{R}^m$ is defined as the Graßmann manifold $Gr(p,m)$.

Instead of representing $Gr(p,m)$ as a homogeneous space [6, 7], we identify $Gr(p,m)$ here with the set of rank $p$ symmetric projection operators on $\mathbb{R}^m$ [8], i.e.,

$$Gr(p,m) := \{P \in \mathbb{R}^{m\times m} | P = P^\top, P^2 = P, \operatorname{tr} P = p\}. \quad (1)$$

Let $SO(m) := \left\{\Theta \in \mathbb{R}^{m\times m} | \Theta^\top\Theta = 1, \det(\Theta) = 1\right\}$ denote the special orthogonal group. One can represent any point $P \in Gr(p,m)$ by $P = \Theta \left[\begin{smallmatrix} I_p & 0 \\ 0 & 0\end{smallmatrix}\right] \Theta^\top$ with a suitable $\Theta \in SO(m)$. Now let us generalise the cost function for ISA already used in [2] as follows,

$$F\colon Gr(p,m) \to \mathbb{R}, \quad F(P) := \tfrac{1}{2}\mathbb{E}_i[G(w_i^\top P w_i)], \quad (2)$$

where $G : \mathbb{R} \to \mathbb{R}$ is a smooth function and $\mathbb{E}_i[\cdot]$ denotes the empirical mean over the index $i$. For the sake of simplicity, in the sequel, we will omit the index $i$. Following [7, 9], we will derive an approximate Newton-like ISA method on $Gr(p,m)$ by optimising the cost function (2).

Recall that a smooth local parameterisation of the Graßmann manifold around $P \in Gr(p,m)$ is as follows

$$\begin{aligned} &\mu_P : \mathbb{R}^{(m-p)\times p} \to Gr(p,m), \\ &\mu_P(Z) := \Theta \exp\begin{bmatrix} 0 & -Z^\top \\ Z & 0 \end{bmatrix} \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix} \exp\begin{bmatrix} 0 & Z^\top \\ -Z & 0 \end{bmatrix} \Theta^\top, \end{aligned} \quad (3)$$

where $\Theta \in SO(m)$ and $P = \mu_P(0) = \Theta \left[\begin{smallmatrix} I_p & 0 \\ 0 & 0\end{smallmatrix}\right] \Theta^\top \in Gr(p,m)$. After composing $F$ with $\mu_P$, the first derivative of $F \circ \mu_P$ at 0 can be computed as follows,

$$\tfrac{\mathrm{d}}{\mathrm{d}\,\varepsilon} \left(F \circ \mu_P\right)(\varepsilon Z)\big|_{\varepsilon=0} = \operatorname{tr} Z^\top K_{21}, \quad (4)$$

where $K_{11} \in \mathbb{R}^{p\times p}$, $K_{21} \in \mathbb{R}^{(m-p)\times p}$ and

$$\begin{bmatrix} K_{11} & K_{21}^\top \\ K_{21} & K_{22} \end{bmatrix} := \mathbb{E}[G'(w^\top P w)\Theta^\top w w^\top \Theta]. \quad (5)$$

Here the term $K_{21}$ is a function in $P$, i.e., $K_{21} : Gr(p,m) \to \mathbb{R}^{(m-p)\times p}$. Thus to characterise the critical points of $F$, we need to study when the expression in (5) vanishes, i.e., when $\operatorname{tr} Z^\top K_{21}(P) = 0$. Due to the fact that this critical point condition depends not only on the function $G$ but also on the statistical properties of the signals, it is hardly possible to characterise all critical points of $F$ in detail. Nevertheless, it can be shown that any correct subspace extraction point $P^* \in Gr(p,m)$ is a critical point of $F$, i.e.,

$$K_{21}(P^*) = 0. \quad (6)$$

It is worthwhile to point out that there might exist more critical points other than $P^*$.

Now we compute the second derivative of $F \circ \mu_P$ at 0 and evaluate at $P = P^*$, i.e.,

$$\tfrac{\mathrm{d}^2}{\mathrm{d}\,\varepsilon^2} \left(F \circ \mu_P\right)(\varepsilon Z)\Big|_{\varepsilon=0, P=P^*} = Z^\top H_{11}(P^*) Z, \quad (7)$$

where $H_{11} : Gr(p,m) \to \mathbb{R}^{p\times p}$ with

$$\begin{aligned} \begin{bmatrix} H_{11} & H_{21}^\top \\ H_{21} & H_{22} \end{bmatrix} :=& 2\mathbb{E}[G''(w^\top P w)\Theta^\top w w^\top \Theta] \\ &- \mathbb{E}[G'(w^\top P w)\Theta^\top w w^\top \Theta] \\ &+ \mathbb{E}[G'(w^\top P w) w^\top P w] I_m. \end{aligned} \quad (8)$$

In the sequel, we make the reasonable assumption that the evaluation of $H_{11}$ at $P = P^*$ is invertible. By smoothness, this implies that within an open neighbourhood of $P^*$, $\mathcal{N}_{P^*} \subset Gr(p,m)$, the evaluation of $H_{11}$ at a point $P \in \mathcal{N}_{P^*}$ is invertible as well. Therefore we suggest to approximate the Hessian of $F \circ \mu_P$ at 0 for arbitrary $P \in \mathcal{N}_{P^*} \subset Gr(p,m)$ using the expression $H_{11}$ as in (8). Note that such approximate Hessian gives the true Hessian at $P^*$.

The approximate Newton direction $Z \in \mathbb{R}^{(m-p)\times p}$ can be computed by solving the following linear equation,

$$Z^\top H_{11}(P) = K_{21}(P). \quad (9)$$

Thus an approximate Newton-like method for solving ISA can be summarised as follows

| Approximate Newton-Like ISA (ANLISA) algorithm |
|---|
| Step 1: Given an initial matrix $\Theta_0 \in SO(m)$ such that $P_0 = \Theta_0 \left[\begin{smallmatrix} I_p & 0 \\ 0 & 0\end{smallmatrix}\right] \Theta_0^\top \in Gr(p,m)$. Set $i = 0$, |
| Step 2: Compute $\begin{bmatrix} K_{11} & K_{21}^\top \\ K_{21} & K_{22} \end{bmatrix} := \mathbb{E}[G'(w^\top P_i w)\Theta_i^\top w w^\top \Theta_i]$, and $\begin{bmatrix} H_{11} & H_{21}^\top \\ H_{21} & H_{22} \end{bmatrix} := 2\mathbb{E}[G''(w^\top P_i w)\Theta_i^\top w w^\top \Theta_i] - \mathbb{E}[G'(w^\top P_i w)\Theta_i^\top w w^\top \Theta_i] + \mathbb{E}[G'(w^\top P_i w) w^\top P_i w] I_m$. |
| Step 3: Compute the Newton direction $Z \in \mathbb{R}^{(m-p)\times p}$ by solving the linear matrix equation $Z^\top H_{11}(P_i) = K_{21}(P_i)$, |
| Step 4: Compute $\Theta_{i+1} = \Theta_i \exp\begin{bmatrix} 0 & -Z^\top \\ Z & 0 \end{bmatrix}$ and $P_{i+1} = \mu_{P_i}(Z)$, |
| Step 5: Set $i = i + 1$ and goto Step 2. |

Again by abusing notations, let us consider the solution $Z$ of the linear system (9) as a function of $P \in Gr(p,m)$, i.e., $Z : Gr(p,m) \to \mathbb{R}^{(m-p)\times p}$. ANLISA can be then restated as the smooth and locally well defined map

$$\eta : \mathcal{N}_{P^*} \to Gr(p,m), \quad P \mapsto \mu_P\left(Z(P)\right). \quad (10)$$

Local convergence properties of the algorithmic map $\eta$ are summarised by the following results.

**Lemma 1** *Consider ANLISA as a map $\eta$ on $Gr(p,m)$ as in* (10) *and let $P^* \in Gr(p,m)$ be a correct subspace separation point. Then $P^*$ is a fixed point of $\eta$.*

PROOF (SKETCH). Recall the result in equation (6), one gets $Z(P^*) = 0$. The lemma then follows. ■

**Theorem 1** *ANLISA considered as the map $\eta$ as in* (10) *is locally quadratically convergent to a correct subspace separation point $P^* \in Gr(p,m)$.*

PROOF (SKETCH). Tedious but direct computations show that $\eta$ is locally smooth in the neighbourhood of $P^*$ and the first derivative of the algorithmic map $\eta$

$$\mathrm{D}\,\eta(P): T_P Gr(p,m) \to T_{\eta(P)} Gr(p,m), \tag{11}$$

vanishes at a correct subspace extraction point $P^* \in Gr(p,m)$. Thus the result follows. ■

## 3. MATRIX SHIFTED FIXED POINT ISA METHODS

It has been shown that FastICA is a scalar shifted version of a simpler fixed point algorithm [10, 5]. In this section we first propose a simple fixed point ISA algorithm and then develop a matrix shifted fixed point ISA method, which obtains local quadratic convergence.

Let denote $St(p,m) := \{X \in \mathbb{R}^{m\times m} | X^\top X = I_p\}$ the Stiefel manifold. We construct a smooth and locally well defined map on $St(p,m)$ as follows

$$\rho: St(p,m) \to St(p,m), \quad X \mapsto (\Psi(X))_Q, \tag{12}$$

where $(Z)_Q$ denotes the $Q$-factor in the $QR$-factorisation $Z = (Z)_Q(Z)_R$ and

$$\Psi: St(p,m) \to \mathbb{R}^{m\times p},\ P \mapsto \mathbb{E}[G'(w^\top X X^\top w) w w^\top] X. \tag{13}$$

It can be shown that for any $\Omega \in SO(p)$,

$$\Psi(X\Omega) = \Psi(X)\Omega \ \Rightarrow\ \operatorname{span}\Psi(X\Omega) = \operatorname{span}\Psi(X). \tag{14}$$

From this we deduce that the map $\rho$ as in (12) is invariant under basis changes.

Recall the representation $P = \Theta \left[\begin{smallmatrix} I_p & 0 \\ 0 & 0 \end{smallmatrix}\right] \Theta^\top \in Gr(p,m)$ with $\Theta \in SO(m)$, and $X = \Theta \left[\begin{smallmatrix} I_p \\ 0 \end{smallmatrix}\right]$. The map $\rho$ can be reconsidered as a smooth and locally well defined map on $Gr(p,m)$

$$\begin{aligned} \widetilde{\rho}: \mathcal{N}_{P^*} &\to Gr(p,m), \\ P &\mapsto \left(\widetilde{\Psi}(P)\right)_Q \left(\widetilde{\Psi}(P)\right)_Q^\top, \end{aligned} \tag{15}$$

where

$$\widetilde{\Psi}: Gr(p,m) \to \mathbb{R}^{m\times p},\ P \mapsto \mathbb{E}[G'(w^\top P w) w w^\top] X. \tag{16}$$

It can be shown easily that a correct subspace extraction point $P^*$ is a fixed point of the algorithmic map $\widetilde{\rho}$.

Now we generalise the algorithmic map (15) by means of a matrix shift, i.e.,

$$\begin{aligned} \zeta: \mathcal{N}_{P^*} &\to Gr(p,m), \\ P &\mapsto \left(\widetilde{\Psi}(P) - X\Phi(P)\right)_Q \left(\widetilde{\Psi}(P) - X\Phi(P)\right)_Q^\top, \end{aligned} \tag{17}$$

where $\Phi : Gr(p,m) \to GL(p)$ is such that, for any $\Omega \in SO(p)$

$$\Phi(X\Omega) = \Omega^\top \Phi(X)\Omega. \tag{18}$$

Straightforwardly, one comes up with the following result

**Lemma 2** *Let $P^* \in Gr(p,m)$ be a correct subspace separation point. Then $P^*$ is a fixed point of the map $\zeta$ as in* (17)*.*

Using the same strategy as developing FastICA as a scalar shifted fixed point method [5], we propose the following shift

$$\Phi_q(P) := 2\mathbb{E}[G''(w^\top P w) X^\top w w^\top X] + \mathbb{E}[G'(w^\top P w)] I_p. \tag{19}$$

A matrix shifted fixed point ISA (MS-ISA) algorithm can then be summarised as follows

| Matrix shifted Fixed Point ISA (MS-ISA) algorithm |
|---|
| Step 1: Given an initial matrix $\Theta_0 \in SO(m)$ such that $X_0 = \Theta_0 \left[\begin{smallmatrix} I_p \\ 0 \end{smallmatrix}\right]$ and $P_0 = X_0 X_0^\top \in Gr(p,m)$. Set $i = 0$, |
| Step 2: Compute $\Psi(P_i) = \mathbb{E}[G'(w^\top P_i w) w w^\top] X_i$, and $\Phi(P_i) = 2\mathbb{E}[G''(w^\top P_i w) X_i^\top w w^\top X_i] + \mathbb{E}[G'(w^\top P_i w)] I_p$ |
| Step 3: Compute $X_{i+1} = (\Psi(P_i) - X_i \Phi(P_i))_Q$ and $P_{i+1} = X_{i+1} X_{i+1}^\top$, |
| Step 4: Set $i = i + 1$ and goto Step 2. |

Local convergence properties of MS-ISA is studied in the following theorem.

**Theorem 2** *MS-ISA considered as the map $\zeta$ as in* (17) *is locally quadratically convergent to a correct subspace separation point $P^* \in Gr(p,m)$.*

Due to the space limit of this paper, the proof will be given in our forthcoming paper. Nevertheless, such result will be verified by numerical evidence in Sec. 4.

### Further Generalisations

Finally it can be easily seen that, if one only takes the diagonal of the expression $\Phi_q(X)$ as in (19) as a new matrix shift, i.e., $\Phi_d(X) = \operatorname{diag}\Phi_q(X)$, it gives exactly the formulation of the original FastISA, referred here as MS-ISA-D. Such diagonalisation strategy is similar to the situation when generalising Rayleigh Quotient Iteration to $St(p,m)$ or $Gr(p,m)$ see [11, 6] and references therein. Likewise, by taking the diagonal of the expression $H_{11}$ as in (8) as a new approximate Hessian, one can easily formulate a new approximate Newton-like method as well, referred here as ANLISA-D. By the same arguments as above, it can be shown that a correct subspace extraction point $P^* \in Gr(p,m)$ is a fixed point of ANLISA-D and MS-ISA-D. Both algorithms are locally linearly convergent to $P^*$.

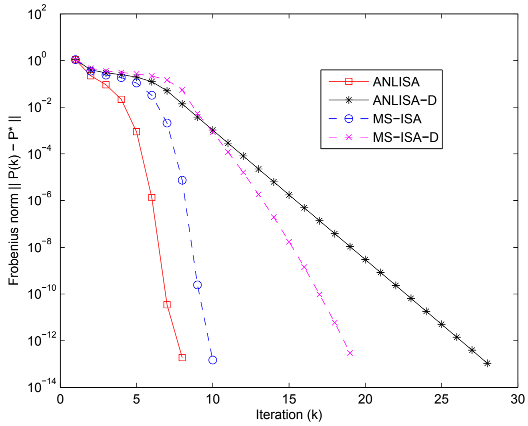


**Fig. 1**. Local convergence properties of ANLISA, ANLISA-D, MS-ISA and MS-ISA-D.

### 4. NUMERICAL EXPERIMENTS

In this section, we investigate performances of the proposed algorithms by numerical experiments, focusing on convergence properties. An ideal dataset is generated to ensure that the approximation (8) gives the true Hessian at a critical point of $F$ corresponding to a correct extraction. Here we specify the non-linear function $G$ as follows

$$G : \mathbb{R} \to \mathbb{R}, \quad G(x) := \log(\cosh(x)). \tag{20}$$

The convergences are measured by the Frobenius norm of the difference between the accumulation point $P^* \in Gr(p, m)$ and the current iterate $P_k \in Gr(p, m)$, i.e., by $\|P_k - P^*\|_F$, with $P^*$ being the computed subspace extraction point.

We need to point out that due to the nature of source components and the choice of $G$, e.g., as in (20), there might exist many local optima of $F$. Thus in this experiment, all algorithms are initialised with the same point, which is close enough to a correct subspace extraction. Numerical results in Fig. 1 show that both ANLISA and MS-ISA are locally quadratically convergent to a correct subspace extraction point $P^* \in Gr(p, m)$. While their diagonal counterparts, ANLISA-D and MS-ISA-D, seem to converge only linearly. It is also worthwhile to point out that, (i) ANLISA and ANLISA-D are of similar computational burden as MS-ISA and MS-ISA-D, respectively; (ii) although ANLISA and MS-ISA require less iterations than their diagonal counterparts, both are a bit more computationally expensive.

## Acknowledgements

This work was partially done during the first author's visits to the Neuroinformatics Group, Helsinki Institute of Information Technology and University of Helsinki, Finland and the Max Planck Institute for Biological Cybernetics, Tübingen, Germany. He wishes to thank both institutions for the generous hospitality, support and invitation. He would also like to thank the PASCAL Network of Excellence for financial support.

### 5. REFERENCES

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

[2] A. Hyvärinen and P. O. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, vol. 12, no. 7, pp. 1705–1720, 2000.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.

[4] A. Hyvärinen and U. Köster, "FastISA: A fast fixed-point algorithm for independent subspace analysis," in *Proceeding of the* $14^{th}$ *European Symposium on Artificial Neural Networks (ESANN 2006)*, Bruges, Belgium, 2006.

[5] H. Shen, K. Hüper, and A.-K. Seghouane, "Geometric optimisation and FastICA algorithms," in *Proceedings of the* $17^{th}$ *International Symposium of Mathematical Theory of Networks and Systems (MTNS 2006)*, Kyoto, Japan, 2006, pp. 1412–1418.

[6] P.-A. Absil, *Invariant Subspace Computation: A Geometric Approach*, Ph.D. thesis, University of Liege, Belgium, 2003.

[7] K. Hüper and J. Trumpf, "Newton-like methods for numerical optimisation on manifolds," in *Proceedings of Thirty-eighth Asilomar Conference on Signals, Systems and Computers*, 2004, pp. 136–139.

[8] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, London, 1994.

[9] U. Helmke, K. Hüper, and J. Trumpf, "Newton's method on Graßmann manifolds," 2007, to be published.

[10] P. Regalia and E. Kofidis, "Monotonic convergence of fixed-point algorithms for ICA," *IEEE Transactions on Neural Networks*, vol. 14, no. 4, pp. 943–949, 2003.

[11] K. Hüper, "A dynamical system approach to matrix eigenvalue algorithms," in *Mathematical Systems Theory in Biology, Communications, Computation*, J. Rosenthal and D. S. Gilliam, Eds., vol. 134 of *The IMA Volumes in Mathematics and its Applications*, pp. 257–274. Springer, New York, 2003.