Natural Conjugate Gradient on Complex Flag Manifolds for Complex Independent Subspace Analysis

Yasunori Nishimori[1], Shotaro Akaho[1], and Mark D. Plumbley[2]

[1] Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST),
AIST Central2, 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan
`y.nishimori@aist.go.jp, s.akaho@aist.go.jp`
[2] Department of Electronic Engineering, Queen Mary University of London,
Mile End Road, London E1 4NS, UK
`mark.plumbley@elec.qmul.ac.uk`

Abstract. We study the problem of complex-valued independent subspace analysis (ISA). We introduce complex flag manifolds to tackle this problem, and, based on Riemannian geometry, propose the natural conjugate gradient method on this class of manifolds. Numerical experiments demonstrate that the natural conjugate gradient method yields better convergence compared to the natural gradient geodesic search method.

1 Introduction

The main aim of this paper is to explain how a new class of complex manifolds, *complex flag manifolds* can be used to solve the complex independent subspace analysis (complex ISA) task. We investigate two geodesic search-type optimization methods on complex flag manifolds. One is the natural gradient geodesic search method (NGS), and the other is the natural conjugate gradient method (NCG). The former method was proposed on the orthogonal group to solve the non-negative ICA task by Plumbley [15], and the latter was proposed by Edelman et al. [5] for the case of real Stiefel and Grassmann manifolds. Although several authors have recently proposed optimization methods on complex manifolds such as the unitary group [1],[6], complex Stiefel and Grassmann manifolds [10], complex flag manifolds have not been explored so far in the signal processing literature. Also, most authors have concentrated on either the natural gradient or the Newton's method on complex manifolds, however, the behavior of the conjugate gradient method over complex manifolds has not been reported.

The organization of this paper is as follows. In section 2 the problem of complex-valued independent subspace analysis is formulated, and how complex flag manifolds naturally arise for solving this problem is explained; subsection 2.1 reviews the Riemannian geodesic line search method as well as the Riemannian conjugate gradient method, subsection 2.2 derives the update rules for both methods on complex flag manifolds. Section 3 illustrates the comparison

of the performance of both methods as applied to a complex ISA problem on synthetic data sets.

2   Complex Independent Subspace Analysis

Complex ISA is a straightforward generalization of real ISA [8]. We observe a linear mixture of source signals $s \in \mathbf{C}^N$:

$$x = As, \text{where } A \in \mathbf{C}^{N \times N} \text{ is nondegenerate.} \tag{1}$$

We assume that s is decomposed into d_j-tuples $(j = 1, \ldots, r+1)$ as

$$s = (s_1, \ldots, s_{d_1}, s_{d_1+1}, \ldots, s_{d_1+d_2}, \ldots, s_n, s_{n+1}, \ldots, s_N)^\top, \quad \sum_{j=1}^{r+1} d_j = N \tag{2}$$

such that signals belonging to different tuples are statistically independent, while signals belonging to the same tuple are allowed to be statistically dependent, and $\top$ denotes the (real) transpose operator. Further, the $(r+1)$st tuple is assumed to be composed of Gaussian noises and $d_{r+1} = N - n$. Then we define the task of complex ISA as estimating n original sources $s_1, \ldots, s_n$, not including $(N-n)$ Gaussian noises $s_{n+1}, \ldots. s_N$ from this statistical assumption (up to ambiguity). Often we deal with the noiseless case $N = n$.

To solve the complex ISA problem, as with ordinary real ICA, we first center and pre-whiten the data:

$$z = Bx, \text{ such that } E_z \left[zz^H \right] = I_N, \tag{3}$$

where H denotes the Hermitian transpose operator. Then solving the complex ISA task reduces to finding a rectangular unitary matrix

$$W = [W_1, W_2, \ldots, W_r] \in \mathbf{C}^{N \times n}, \text{ where } W^H W = I_n, W_j \in \mathbf{C}^{N \times d_j} \tag{4}$$

such that the output vector

$$y = W^H z = W^H BAs \tag{5}$$

satisfies the statistical assumption of the original sources. Therefore the complex ISA task can be solved by optimization on the complex Stiefel manifold:

$$\mathrm{St}(N, n\,;\mathbf{C}) = \left\{ W \in \mathbf{C}^{N \times n} | W^H W = I_n \right\}. \tag{6}$$

Note that the complex Stiefel manifold $\mathrm{St}(N, n\,;\mathbf{C})$ includes the unitary group

$$U(N) = \left\{ W \in \mathbf{C}^{N \times N} | W^H W = I_N \right\} \tag{7}$$

as a special case when $N = n$.

For a cost function, we use

$$f(W) = \sum_{j=1}^{r} E_z \left[\log p(\sum_{l=1}^{d_j} |w_{jl}^H z|^2) \right], \tag{8}$$

where E_z denotes the empirical mean with respect to z, and p is a probability density function of the power of the projection of the observed signals to the subspace spanned by w_{jl}'s ($W_j = w_{j1}, \ldots, w_{jd_j}$), that generalizes the maximum likelihood-based parametric cost function for real ISA proposed by Hyvärinen and Hoyer [8]. However, this cost function is invariant to the unitary transformations of each W_j ($j = 1, \ldots, r$) :

$$W \mapsto W \operatorname{diag}[U_1, U_2, \ldots, U_r], \; U_j \in U(d_j),$$

so an additional structure is present in the target manifold over which a solution matrix is searched. The target manifold is obtained by identifying those points on $\operatorname{St}(N, n \,; \mathbf{C})$ related by the following equivalence relation $\sim$

$$W_1 \sim W_2 \iff \exists U_i \in U(d_i) \text{ s.t. } W_2 = W_1 \operatorname{diag}[U_1, U_2, \ldots, U_r]. \tag{9}$$

This quotient space is called the *complex flag manifold*, and we denote it as $\operatorname{Fl}(n, d_1, \ldots, d_r \,; \mathbf{C})$. $\operatorname{Fl}(n, d_1, \ldots, d_r \,; \mathbf{C})$ is isomorphic to $U(N)/U(d_1) \times \cdots \times U(d_r) \times U(N - n)$ as a homogeneous space, and includes the complex Grassmann manifold as a special case when $r = 1$. We will exploit this geometric structure for the optimization of the complex ISA cost function. For instance, the Riemannian Hessian of the ISA cost function on the tangent space of simpler complex Stiefel manifold is degenerate, therefore we cannot directly apply the Riemannian Newton's method on the Stiefel manifold to ISA.

2.1   Geodesic Search Method

The update rules of both geodesic search type methods for minimizing a real-valued smooth cost function f over a smooth manifold M go as follows. We denote the equation of the geodesic emanating from a point $W \in M$ in the direction of V as $\varphi_M(W, V, t)$, where t is a time parameter s.t.

$$\varphi_M(W, V, 0) = W, \quad \frac{d}{dt}\varphi_M(W, V, 0)|_{t=0} = V \tag{10}$$

hold. Although a geodesic is a generalization of a straight line to a manifold and is computed by a Riemannian metric g on the manifold, we use one metric for each manifold in this paper, and do not refer which metric to use with this notation. The formula [12]

$$\varphi_{\operatorname{St}(N,n)}(W, V, t) = \exp(t(DW^\top - WD^\top))W, \text{ where } D = (I - \frac{1}{2}WW^\top)V \tag{11}$$

is used in the subsequent part. We extend both methods such that they are applicable to complex flag manifolds. At each iteration, both methods search

for a point where the cost function takes its minimum along the geodesic in the direction of each search direction as follows. The search direction for NGS is the natural gradient of the cost function, while the NCG search direction is determined by using the parallel transportation of the one-step former search direction to the current position as well as the natural gradient of the cost function.

– the natural gradient geodesic search method (NGS)

$$t^k_{\min} = \arg \min_t f(\varphi_M(W_k, G_k, t)) \tag{12}$$

$$G_k = - \operatorname{grad}^M_{W_k} f(W_k) \tag{13}$$

$$W_{k+1} = \varphi_M(W_k, G_k, t^k_{\min}) \tag{14}$$

– natural conjugate gradient method (NCG)

$$H_0 = - \operatorname{grad}^M_{W_0} f(W_0), \tag{15}$$

$$H_k = - \operatorname{grad}^M_{W_k} f(W_k) + \gamma_k \tau H_{k-1} \tag{16}$$

$$t^k_{\min} = \arg \min_t f(\varphi_M(W_k, H_k, t)) \tag{17}$$

$$W_{k+1} = \varphi_M(W_k, H_k, t^k_{\min}) \tag{18}$$

Here τH_{k-1} denotes the parallel transportation of H_k along $\varphi_M(W_{k-1}, H_{k-1}, t)$ to W_k, and[1]

$$\gamma_k = \frac{g_{W_k}(\operatorname{grad}^M_{W_k} f(W_k) - \operatorname{grad}^M_{W_{k-1}} f(W_{k-1}), \operatorname{grad}^M_{W_k} f(W_k))}{g_{W_{k-1}}(\operatorname{grad}^M_{W_{k-1}} f(W_{k-1}), \operatorname{grad}^M_{W_{k-1}} f(W_{k-1}))}. \tag{19}$$

Since $\varphi_{\operatorname{St}(N,n)}(W_k, H_k, t)$ is a geodesic, the parallel transportation vector τH_k is equal to the velocity vector

$$\frac{d}{dt} \varphi_{\operatorname{St}(N,n)}(W_k, H_k, t) = (D_k W_k^\top - W_k^\top D_k) \exp(t(D_k W_k^\top - W_k^\top D_k)) \tag{20}$$

at $t = t^k_{\min}$. Therefore

$$\tau H_k = (D_k W_k^\top - W_k^\top D_k) W_{k+1}, \tag{21}$$

where $D_k = (I - \frac{1}{2} W_k W_k^\top) H_k$.

2.2   Geometry of Complex Flag Manifolds

We first consider an optimization problem over a complex Stiefel manifold:

$$\operatorname{St}(N, n \,; \mathbb{C}) \ni W = W^{\Re} + iW^{\Im} \mapsto F(W) \in \mathbb{R}. \tag{22}$$

---

[1] There are several versions of selecting this γ_k for the conjugate gradient method. Here we use the Fletcher-Reeves rule. More precisely, $\tau \operatorname{grad}_{W_{k-1}} f(W_{k-1})$ should be used instead of $\operatorname{grad}_{W_{k-1}} f(W_{k-1})$ in the numerator, however, it is not known how to compute it cheaply, so we use this approximated version as is used in [5].
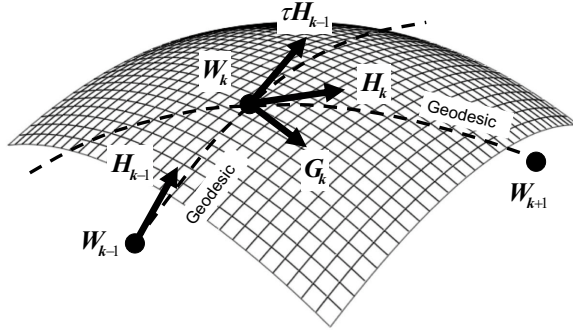
Fig. 1. Natural conjugate gradient update

Since the cost function F is real-valued, we regard $\mathrm{St}(N, n\,;\mathbb{C})$ as a real manifold; taking the real and imaginary parts of the defining equation

$$W^H W = (W^{\Re\top} - iW^{\Im\top})(W^\Re + iW^\Im) = I_n \tag{23}$$

separately, we get a real manifold M underlying $\mathrm{St}(N, n\,;\mathbb{C})$, which is a submanifold in $\mathbb{R}^{2N \times n}$ defined by the constraints:

$$M := \{W' = \begin{pmatrix} W^\Re \\ W^\Im \end{pmatrix} \in \mathbb{R}^{2N \times n} |$$
$$W^{\Re\top} W^\Re + W^{\Im\top} W^\Im = I_n, W^{\Re\top} W^\Im - W^{\Im\top} W^\Re = O_n\}. \tag{24}$$

We further embed M by the following map:

$$v : W' \mapsto \tilde{W} = \begin{pmatrix} W^\Re & -W^\Im \\ W^\Im & W^\Re \end{pmatrix}. \tag{25}$$

From the relation (24), $\tilde{W}^\top \tilde{W} = I_{2n}$ holds. In fact, this can be elaborated; $\tilde{M} = v(M)$ is a *totally geodesic*, or *autoparallel* submanifold of $\mathrm{St}(2N, 2n\,;\mathbf{R})$ relative to the normal metric. That is, every geodesic on $\tilde{M}$ becomes a geodesic on $\mathrm{St}(2N, 2n\,;\mathbf{R})$ as well, as follows:

$$\begin{array}{ccccc} \mathrm{St}(N, n\,;\mathbf{C}) \to & M & \xrightarrow{v} \tilde{M} \to & \mathrm{St}(2N, 2n\,;\mathbf{R}) \\ W & \mapsto W' \mapsto & \tilde{W} \end{array} \tag{26}$$

Thus, minimizing F over $\mathrm{St}(N, n\,;\mathbb{C})$ can be solved through the minimization of $f(\tilde{W}) := F(W)$ on $\tilde{M}$. Though we start from this real manifold $\tilde{M}$ for the optimization, the update rule can be expressed by the original complex variables using the following nice properties the map $u : W \mapsto \tilde{W}$ possesses:

1. $u(I_m) = I_{2m}$
2. $u(W_1 W_2) = u(W_1)u(W_2)$, where $W_1 \in \mathbf{C}^{l \times m}, W_2 \in \mathbf{C}^{m \times n}$
3. $u(W_1 + W_2) = u(W_1) + u(W_2)$, where $W_1, W_2 \in \mathbf{C}^{m \times n}$

4. $u(W^H) = u(W)^\top$, where $W \in \mathbf{C}^{m \times m}$
5. $u(\exp(W)) = \exp(u(W))$, where $W \in \mathbf{C}^{m \times m}$

For instance, the geodesic

$$\varphi_{\mathrm{St}(2N,2n;\mathbf{R})}(\tilde{W}, \tilde{V}, t) = \exp(t(\tilde{D}\tilde{W}^\top - \tilde{W}\tilde{D}^\top))\tilde{W}, \ \tilde{D} = (I - \frac{1}{2}\tilde{W}\tilde{W}^\top)\tilde{V}, \quad (27)$$

on $\tilde{M}$ is readily expressed in the complex form by switching the real transpose operator $\top$ to the Hermitian transpose operator H :

$$\varphi_{\mathrm{St}(N,2;\mathbf{C})}(W, V, t) = \exp(t(DW^H - WD^H))W, \quad (28)$$

where $D = (I - \frac{1}{2}WW^H)V, V = u^{-1}\tilde{V}$.

Next we derive another ingredient necessary for the optimization: the natural gradient. Since $\tilde{M}$ is a submanifold of $\mathrm{St}(2N, 2n\,; \mathbf{R})$, the natural gradient of f on $\tilde{M}$ is obtained by projecting the natural gradient of f on $\mathrm{St}(2N, 2n;\mathbf{R})$ at $\tilde{W}$ to the tangent space of $\tilde{M}$ at $\tilde{W}$. It is easily verified that the natural gradient of f on $\mathrm{St}(2N, 2n\,; \mathbf{R})$ at $\tilde{W}$ always lies in the tangent space of $\tilde{M}$ at $\tilde{W}$, hence we get

$$\mathrm{grad}_{\tilde{W}}^{\tilde{M}} f(\tilde{W}) = F(\tilde{W}) = \tilde{\nabla}F - \tilde{W}^\top \tilde{\nabla}F\tilde{W}^\top, \text{where} \tilde{\nabla}F = \begin{pmatrix} \frac{\partial F}{\partial W^{\Re}} & -\frac{\partial F}{\partial W^{\Im}} \\ \frac{\partial F}{\partial W^{\Im}} & \frac{\partial F}{\partial W^{\Re}} \end{pmatrix}. \quad (29)$$

$\tilde{\nabla}F$ comes from the Euclidean gradient of F on M;

$$\tilde{\nabla}F = v\left(\left(\frac{\partial F}{\partial W^{\Re}}, \frac{\partial F}{\partial W^{\Im}}\right)^\top\right), \quad (30)$$

and the corresponding complex gradient becomes[2]

$$2\frac{\partial F}{\partial \overline{W}} := \frac{\partial F}{\partial W^{\Re}} + i\frac{\partial F}{\partial W^{\Im}}. \quad (31)$$

This complex gradient has been used for the optimization on complex manifolds by many authors including [1], [2], [4], [6], [9], [10], [17]. Finally, the following formula for the complex form of the natural gradient is obtained by replacing

$$\tilde{\nabla}F \longmapsto 2\frac{\partial F}{\partial \overline{W}}, \qquad \text{and} \qquad \top \longmapsto H; \quad (32)$$

$$\mathrm{grad}_W^{\mathrm{St}(N,n;\mathbf{C})} F = 2\left(\frac{\partial F}{\partial \overline{W}} - W\frac{\partial F}{\partial \overline{W}}^H W\right). \quad (33)$$

---

[2] This notation using $\overline{W}$(the complex conjugate of W) is due to the Wirtinger calculas [9].

In the same manner as for the complex Stiefel manifold, we can take the complex flag manifold $\mathrm{Fl}(n, d_1, \ldots, d_r\,;\mathbb{C})$ as a totally geodesic submanifold by using the following embedding

$$v : W = [W_1, W_2, \ldots, W_r] \mapsto \tilde{W}$$

$$= \begin{pmatrix} W_1^{\Re} & -W_1^{\Im} & W_2^{\Re} & -W_2^{\Im} & \cdots & W_r^{\Re} & -W_r^{\Im} \\ W_1^{\Im} & W_1^{\Re} & W_2^{\Im} & W_2^{\Re} & \cdots & W_r^{\Im} & W_r^{\Re} \end{pmatrix}, \quad (34)$$

where $W_k = W_k^{\Re} + iW_k^{\Im} \in \mathbf{C}^{N \times d_k}$, $\sum_{k=1}^{r} d_r = n$, of $\mathrm{Fl}(n, 2d_1, \ldots, 2d_r\,;\mathbb{R})$. Thus, the formula for the natural gradient and geodesics of real flag manifolds we obtained in [14] can also be complexified by replacing the real Euclidean gradient by the complex gradient and the transpose operator by the Hermitian transpose operator;

$$\mathrm{grad}_W^{\mathrm{Fl}(n,d_1,\ldots,d_r;\mathbb{C})} F = \left[V_1^{\mathbf{C}}, V_2^{\mathbf{C}}, \ldots, V_r^{\mathbf{C}} \right], \quad (35)$$

$$V_j^{\mathbf{C}} = X_j^{\mathbf{C}} - (W_j W_j^{H} X_j^{\mathbf{C}} + \sum_{l \neq j} W_l X_l^{\mathbf{C}^H} W_j), \quad (36)$$

where $X_j^{\mathbf{C}} = \frac{\partial F}{\partial W_j^{\Re}} + i\frac{\partial F}{\partial W_j^{\Im}}$.

$$\varphi_{\mathrm{Fl}(n,d_1,\ldots,d_r;\mathbb{C})}(W, V, t) = \exp(t(DW^H - WD^H))W, \quad (37)$$

where $D = (I - \frac{1}{2}WW^H)V$.

3   Numerical Experiments

We applied the Riemannian conjugate gradient method to the complex-valued ISA problem using the following synthetic data sets and compared its performance with that of the natural gradient geodesic search method. Let $s = (s_1, s_2, \ldots, s_8)^\top = (\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4)^\top \in \mathbb{C}^8$ be a source signals, where $\mathbf{s}_i \in \mathbb{C}^2$ and $\mathbf{s}_1, \mathbf{s}_2$ are sampled from a generalization of PSK8 signal, $\mathbf{s}_3$, sampled from a generalization of QAM(4) signal, and $\mathbf{s}_4$, sampled from a two dimensional complex Gaussian noise.[3] $\mathbf{s}_1, \mathbf{s}_2$ take either of the following 24 values; $(\pm 1, 0), (0, \pm 1), (\pm i, 0), (0, \pm i), (\pm 1 \pm i, \pm 1 \pm i)/2$ independently at a time. The codomain of a PSK8 signal lies on a circle equiangularly in the complex plane $\simeq \mathbb{R}^2$, while that of this generalized signal lies on a hypersphere in $\mathbb{C}^2 \simeq \mathbb{R}^4$ equiangularly. QAM(4) signal takes either of 4 values $(\pm 1 \pm i)$ independently, while $\mathbf{s}_3 \in \mathbb{C}^2$ takes either of the following 16 values $(\pm 1 \pm i, \pm 1 \pm i)$ independently. We observe an instantaneous linear mixture of the signals $x = As$, where A is a regular 8×8 complex matrix. After whitening $z = Bx$, the complex ISA task in the sense of this paper can be solved by optimization of a cost function over the complex flag manifold $\mathrm{Fl}(8, 2, 2, 2;\mathbb{C})$ such that $y = W^H z$ give recovered signals, where $W \in \mathrm{Fl}(8, 2, 2, 2\,;\mathbb{C})$. We use the following Kurtosis-like higher order statistics for a cost function.
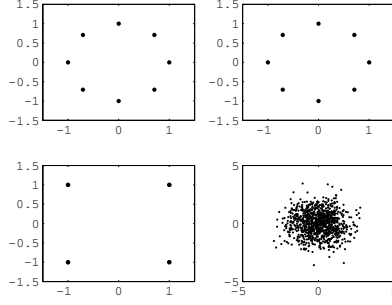
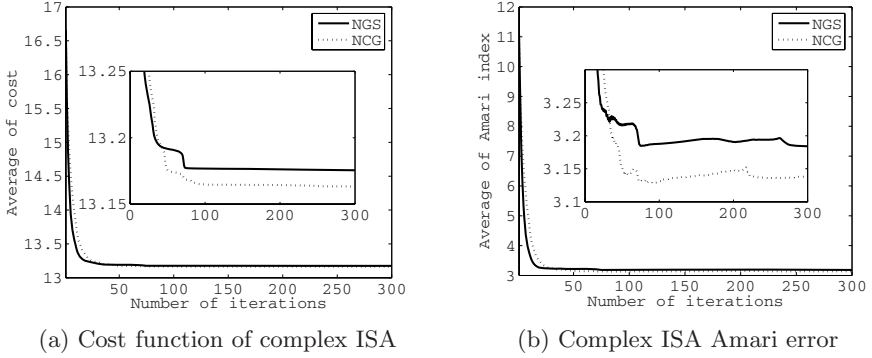**Fig. 2.** Source signals of complex ISA (schematic)

(a) Cost function of complex ISA   (b) Complex ISA Amari error

Fig. 3. Numelical simulation results

$$f(W) = \sum_{i=1}^{3} E_z(||\mathbf{y}_i||^4), \tag{38}$$

where $y = (y_1, y_2, \ldots, y_6)^\top = (\mathbf{y}_1^\top, \mathbf{y}_2^\top, \mathbf{y}_3^\top)^\top, \mathbf{y}_i \in \mathbb{C}^2$. Thus we get the natural gradient of f by substituting

$$\frac{\partial f}{\partial W_i^\Re} = E_z(||\mathbf{y}_i||^2 (z\mathbf{y}_i^H + \bar{z}\mathbf{y}_i^\top)) \tag{39}$$

$$\frac{\partial f}{\partial W_i^\Im} = E_z(||\mathbf{y}_i||^2 (z\mathbf{y}_i^H - \bar{z}\mathbf{y}_i^\top)) \tag{40}$$

into (35). Also, to evaluate the performance, the following generalized Amari error was used. Let us put $C = W^H B A \in \mathbb{C}^{6\times8}$, decompose it into 3×4 block matrices $D_{ij} \in \mathbb{C}^2 (i = 1, \ldots, 3, j = 1, \ldots, 4)$, and make a new 4×4 matrix $\tilde{D}$ such that $\tilde{D}_{ij} = \max_{||x||=1} ||D_{ij}x||, (i = 1, \ldots, 3, j = 1, \ldots, 4), \tilde{D}_{4,4} = 1, \tilde{D}_{i,j} = 0,$

---

[3] Note that the data sets s_i actually take value in $\mathbb{C}^2$ instead of $\mathbb{C}$. Simplified versions are drawn in Figure 2 for visualization.

otherwise. The Amari error applied to this matrix yields an appropriate quantity to measure how close C is block diagonal up to scaling and permutation.

$$E(\tilde{D}) = \sum_{i=1}^{4} \sum_{j=1}^{4} \frac{\tilde{D}_{ij}}{\max_k \tilde{D}_{ik}} + \frac{\tilde{D}_{ij}}{\max_k \tilde{D}_{kj}} - 2 \times 4 \qquad (41)$$

Figure 3 show average behaviors of two algorithms over 100 trials. Note that a zoomed figure around the convergence point is superimposed for each of the figures. We observed that the natural gradient geodesic search (NGS) decreased the cost function more quickly at early stages of learning, yet the natural conjugate gradient method (NCG) exhibited faster and better convergence around the convergence points. One possible reason for this difference is because, around the convergence point, the cost function can be well approximated by a quadratic function. This assumption is exploited to derive the conjugate gradient method. Similar observation in favour of the natural conjugate gradient on a manifold of probability distributions was recently reported by Honkela et al.[7]. Also the effectiveness of the conjugate gradient method in real ICA was shown by Martin-Clemente et al.[11].

Lastly we would like to mention that the problem of complex ISA was also discussed by Szabo et al.[16]. Their proposed learning algorithm consists of 2 steps: first seek an ordinary complex ICA solution, then grouping dependent components by swapping two basis vectors belonging to different subspaces until the swapping does not lower the cost function. Here we didn't compare their algorithm with ours because we consider the case when the effect of the local minima of the cost function is not so serious. When the effect is serious, we also needed to swap basis vectors to get out of the local minima in [13]. From a computational complexity point of view, the optimization of a parametric ICA cost function is similar to the optimization of a parametric ISA cost function; the second greedy search step in their algorithm costs more expensive.

4   Conclusions

We introduced complex flag manifolds to tackle complex version of independent subspace analysis and proposed the Riemannian conjugate gradient method for the optimization on this class of complex manifolds. The Riemannian conjugate gradient method outperformed simpler natural gradient geodesic search. This Riemannian method is applicable to other signal processing problems that are formalized as optimization problems on complex flag manifolds.

Acknowledgements

References

1. Abrudan, T., Eriksson, J., Koivunen, V.: Steepest Descent Algorithms for Optimization under Unitary Matrix Constraint (preprint)
2. Adali, T., Li, H.: A practical formulation for computation of complex gradients and its application to maximum likelihood. In: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Honolulu, Hawaii (April 2007)
3. Amari, S.: Natural gradient works efficiently in learning. Neural Computation 10, 251–276 (1998)
4. Brandwood, D.H.: A Complex Gradient Operator and its Application in Adaptive Array Theory. IEE Proceedings H (Microwaves, Optics, and Antennas) 130(1), 11–16 (1983)
5. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM Journal on Matrix Analysis and Applications 20(2), 303–353 (1998)
6. Fiori, S.: A Study on Neural Learning on Manifold Foliations: The Case of the Lie Group SU(3). Neural Computation 4, 1091–1197 (2008)
7. Honkela, A., Tornio, M., Raiko, T., Karhunen, J.: Natural Conjugate Gradient in Variational Inference. In: Ishikawa, M., Doya, K., Miyamoto, H., Yamakawa, T. (eds.) ICONIP 2007, Part II. LNCS, vol. 4985. Springer, Heidelberg (to appear, 2008)
8. Hyvärinen, A., Hoyer, P.O.: Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. Neural Computation 12(7), 1705–1720 (2000)
9. Kreutz-Delgado, K.: The Complex Gradient Operator and the CR -Calculus, Technical report, University of California, San Diego (2003)
10. Manton, J.: Optimization algorithms exploiting unitary constraints. IEEE Transactions on Signal Processing 50(3), 635–650 (2002)
11. Clemente, R.M., Puntonet, C.G., Catalina, J.I.A.: Blind Signal Separation Based on the Derivatives of the Output Cumulants and a Conjugate Gradient Algorithm. In: Proceedings of Int. Conf. Independent Component Analysis Signal Separation, San Diego, pp. 390–393 (2001)
12. Nishimori, Y., Akaho, S.: Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. Neurocomputing 67, 106–135 (2005)
13. Nishimori, Y., Akaho, S., Plumbley, M.D.: Riemannian optimization method on the flag manifold for independent subspace analysis. In: Rosca, J.P., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 295–302. Springer, Heidelberg (2006)
14. Nishimori, Y., Akaho, S., Abdallah, S., Plumbley, M.D.: Flag manifolds for subspace ICA problems. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007), Honolulu, Hawaii, USA, April 15-20, vol. IV, pp. 1417–1420 (2007)
15. Plumbley, M.D.: Algorithms for non-negative independent component analysis. IEEE Transactions on Neural Networks 14(3), 534–543 (2003)
16. Szabo, Z., Lorincz, A.: Real and Complex Independent Subspace Analysis by Generalized Variance, arXiv:math/0610438v1
17. van den Bos, A.: Complex Gradient and Hessian. IEE Proc.-Vis. Image Signal Processing 141(6), 380–382 (1994)