Abstract. In this paper we introduce a new class of manifolds, generalized flag manifolds, for the complex and subspace ICA problems. A generalized flag manifold is a manifold consisting of subspaces which are orthogonal to each other. The class of generalized flag manifolds include the class of Grassmann manifolds. We extend the Riemannian optimization method to include this new class of manifolds by deriving the formulas for the natural gradient and geodesics on these manifolds. We show how the complex and subspace ICA problems can be solved by optimization of cost functions on a generalized flag manifold. Computer simulations demonstrate our algorithm gives good performance compared with the ordinary gradient descent method.

Key Words: Independent subspace analysis, Complex ICA, Natural gradient, Geodesics, Generalized flag manifolds, Riemannian optimization.

INTRODUCTION

Many neural networks and signal processing tasks, including independent component analysis (ICA), involve optimization of a cost function over matrices subject to some constraints, such as orthonormality. This type of problem can be tackled by optimization over manifolds, and we often deal with manifolds related to the orthogonal group $O(n)$, such as the Stiefel and Grassmann manifolds.

Some Euclidean optimization methods, such as steepest gradient descent, can be used for optimization over manifolds, but they need to be properly modified to do so. Firstly, the Euclidean gradient depends on the way the manifold is parametrized, which can lead to different ‘steepest’ directions. We therefore introduce a Riemannian metric on the manifold itself: the steepest direction with respect to this metric is called the Riemannian gradient vector, also known as the natural gradient in the neural networks community [2]. Secondly, because a manifold is ‘curved’, the usual ‘add’ update step used in the Euclidean space does not keep the current point constrained on the manifold. To overcome this, we instead ensure our updates follow geodesics on the manifold. A geodesic joining two nearby points on a manifold is the shortest path between those points. It is determined by the Riemannian metric, and is a generalization of the Euclidean concept of a straight line.

Putting these ideas together, the Riemannian optimization method operates as follows
that St is the isotropy subgroup of the Riemannian gradient of orthogonal rectangular matrices. The subgroup of the Lie group of orthogonal matrices computing $[1, 3, 7, 9]$. The most fundamental is the orthogonal group investigated in neural networks, signal processing, numerical analysis, and scientific analysis-type problems such as subspace or complex ICA. Simulations are carried out that generalized flag manifolds arise naturally when we consider manifolds using our previous geodesic formula for Stiefel manifolds [9]. We will show previous manifolds, and extend the Riemannian optimization method to generalized flag manifolds [1, 3, 4, 5, 9, 10, 11]. The aim of the present paper is to introduce a new class of manifolds which generalize Grassmann manifolds [3]. Firstly, an appropriate Riemannian metric $g$ is introduced into a manifold $M$. Next, the Riemannian gradient $V = \text{grad}_W f(W)$ is used in place of the usual Euclidean gradient $\nabla f$. Finally, the current point $W_k$ is updated along the geodesic in the direction $-V_k$, to the point $: W_{k+1} = \varphi_M(W_k, -\text{grad}_W f(W_k), \eta_k)$, where $\gamma(t) = \varphi_M(W, V, t)$ denotes the equation of the geodesic on a manifold $M$ starting from $W \in M$ (i.e. $\gamma(0) = W$) in direction $V \in T_W M$ (i.e. $\gamma(0) = V$) with respect to a Riemannian metric $g$ on $M$.

The use of such Riemannian geometrical techniques for optimization on manifolds has been explored by recent authors, mainly over the real Stiefel and Grassmann manifolds [1, 3, 4, 5, 9, 10, 11]. The aim of the present paper is to introduce a new class of manifolds, generalized (or partial) flag manifolds, which generalize Grassmann manifolds. We will describe the relationships between this new class of manifolds and the previous manifolds, and extend the Riemannian optimization method to generalized flag manifolds using our previous geodesic formula for Stiefel manifolds [9]. We will show that generalized flag manifolds arise naturally when we consider dependent component analysis-type problems such as subspace or complex ICA. Simulations are carried out to compare the Riemannian optimization method with the ordinary gradient method.

**GENERALIZED FLAG MANIFOLDS**

We summarize the relationships between manifolds (Fig. 1) which have been recently investigated in neural networks, signal processing, numerical analysis, and scientific computing [1, 3, 7, 9]. The most fundamental is the orthogonal group $O(n)$, which is the Lie group of orthogonal matrices $\left\{ \tilde{W} \in \mathbb{R}^{n \times n} | \tilde{W}^T \tilde{W} = I_n \right\}$. The standard ICA problem can be solved by minimizing a cost function over $O(n)$. Other manifolds discussed in this paper are all descendants of $O(n)$. If we consider the manifold which is the set of orthogonal rectangular matrices $\{W = (w_1, \ldots, w_p) \in \mathbb{R}^{n \times p} | W^T W = I_p, n \geq p\}$, we get a (real) Stiefel manifold $\text{St}(n, p; \mathbb{R})$. $O(n)$ acts transitively on $\text{St}(n, p; \mathbb{R})$ by matrix multiplication. The subgroup of $O(n)$ which fixes a point $\tilde{W} \in \text{St}(n, p; \mathbb{R})$ is called the isotropy subgroup of $\tilde{W}$, which is isomorphic to $O(n-p)$. Manifold theory tells us that $\text{St}(n, p; \mathbb{R})$ can be regarded as the quotient space $O(n)/O(n-p)$. In other words,
we can interpret $O(n)$ as a fiber bundle over $\text{St}(n,p;\mathbb{R})$ whose fiber is isomorphic to $O(n-p)$. A manifold expressed as $G/H$ is called a homogeneous space, where $G$ is a Lie group and $H$ is a closed subgroup of $G$. Next comes the (real) Grassmann manifold $\text{Gr}(n,p;\mathbb{R})$, which is defined to be the set of $p$-dimensional subspaces in $\mathbb{R}^n$; $\text{Gr}(n,p;\mathbb{R})$ concerns a subspace in $\mathbb{R}^n$ spanned by $w_1,\ldots,w_p$ instead of the individual frame vectors $(w_1,\ldots,w_p)$ themselves. In other words, any two matrices $W_1,W_2 \in \text{st}(n,p;\mathbb{R})$ related by $W_2 = W_1 R$, where $R \in O(p)$, correspond to the same point on $\text{Gr}(n,p;\mathbb{R})$; we say we identify these two matrices. More formally said, $\text{St}(n,p;\mathbb{R})$ is a fiber bundle over $\text{Gr}(n,p;\mathbb{R})$, whose fiber is isomorphic to $O(p)$. Therefore, as a homogeneous space, $\text{Gr}(n,p;\mathbb{R}) \simeq O(n)/O(p) \times O(n-p)$.

Let us introduce a generalized flag manifold $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$, which is by definition the set of the direct sum of the subspaces $V = V_1 \oplus V_2 \oplus \cdots \oplus V_r \subset \mathbb{R}^n$, where $r$ and each $\dim V_i := d_i$ are fixed ($\sum_i d_i = p$).\footnote{This definition is slightly different from the standard definition of a generalized flag manifold, yet both are diffeomorphic to each other. For more details, see [10].} We represent a point on this manifold by $W \in \text{St}(n,p;\mathbb{R})$, which can be decomposed as $W = (W_1,W_2,\ldots,W_r),W_i = (w_i^1,w_i^2,\ldots,w_i^{d_i})$, where $w_i^k \in \mathbb{R}^n$, $k = 1,\ldots,d_i$ for some $i$, form the orthogonal basis of $V_i$. As in the case of $\text{Gr}(n,p;\mathbb{R})$, we are concerned about each subspace $V_i$ rather than frame vectors $w_i^k$ themselves, hence, as a point on $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$, any two matrices $W_1,W_2 \in \text{St}(n,p;\mathbb{R})$ related by $W_2 = W_1 \text{diag}(R_1,R_2,\ldots,R_r)$, are identified, where $R_i \in O(d_i)$; namely, $\text{St}(n,p;\mathbb{R})$ is a fiber bundle over $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$, whose fiber is isomorphic to $O(d_1) \times \cdots \times O(d_r)$. As a homogeneous space, $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R}) \cong O(n)/O(d_1) \times \cdots \times O(d_r) \times O(n-p)$. $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$ is locally isomorphic to $\text{St}(n,p;\mathbb{R})$ as a homogeneous space when all $d_i$ (1 $\leq i \leq r$ = 1), and it reduces to a Grassmann manifold if $r = 1$.

To derive the update rule for the Riemannian gradient descent geodesic method, we need to obtain the formulas for the natural gradient and geodesics on $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$. By differentiating the constraints on the generalized flag manifold, we see a tangent vector $V = (V_1,\ldots,V_r)$ of $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$ at $W = (W_1,\ldots,W_r)$ is characterized by

$$W^TV + V^TW = O, W_i^TV_i = O, \quad i = 1,\ldots,r.$$ (1)

First, let us derive the equation of a geodesic on a generalized flag manifold; it can be obtained based on our geodesic formula for the Stiefel manifold with respect to the normal metric $g_{\text{St}(n,p;\mathbb{R})}(V_1,V_2) = \text{tr}V_1^T(I - \frac{1}{2}WW^T)V_2$, where $V_1,V_2 \in T_W\text{St}(n,p;\mathbb{R})$ [9]:

$$\varphi_{\text{St}(n,p;\mathbb{R})}(W_i - \text{grad}_{W}^{\text{St}(n,p;\mathbb{R})}f,t) = \exp(-t(\nabla f(W)W^T - W \nabla f(W^T)))W.$$ (2)

Here we recall the following theorem: Let $p: \tilde{M} \rightarrow M$ be a Riemannian submersion (see Fig. 2), that is, for any $\tilde{m} \in \tilde{M}$, $(dp)_{\tilde{m}}$ is an isometry between $H_{\tilde{m}}$ and $T_{p(\tilde{m})}M$, where $H_{\tilde{m}}$ is the horizontal space in $T_{\tilde{m}}\tilde{M}$. Let $\tilde{c}(t)$ be a geodesic of $(\tilde{M},\tilde{g})$. If the vector $\tilde{c}(0)$ is horizontal, $\tilde{c}(t)$ is horizontal for any $t$, and the curve $p(\tilde{c}(t))$ is a geodesic of $(M,g)$ of the same length as $\tilde{c}(t)$ [6, 9]. Because the projection $\pi: \text{St}(n,p;\mathbb{R}) \rightarrow \text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$ is
a Riemannian submersion, and any tangent vector $V \in T_W \text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$ belongs to $H_W$ in $T_W\text{St}(n,p;\mathbb{R})$, this theorem ensures that the normal metric $g_W^{\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})}$ coincides with $g_W^{\text{St}(n,p;\mathbb{R})}$ on $T_W \text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$, and that

$$\varphi_{\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})}(W,-\text{grad}_{W}^{\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})}f,t) = \varphi_{\text{St}(n,p;\mathbb{R})}(W,-\text{grad}_{W}^{\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})}f,t).$$  (3)

Next, using the following notations: $G = I - \frac{1}{2}WW^\top$, $X = \nabla_Wf = \left(\frac{\partial f}{\partial w_{ij}}\right)$, $(X_1,\ldots,X_r)$, $Y = G^{-1}\nabla_WY, Y = G^{-1}X_i$, we can get the natural gradient $V$ of a function $f$ on $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$ at $W$ with respect to $g^{\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})}$ by the orthogonal projection of $Y$ onto $T_W\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$ relative to $g_W^{\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})}$. In other words, $V$ is obtained by minimizing $\text{tr}\left\{(V-Y)^\top G(V-Y)\right\}$ under the tangency constraints of $V$. This can be solved by the Lagrangian multiplier method and we get:

$$V_i = X_i - (W_iW_i^\top X_i + \sum_{j \neq i} W_jX_j^\top W_i).$$  (4)

**SUBSPACE ICA**

Subspace ICA (a.k.a. independent subspace analysis) was proposed by Hyvärinen and Hoyer [8] by relaxing the assumption of standard ICA, namely each source signal is statistically independent. The subspace ICA task is to decompose a gray-scale image $I(x,y)$ into linear combination of basis images $a_i(x,y)$: $I(x,y) = \sum_{i=1}^{n} s_i a_i(x,y)$, where $s_i$ is a coefficient. Let the inverse filter of this model be $s_i = \langle w_i, I \rangle = \sum_{x,y} w_i(x,y)I(x,y)$. The goal is to estimate $s_i$ (or equivalently $w_i(x,y)$) from a set of given images. In the subspace ICA model, we assume $s = (s_1,\ldots,s_n)^\top$ is decomposed into disjoint subspaces $S_1,\ldots,S_r$, $(\dim S_i = d_i)$, where signals within each subspace are allowed to be dependent on each other, and signals belonging to different subspaces are statistically independent. As a cost function to solve this task, we take the negative log-likelihood:

$$f(\{w_i\}) = -\sum_{k=1}^{K} \log L(I_k;\{w_i\}) = -\sum_{k=1}^{K} \sum_{j=1}^{r} \log p \left( \sum_{i \in S_j} \langle w_i, I_{k} \rangle^2 \right)$$  (5)

where $k$ denotes the index of sample images and $p$ denotes the exponential distribution $p(x) = \alpha \exp(-\alpha x)$. Since the subspace ICA algorithm uses pre-whitening, solving the subspace ICA task reduces to minimizing $f$ over the orthogonal group $O(n)$, as standard ICA. However, because of the statistical dependence of signals within each $S_i$, the objective function $f$ is invariant under rotation within each subspace: $W \mapsto W\text{diag}(R_1,\ldots,R_r)$, where $R_i \in O(d_i)$. Therefore, the subspace ICA task should be regarded as optimization on the generalized flag manifold $\text{Fl}(n,d_1,\ldots,d_r;\mathbb{R})$ instead of simply $O(n)$.

To demonstrate the Riemannian optimization method is effective, we applied it to the following subspace ICA task: We prepared 10000 image patches of
16 × 16 pixels at random locations extracted from monochrome photographs of natural images. (The dataset and subspace ICA code is distributed by Hyvärinen http://www.cis.hut.fi/projects/ica/data/images). As a preprocessing step, the mean gray-scale value of each image patch was subtracted, then the dimension of the image was reduced from 256 to 160 by PCA \((n = 160)\), and the data were whitened. We performed subspace ICA on this dataset; the 160-dimensional vector space was decomposed into \(40 \times 4\)-dimensional subspaces \((i.e., r = 40, d_i = 4)\) by minimizing \(f\) over Fl\((160, 4, \ldots, 4; \mathbb{R})\). We compared the Riemannian optimization method with the standard gradient descent method used in [8] for this minimization problem. The former is:

\[
W_{k+1} = \varphi_{\text{Fl}(160, 4, \ldots, 4; \mathbb{R})}(W_k, -\nabla_{W_k} f_{\text{Fl}(160, 4, \ldots, 4; \mathbb{R})})(W_k, \eta_k) := \gamma_1(\eta_k),
\]

while the latter

\[
W_{s+1} = \text{proj}(W_s - \mu_s \Delta s) := \gamma_2(\mu_s),
\]

where \(\text{proj}\) means the projection onto \(O(160)\) by SVD. The learning constant \(\eta_k, \mu_s\) was chosen at each iteration based on the Armijo rule such that

\[
f(W_k) - f(\gamma_1(\eta_k)) \geq \frac{1}{2} \eta_k g_{W_k}^\text{Fl}(\Delta_k, \Delta_k), f(W_k) - f(\gamma_1(2\eta_k)) \leq \eta_k g_{W_k}^\text{Fl}(\Delta_k, \Delta_k)
\]

\[
f(W_s) - f(\gamma_2(\mu_s)) \geq \frac{1}{2} \mu_s \langle \delta_s, \delta_s \rangle, f(W_s) - f(\gamma_2(2\mu_s)) \leq \mu_s \langle \delta_s, \delta_s \rangle
\]

are satisfied, where \(\text{Fl}\) denotes \(\text{Fl}(160, 4, \ldots, 4; \mathbb{R})\), \(\Delta_k = -\nabla_{W_k} f_{\text{Fl}(160, 4, \ldots, 4; \mathbb{R})}(W_k)\), \(\delta_s\) denotes the orthogonal projection of \(-\frac{\partial f}{\partial W_s}\) onto \(T_{W_s}O(160)\) with respect to the Euclidean metric \(\langle \cdot, \cdot \rangle\). The behavior of these algorithms is shown in Fig. 3(a). In the early stages of learning, the geodesic method decreased the cost much faster than the standard gradient method. The inverse filters recovered by the geodesic method \(w_i(x, y)\) are shown in Fig. 3(b). We obtained complex cell-like filters, which were grouped into 4-dimensional subspaces. We found no significant difference between the points of convergence of the two methods, and neither method appeared to get ‘stuck’ in a local minimum.
**COMPLEX ICA**

Let us consider an optimization problem on the class of complex Stiefel manifold.

$$F : \text{St}(n, p; \mathbb{C}) \rightarrow \mathbb{R},$$

where $\text{St}(n, p; \mathbb{C}) = \{ W = (w_1, \ldots, w_p) = W_\mathbb{R} + iW_\mathbb{I} \in \mathbb{C}^{n \times p} \mid |W^\mathbb{R}W - I_p| \}$ (H denotes the Hermitian transpose operator). We assume $F$ is a smooth function of the norm of column vectors $||w_i||$ ($i = 1, \ldots, p$), which is satisfied by many signal processing tasks including complex ICA.

Because the cost function $F$ is real-valued, $\text{St}(n, p; \mathbb{C})$ should be regarded as a *real manifold* rather than a complex manifold. The real manifold underlying $\text{St}(n, p; \mathbb{C})$ is a submanifold $M$ in $\mathbb{R}^{2n \times p}$ defined by the constraints:

$$M := \{ \bar{W} = \left( \begin{array}{c} W_\mathbb{R} \\ W_\mathbb{I} \end{array} \right) \in \mathbb{R}^{2n \times p} | W_\mathbb{R}^T W_\mathbb{R} + W_\mathbb{I}^T W_\mathbb{I} = I_p, W_\mathbb{R}^T W_\mathbb{I} = W_\mathbb{I}^T W_\mathbb{R} = 0 \} .$$

The cost function $F$ over $\text{St}(n, p; \mathbb{C})$ corresponds to the function $F'(\bar{W}) := F(\bar{W})$ over $M$. However, it is difficult to deal with the constraints (9) as is; we embed $M$ into $\mathbb{R}^{2n \times 2p}$ by the following map:

$$\tau : \left( \begin{array}{c} W_\mathbb{R} \\ W_\mathbb{I} \end{array} \right) = \left( \begin{array}{c} w_1^\mathbb{R} \cdots w_p^\mathbb{R} \\ w_1^\mathbb{I} \cdots w_p^\mathbb{I} \end{array} \right) \mapsto \tilde{W} = \left( \begin{array}{c} w_1^\mathbb{R} - w_1^\mathbb{I} w_2^\mathbb{R} - w_2^\mathbb{I} \cdots w_p^\mathbb{R} - w_p^\mathbb{I} \\ w_1^\mathbb{I} w_2^\mathbb{R} - w_1^\mathbb{R} \cdots w_p^\mathbb{I} w_p^\mathbb{R} \end{array} \right).$$

We consider the embedded manifold $N = \tau(M)$ in $\mathbb{R}^{2n \times 2p}$ and the function $f : N \rightarrow \mathbb{R}$ associated with the embedding $\tau$. $\tilde{W} \mapsto f(\tilde{W}) := F'(\tilde{W})$. If $\bar{W} \in M$, then $\tilde{W} \in \text{St}(2n, 2p; \mathbb{R})$ holds. It turns out that $N = \text{St}(2n, 2p; \mathbb{R}) \cap T$, where $T = \tau(\mathbb{R}^{2n \times p})$ forms a subspace in $\mathbb{R}^{2n \times 2p}$. As such, minimizing $F$ over $\text{St}(n, p; \mathbb{C})$ is transformed to minimizing $f$ over $N$.

Furthermore, the assumption of $F$ gives $N$ an additional structure. We see the transformation on $\text{St}(n, p; \mathbb{C})$:

$$W = (w_1, \ldots, w_p) \mapsto (e^{i\theta_1}w_1, \ldots, e^{i\theta_p}w_p)$$

and $F$ is invariant under the transformation (11) from the assumption. Thus the function $f$ is also invariant under the transformation (12). Therefore, $f$ can be interpreted as a function over a submanifold of a generalized flag manifold: $N' = \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \cap T^2$.

In fact, the following two facts allow us to consider just $\text{Fl}(2n, 2, \ldots, 2; \mathbb{R})$ instead of its submanifold $N'$. First, $N'$ is a totally geodesic submanifold of $\text{Fl}(2n, 2, \ldots, 2; \mathbb{R})$, yet both are locally isomorphic to each other as a homogeneous space, and we use $\text{Fl}(2n, 2, \ldots, 2; \mathbb{R})$ by abuse of notation.

---

2 Strictly speaking $\text{Fl}(2n, 2, \ldots, 2; \mathbb{R})$ should be replaced with $SO(2n)/SO(2) \times \ldots \times SO(2)$, yet both are locally isomorphic to each other as a homogeneous space, and we use $\text{Fl}(2n, 2, \ldots, 2; \mathbb{R})$ by abuse of notation.
that is, a geodesic on \( \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \) emanating from \( \tilde{W} \in N' \) in direction \( \tilde{V} \in T_{\tilde{W}}N' \) is always contained in \( N' \). Second, the natural gradient of \( f \) on \( N' \) at \( \tilde{W} \) coincides with the natural gradient of \( \tilde{f} \) on \( \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \) at \( \tilde{W} \), that is, we can obtain \( \text{grad}_{\tilde{W}}^{N'} f \) by substituting \( \left( \frac{\partial f}{\partial u^{i}} - i \frac{\partial f}{\partial u^{3}} \right) \) for \( X_i \) in the formula for the natural gradient of \( \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \), \( (d_i = 2, r = p) \). Note that \((X_1, \ldots, X_p)\) is the gradient of \( f \) in \( T \) relative to the Euclidean metric. To summarize, minimizing \( F \) over \( \text{St}(n, p; \mathbb{C}) \) can be solved by minimizing the function \( f \) over the submanifold \( N' \) of \( \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \); for minimizing \( f \) on \( N' \), we have only to apply the Riemannian optimization method for \( \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \) to \( f \).

To explore the behavior of the Riemannian gradient descent geodesic method on the complex Stiefel manifold as described above, we performed a numerical experiment for complex ICA. Let us assume we are given 9 source signals \( x = (x_1, \ldots, x_9) \top \) (Fig. 4(b)) which are complex-valued instantaneous linear mixture of four independent QAM16 signals \( s = (s_1, \ldots, s_4) \top \) and five complex-valued Gaussian noise signals \( u = (s_5, \ldots, s_9) \top \) (Fig. 4(a)) such that \( x = A \begin{pmatrix} s \\ u \end{pmatrix} \), where \( A \) is a randomly generated nonsingular \( 9 \times 9 \) matrix. We assume we know in advance the number of the noise signals. The task of complex ICA under this assumption is to recover only non-noise signals \( y = (y_1, \ldots, y_4) \top \) so that \( y = W \top x \). As a preprocessing stage, we first center the data and then whiten it by SVD. Thus, \( n \times p \) demixing matrix \( W \) can be regarded as a point on the complex Stiefel manifold \( \text{St}(n, p; \mathbb{C}) \), namely \( W^H W = I_p \). As an object function, we use a kurtosis-like higher-order statistics: \( F(W) = \sum_{i=1}^{4} \mathbb{E} \left[ ||y_i(t)||^4 \right] \) [4], then by minimizing \( F(W) \) over \( \text{St}(n, p; \mathbb{C}) \) we can solve the task.

We compared two algorithms for optimizing \( F(W) \) over \( \text{St}(n, p; \mathbb{C}) \). One is the Riemannian optimization method: \( \tilde{W}_{k+1} = \varphi_{\text{Fl}(2n, 2, \ldots, 2; \mathbb{R})} (\tilde{W}_k, - \text{grad}_{\tilde{W}_k} f(\tilde{W}_k), \eta_k) := \gamma_1(\eta_k) \), and another is the standard gradient descent method followed by projection: \( W_{s+1} = \text{pro}(W_s - \mu_s \frac{\partial f}{\partial W_s}) := \gamma_2(\mu_s) \), where \( \frac{\partial f}{\partial W_s} \) denotes \( \frac{\partial f}{\partial W^{\Re}} + i \frac{\partial f}{\partial W^{\Im}} \), and pro means the projection onto \( \text{St}(n, p; \mathbb{C}) \) via complex SVD. Both \( \text{grad}_{\tilde{W}_k} f(\tilde{W}_k) \) and \( \frac{\partial f}{\partial W} \) are computed by substituting \( \frac{\partial ||y||^4}{\partial u_i} = 2 ||y||^2 (y^* x + y_i x^*) \) and \( \frac{\partial ||y||^4}{\partial u_i} = 2 ||y||^2 (y^* x - y_i x^*) \).

Recall that we map \( \text{St}(n, p; \mathbb{C}) \) to \( \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \) and the Riemannian optimization method for \( f \) updates the matrices on \( \text{Fl}(2n, 2, \ldots, 2; \mathbb{R}) \) using the correspondence between \( W \) and \( \tilde{W} \) (10). After \( \tilde{W} \) converges to \( \tilde{W}_\infty, \tilde{W}_\infty \) is pulled back to \( \text{St}(n, p; \mathbb{C}) \) to give a demixing matrix \( W_\infty \). We used the Armijo rule to set the learning constant at each iteration as the subspace ICA experiment. The separation result is shown in Fig. 4(c). The QAM 16 constellation was well-recognized after recovery. Both algorithms were tested for 100 trials. On each trial, a random nonsingular matrix was used to generate the data; a random unitary matrix was chosen as a initial demixing matrix; we iterated for 200 steps. The plots of Fig. 4(d) show the average behavior of these two algorithms over 100 trials. We observed that the Riemannian optimization method outperformed the standard gradient descent method followed by projection, particularly in the early stages of learning much the same way as the subspace ICA experiment.
ACKNOWLEDGEMENTS

This work is partly supported by JSPS Grant-in-Aid for Exploratory Research 16650050, and MEXT Grant-in-Aid for Scientific Research on Priority Areas 17022033.

REFERENCES


FIGURE 4. Complex ICA experiment