

Averaging Complex Subspaces via a Karcher Mean Approach

K. Hüper^{1a}, M. Kleinstaubert^{2b}, H. Shen^{3b}

^a*Department of Mathematics, Julius-Maximilians-Universität Würzburg,
97074 Würzburg, Germany.*

^b*Department of Electrical Engineering and Information Technology,
Technische Universität München, 80333 Munich, Germany.*

Abstract

We propose a conjugate gradient type optimization technique for the computation of the Karcher mean on the set of complex linear subspaces of fixed dimension, modeled by the so-called Grassmannian. The identification of the Grassmannian with Hermitian projection matrices allows an accessible introduction of the geometric concepts required for an intrinsic conjugate gradient method. In particular, proper definitions of geodesics, parallel transport, and the Riemannian gradient of the Karcher mean function are presented. We provide an efficient step-size selection for the special case of one dimensional complex subspaces and illustrate how the method can be employed for blind identification via numerical experiments.

Keywords: conjugate gradient algorithm, Grassmannian, complex projective space, complex linear subspaces, Karcher mean.

1. Introduction

In a wide range of signal processing applications and methods, subspaces of a fixed dimension play an important role. Signal and noise subspaces of covariance matrices are well studied objects in classical applications, such as subspace tracking [1] or direction of arrival estimation [2]. More recently, a significant amount of work is focussed on applying subspace based methods to image and video analysis [3], as well as to matrix completion problems [4]. One fundamental challenge amongst these works is the study of the statistical properties of distributions of subspaces. Specifically, in the present work, we

¹hueper@mathematik.uni-wuerzburg.de

²Corresponding author, kleinstaubert@tum.de, web: www.gol.ei.tum.de

³hao.shen@tum.de

This work has partially been supported by the Cluster of Excellence *CoTeSys - Cognition for Technical Systems*, funded by the German Research Foundation (DFG).

are interested in computing the mean of a set of subspaces of equal dimension via averaging.

The averaging process, considered in this paper, employs the intrinsic geometric structure of the underlying set and is also known as the computation of the *Karcher mean* (in differential geometry, [5]), Fréchet mean or barycentre (statistics), geometric mean (linear algebra and matrix analysis), or center of mass (physics). General concepts of a geometric mean have been extensively studied from both theoretical and practical points of view. To mention just a few, they include probability theory and shape spaces [6, 7], imaging [8], linear algebra and matrix analysis [9], interpolation [10], and convex and differential geometry [11, 12].

An appropriate mathematical framework is given by the so-called Grassmannian, which assigns a differentiable manifold structure to the set of subspaces of equal dimension. Usually, this is achieved by identification with a matrix quotient space.⁴ In this work, we do not follow such an approach. By following [14] instead, we identify the set of subspaces of equal dimension with a set of matrices. More precisely, we consider the set of Hermitian projectors of fixed rank, which inherits its differentiable structure from the surrounding vector space of Hermitian matrices. In contrast to [14], we consider the complex case here. The identification of the complex Grassmannian with Hermitian projection matrices allows an accessible introduction of the geometric concepts such as geodesics, parallel transport, and the Riemannian gradient of the Karcher mean function.

In general, computing the Karcher mean on a smooth manifold involves a process of optimization, which by its own is of both theoretical and practical interest. Various numerical methods have been developed on the Grassmannian, such as a direct method [15], gradient descent algorithms [16], Newton's method [14], and conjugate gradient methods [17, 18].

In this work, we focus on the development of conjugate gradient methods. These methods have been proven to be efficient in many applications due to their trade-off between computational complexity and excellent convergence properties. In particular, we propose an efficient step-size selection for the interesting case where the Grassmannian is equal to the complex projective space. Moreover, we outline how the developed method can be employed for blind identification.

The paper is organized as follows. Section 2 recalls some basic concepts in differential geometry, which make the present work intuitive and self-contained. An abstract framework of conjugate gradient methods on smooth manifolds is given in Section 3. In Section 4, the geometry of the Grassmannian is presented, followed by a detailed analysis of the of the Karcher mean function in Section 5. A geometric CG algorithm is given in Section 6 for the computation of the Karcher mean on the Grassmannian in general, together with a particularly

⁴The set of m -dimensional subspaces of \mathbb{C}^n is identified with $\mathbb{C}_*^{n \times m} / GL(m)$, cf. [13], $\mathbb{C}_*^{n \times m}$ is the set of full rank $(n \times m)$ -matrices, and $GL(m)$ are the complex invertible $(m \times m)$ -matrices. The equivalence relation is defined by $X \sim Y \Leftrightarrow X = gY$ for some $g \in GL(m)$.

efficient step-size selection for the special case of the complex projective space. In Section 7, we outline how the proposed approach of averaging subspaces is evidenced to be useful in blind identification and a conclusion is drawn in Section 8.

2. Differential geometric concepts

In this section, we shortly recall and explain the differential geometric concepts that are needed for this work. We refer to [20] for a detailed insight into differential and Riemannian geometry and for the formal definitions of the mathematical objects, and to [13] for an introduction of the topic with a focus on matrix manifolds.

Strictly speaking, a *manifold* M is a topological space that can locally be continuously mapped to some linear space, where this map has a continuous inverse. These maps are called *charts*, and since charts are invertible, we can consider the change of two charts around any point in M as a local map from the linear space into itself. M is a *differentiable* or *smooth* manifold, if these maps are smooth for all points in M . Many data sets considered in signal processing are subsets of such a manifold. Important examples are matrix groups, the set of subspaces of fixed dimension, the set of matrices with orthonormal columns (so-called Stiefel manifold), the set of positive definite matrices, etc.

To every point x in the smooth manifold M one can assign a *tangent space*, consisting of all velocities of smooth curves in M that pass x . Formally, we define

$$T_x M := \{\dot{\alpha}_x(0) \mid \alpha(t) \subset M, \alpha_x(0) = x\}. \quad (1)$$

Intuitively, $T_x M$ contains all possible directions in which one can tangentially pass through x . The elements of $T_x M$ are called *tangent vectors* at x .

A *Riemannian manifold* M is a smooth manifold with a scalar product $g_x(\cdot, \cdot)$ assigned to each tangent space $T_x M$ that varies smoothly with x , the so called *Riemannian metric*. We drop the subscript x if it is clear from the context which tangent space g refers to. The corresponding norm will be denoted by $\|\cdot\|_g$. The Riemannian metric allows to measure the distance on the manifold. As a natural extension of a straight line in the Euclidean space, a *geodesic* is defined to be a smooth curve in M that connects two sufficiently close points with shortest length. The length of a smooth curve $\alpha: (a, b) \rightarrow M$ on a Riemannian manifold is defined as

$$L(\alpha) = \int_a^b \sqrt{g_{\alpha(t)}(\dot{\alpha}(t), \dot{\alpha}(t))} dt. \quad (2)$$

In Euclidean space, two velocities at different locations are both vectors in this space. This allows to form linear combinations and scalar products of these vectors. In the manifold setting, however, this is not possible, since these velocities are elements in different (tangent) spaces. We hence need a way to identify tangent vectors at $x \in M$ with tangent vectors at $y \in M$ if $x \neq y$. To that end, we assume that there is a unique geodesic in M that connects x and y , say $\gamma(t)$, with $\gamma(0) = x$ and $\gamma(\tau) = y$, being possible if x, y are not too far apart.

The *parallel transport* along $\gamma(t)$ admits one way of identifying $T_x M$ with $T_y M$. A rigorous definition is beyond the scope of this work, but loosely speaking, the transportation is done in such a way that during the transportation process, there is no intrinsic rotation of the transported vector. In particular, this leaves the scalar product between the transported vector and the velocity of the curve invariant.

Certainly, such an identification of tangent vectors depends on the geodesic. Consider for example a sphere with two different geodesics connecting the south with the north pole (i.e. two meridians) that leave the south pole by an angle of $\pi/2$. Parallel transporting the same vector along both meridians from the south pole to the north will result in two antiparallel vectors at the north pole. Note that the identification of different tangent spaces via parallel transport along a geodesic is just *one* particular instance of a more general concept termed *vector transport* in [13].

In order to minimize a real valued function on M , we have to extend the notion of a gradient to the Riemannian manifold setting. To that end, recall that if $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is smooth in x , there is a unique vector $G(x)$ such that

$$\frac{d}{dt}f(x + tH)|_{t=0} = G(x)^\top H =: \langle G(x), H \rangle_{\text{Euclid}} \quad \text{for all } H \in \mathbb{R}^n, \quad (3)$$

where $(\cdot)^\top$ denotes transpose. Typically, we write $\nabla f(x) := G(x)$ and call it the *gradient* of f at x . This coordinate free definition of a gradient can be straightforwardly adapted to the manifold case. Let

$$f: M \rightarrow \mathbb{R} \quad (4)$$

be smooth in $x \in M$. There is a unique tangent vector $G(x) \in T_x M$ such that

$$Df(x)H := \frac{d}{dt}(f \circ \gamma)(t)|_{t=0} = g_x(G(x), \dot{\gamma}(0)) \quad (5)$$

for all geodesics γ with $\gamma(0) = x$ and $\dot{\gamma}(0) = H$. We denote the *Riemannian gradient* as $\text{grad } f(x) := G(x)$. Note, that the Riemannian gradient is a tangent vector in the respective tangent space that depends on the chosen Riemannian metric. It is unique due to the Riesz representation theorem. The following special case is of particular interest. Let M be a submanifold of some Euclidean space E with Riemannian metric induced by the surrounding space, i.e. the Riemannian metric is obtained by restricting the scalar product from E to the tangent spaces. In this setting, the *normal subspace* is the orthogonal complement of the tangent space. Assume furthermore, that the function f that is to be minimized in (4) is in fact the restriction of a function \hat{f} that is globally defined on the entire surrounding Euclidean space. If Π_x denotes the orthogonal projection from E onto the tangent space $T_x M$, then the Riemannian gradient is just the projection of the gradient of \hat{f} in E . In formulas, this reads as

$$\text{grad } f(x) = \Pi_x \nabla \hat{f}(x). \quad (6)$$

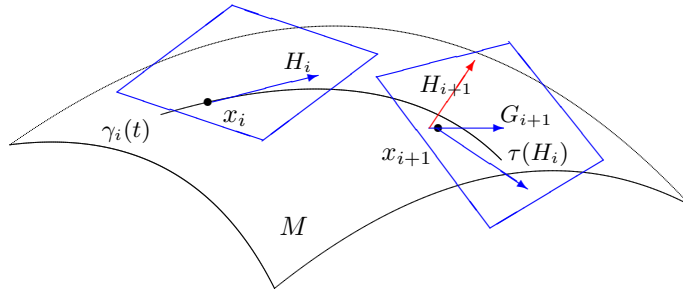


Figure 1: Illustration of the geometric conjugate gradient method.

3. A conjugate gradient method on manifolds

In this section, we recall one possibility of how to transfer the concept of conjugate gradient (CG) methods to the manifold setting. We refer to [13] for a more general approach that uses retractions on manifolds. Ultimately, the latter approach might lead to a whole set of general methods to minimize (4). The CG method is initialized by some $x_0 \in M$ and the descent direction $H_0 := -\text{grad } f(x_0)$ is given by the Riemannian gradient. Subsequently, sweeps are iterated that consist of two steps, a *line search* in a given direction (i.e. along a geodesic in that direction) followed by an update of the *search direction*. We illustrate the CG method on manifolds in Figure 2. Several different possibilities for these steps lead to different CG methods. Assume now that x_i , H_i , and $G_i := \text{grad } f(x_i)$ are given.

3.1. Line search

Given a geodesic γ_i with $\gamma_i(0) = x_i$ and $\dot{\gamma}_i(0) = H_i$, the line search aims to find $a_i \in \mathbb{R}$ that minimizes $f \circ \gamma: t \rightarrow \mathbb{R}$. We propose two approximations. The first is based on the assumption that $f \circ \gamma$ has its minimum near 0, which under certain mild conditions follows from the fact that x_i is already near the optimum. The step-size is chosen via a one dimensional Newton step, cf. [17], i.e.

$$a_i^{\text{Newton}} := -\frac{\frac{d}{dt}(f \circ \gamma)(t)|_{t=0}}{\left| \frac{d^2}{dt^2}(f \circ \gamma)(t)|_{t=0} \right|}. \quad (7)$$

The absolute value in the denominator is chosen for the following reason. While being an unaltered one-dimensional Newton step in a neighborhood of a minimum the step size is the negative of a regular Newton step if $\frac{d^2}{dt^2}(f \circ \gamma)(t)|_{t=0} < 0$ and thus yields non-attractiveness for critical points that are no minima, cf. [21].

This approach, however, uses second order information of the cost function and is often computationally too expensive. An alternative approach is the Riemannian adaption of the backtracking line search, described in Algorithm 1 below. The new iterate is then given by

$$x_{i+1} = \gamma(a_i), \quad (8)$$

where a_i is either obtained by backtracking or by Eq. (7).

Algorithm 1: Backtracking line search on manifolds

Step 1: Choose $\bar{a} > 0$, $c, \rho \in (0, 1)$ and set $a \leftarrow \bar{a}$;
Step 2: **repeat** until $(f \circ \gamma)(a) \leq f(x_i) + c a g_x(G_i, H_i)$;
 $a \leftarrow \rho a$;
end repeat;
Step 3: Choose step-size $a_i^{\text{backtrack}} := a$;

3.2. Search direction update

In order to compute the new search direction $H_{i+1} \in T_{x_{i+1}}M$, we need to transport H_i and G_i , which are tangent to x_i , to the tangent space $T_{x_{i+1}}M$. This is done via parallel transport along the geodesic γ , which we denote by

$$\tau: T_{x_i}M \rightarrow T_{x_{i+1}}M. \quad (9)$$

The updated search direction is now chosen according to a Riemannian adaption of the Hestenes-Stiefel formula, or any other CG formula known from the Euclidean case, cf. [22]. Specifically, we have

$$H_{i+1} = -G_{i+1} + r_i \tau H_i, \quad (10)$$

where the most common formulas for r_i read in the manifold setting as

$$\begin{aligned} r_i^{HS} &= \frac{g(G_{i+1}, G_{i+1} - \tau G_i)}{g(\tau H_i, G_{i+1} - \tau G_i)} && \text{(Hestenes-Stiefel)} \\ r_i^{PR} &= \frac{g(G_{i+1}, G_{i+1} - \tau G_i)}{\|G_i\|_g^2} && \text{(Polak-Ribière)} \\ r_i^{FR} &= \frac{\|G_{i+1}\|_g^2}{\|G_i\|_g^2} && \text{(Fletcher-Reeves)} \\ r_i^{DY} &= \frac{\|G_{i+1}\|_g^2}{g(\tau H_i, G_{i+1} - \tau G_i)} && \text{(Dai-Yuan)} \\ r_i^* &= -\frac{g(G_{i+1}, G_{i+1} - \tau G_i)}{g(H_i, G_i)}. \end{aligned} \quad (11)$$

Albeit the nice performance in applications, convergence analysis of CG methods on smooth manifolds is still an open problem. To the best of the authors' knowledge, the only partial convergence result is provided in [23].

4. Geometry of the Grassmannian

The complex Grassmannian $\text{Gr}_{m,n}$ is defined as the set of complex m -dimensional \mathbb{C} -linear subspaces of \mathbb{C}^n . It provides a natural generalization of the familiar complex projective spaces. We denote the unitary group by

$$\text{U}_n := \{X \in \mathbb{C}^{n \times n} | X^H X = I_n\}, \quad (12)$$

where $(\cdot)^H$ denotes complex conjugate transpose, and I_n is the $(n \times n)$ -identity matrix. For computational purposes it makes sense to identify the Grassmannian $\text{Gr}_{m,n}$ with a set of self-adjoint Hermitian projection operators as

$$\text{Gr}_{m,n} := \{P \in \mathbb{C}^{n \times n} \mid P^H = P, P^2 = P, \text{tr } P = m\}, \quad (13)$$

i.e. the smooth manifold of rank m Hermitian projection operators of \mathbb{C}^n . Here, $\text{tr}(\cdot)$ is the trace of a matrix. In the sequel we describe the Riemannian geometry directly for the submanifold $\text{Gr}_{m,n}$ of $\mathbb{C}^{n \times n}$. As we will see, this approach has advantages that simplify both the analysis and the design of CG-based algorithms. We begin by recalling facts about the complex Grassmannian [24, 25]. Let

$$\mathfrak{u}_n := \{\Omega \in \mathbb{C}^{n \times n} \mid \Omega^H = -\Omega\} \quad \text{and} \quad \text{Herm}_n := \mathfrak{i} \mathfrak{u}_n \quad (14)$$

denote the *real* n^2 -dimensional vector spaces of skew-Hermitian and Hermitian matrices, respectively. Here we follow the terminology in group theory, with \mathfrak{u}_n being the Lie algebra for the unitary Lie group U_n . In particular, $e^{\mathfrak{u}_n} = U_n$, where $e^{(\cdot)}$ is the matrix exponential function.

Theorem 1. *The Grassmannian $\text{Gr}_{m,n}$ is a real, smooth, and compact submanifold of Herm_n of real dimension $2m(n-m)+1$. Moreover, the tangent space at an element $P \in \text{Gr}_{m,n}$ is given as*

$$T_P \text{Gr}_{m,n} = \{P\Omega - \Omega P \mid \Omega \in \mathfrak{u}_n\}. \quad (15)$$

It is useful for further analysis to define the linear operator

$$\text{ad}_P : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}, \quad \text{ad}_P(X) := [P, X] := PX - XP. \quad (16)$$

Lemma 1. *([14] for the real case) For any $P \in \text{Gr}_{m,n}$ the minimal polynomial of ad_P is equal to $s^3 - s$. Thus $\text{ad}_P^3 = \text{ad}_P$, i.e.,*

$$\text{ad}_P^2 H = [P, [P, H]] = H \quad \forall H \in T_P \text{Gr}_{m,n}. \quad (17)$$

In the sequel, we will always endow Herm_n with the Frobenius inner product, defined by

$$\langle X, Y \rangle := \text{tr}(XY). \quad (18)$$

The Euclidean Riemannian metric g_P on $\text{Gr}_{m,n}$ induced by the embedding space Herm_n is defined by the restriction of (18) to the tangent spaces, i.e.

$$g_P(H_1, H_2) = \text{tr}(H_1 H_2) \quad \forall P \in \text{Gr}_{m,n} \quad \text{and} \quad \forall H_1, H_2 \in T_P \text{Gr}_{m,n}. \quad (19)$$

Lemma 2. *([14] for the real case) Let $P \in \text{Gr}_{m,n}$ be arbitrary. The normal subspace at P in Herm_n is given by $N_P \text{Gr}_{m,n} = \{X - \text{ad}_P^2 X \mid X \in \text{Herm}_n\}$. The linear map*

$$\Pi_P : \text{Herm}_n \rightarrow \text{Herm}_n, \quad X \mapsto \text{ad}_P^2 X = [P, [P, X]] \quad (20)$$

is the self-adjoint Hermitian projection operator onto $T_P \text{Gr}_{m,n}$ with kernel $N_P \text{Gr}_{m,n}$.

In general, a geodesic is a minimizer of the variational problem (2), i.e. it is the solution of the corresponding Euler-Lagrange equation, the latter being a second order ordinary differential equation. The following result characterizes the geodesics on $\text{Gr}_{m,n}$.

Theorem 2. ([14] for the real case) *The geodesics of $\text{Gr}_{m,n}$ are exactly the solutions of the second order differential equation $\ddot{P} + [\dot{P}, [\dot{P}, P]] = 0$. The unique geodesic $P(t)$ with initial conditions $P(0) = P_0 \in \text{Gr}_{m,n}$, $\dot{P}(0) = H \in T_{P_0} \text{Gr}_{m,n}$ is given by*

$$P_H(t) = e^{t[H, P_0]} P_0 e^{-t[H, P_0]}. \quad (21)$$

Definition 1. *We define the Riemannian exponential map as*

$$\exp_{P_0} : T_{P_0} \text{Gr}_{m,n} \rightarrow \text{Gr}_{m,n}, \quad H \mapsto P_H(1). \quad (22)$$

As outlined above, we need the concept of parallel transport along geodesics to give vector addition a well defined meaning.

Lemma 3 ([26] for the real case). *For $P \in \text{Gr}_{m,n}$ and $G_0 \in T_P \text{Gr}_{m,n}$ the parallel transport of G_0 along the geodesic $P_H(t)$ is given by*

$$G_H(t) = e^{t[H, P]} G_0 e^{-t[H, P]}. \quad (23)$$

Note, that the complex case considered here follows by a straightforward adaption of the proof of the real case in [26].

5. Karcher mean

5.1. The distance between two complex subspaces

In the first step we investigate the Riemannian distance of two complex subspaces. For convenience, we denote the standard projector by

$$\mathcal{I} := \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}. \quad (24)$$

Let $P \in \text{Gr}_{m,n}$, and assume for the moment that P is sufficiently close to \mathcal{I} , i.e. that there is a *unique* geodesic emanating from \mathcal{I} to P . Together with Eq. (21) this implies the existence of a *unique* $Z = Z(P) \in B_0 \subset \mathbb{C}^{(n-m) \times n}$ where B_0 is a sufficiently small open ball around the zero matrix 0, such that

$$P = e^{\begin{bmatrix} 0 & -Z^H \\ Z & 0 \end{bmatrix}} \mathcal{I} e^{-\begin{bmatrix} 0 & -Z^H \\ Z & 0 \end{bmatrix}}. \quad (25)$$

Hence, Z can be considered as a function of P , implicitly defined by (25).

Lemma 4 (cf. Fig. ??). *With*

$$\mathcal{Z}(P) := \begin{bmatrix} 0 & -Z^H(P) \\ Z(P) & 0 \end{bmatrix}, \quad (26)$$

the geodesic distance from \mathcal{I} to P is given by $\text{dist}(\mathcal{I}, P) = \|\mathcal{Z}(P), \mathcal{I}\|$.

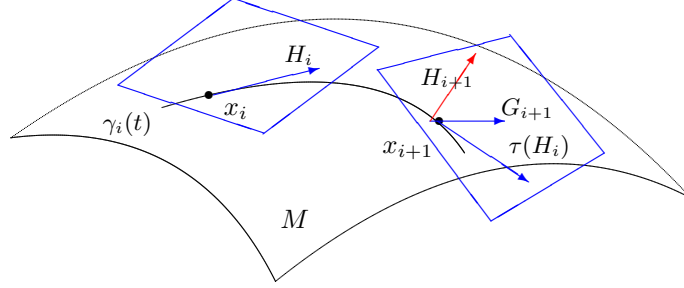


Figure 2: Illustration of the geometric conjugate gradient method.

Proof. Let $\gamma(t) = e^{tZ(P)} \mathcal{I} e^{-tZ(P)}$ be the geodesic (21) emanating from \mathcal{I} in direction $\text{ad}_{Z(P)}(\mathcal{I})$ with $\gamma(1) = P$. With $\dot{\gamma}(t) = e^{tZ(P)} [Z(P), \mathcal{I}] e^{-tZ(P)}$ and by Eq. (2) we get

$$\int_0^1 \|\dot{\gamma}(t)\| dt = \int_0^1 \| [Z(P), \mathcal{I}] \| dt = \| [Z(P), \mathcal{I}] \| = \sqrt{\text{tr}[Z(P), \mathcal{I}]^2} \quad (27)$$

and the result follows. \square

Let $U \subset \text{Gr}_{m,n}$ be a neighborhood around \mathcal{I} and $B \subset T_{\mathcal{I}} \text{Gr}_{m,n}$ such that $\exp_{\mathcal{I}}: B \rightarrow U$ is one-one and onto. Consider the function

$$f: U \rightarrow \mathbb{R}, \quad P \mapsto \text{dist}^2(\mathcal{I}, P). \quad (28)$$

To calculate the derivative we use an equivalent expression of f , namely

$$f(P) = -\text{tr} \left(\mathcal{I} \text{ad}_{Z(P)}^2(\mathcal{I}) \right). \quad (28')$$

Let $H \in T_P \text{Gr}_{m,n}$ be an arbitrary tangent vector. For the directional derivative we will use the abbreviation $Z' := DZ(P)H$, cf. Eq. (5). Therefore

$$Df(P)H = -2 \text{tr} (\mathcal{I} \text{ad}_Z(\text{ad}_{Z'}(\mathcal{I}))). \quad (29)$$

For computing the Riemannian gradient, we need an expression for $\text{ad}_{Z'}$. To that end, note that Eq. (25) is equivalent to

$$P = e^{\text{ad}_{Z(P)}}(\mathcal{I}) \quad (30)$$

and thus differentiating (30) with respect to P in direction H yields

$$D(P)H = H = e^{\text{ad}_Z} \left(\left(\frac{\text{id} - e^{-\text{ad}_{\text{ad}_Z}}}{\text{ad}_{\text{ad}_Z}} (\text{ad}_{Z'}) \right) (\mathcal{I}) \right). \quad (31)$$

Here, the operators e^{ad_Z} and $\frac{\text{id} - e^{-\text{ad}_{\text{ad}_Z}}}{\text{ad}_{\text{ad}_Z}}$ have to be understood via their series expansion and acting on the right as usual. By parallel transporting $\begin{bmatrix} 0 & K^H \\ K & 0 \end{bmatrix} \in T_{\mathcal{I}} \text{Gr}_{m,n}$ with $K \in \mathbb{C}^{(n-m) \times m}$ along the unique geodesic connecting \mathcal{I} and P it is easily seen that $H \in T_P \text{Gr}_{m,n}$ has the representation

$$H = e^{\text{ad}_Z} \begin{bmatrix} 0 & K^H \\ K & 0 \end{bmatrix}. \quad (32)$$

Using

$$\mathcal{K} := \begin{bmatrix} 0 & -K^H \\ K & 0 \end{bmatrix} \quad (33)$$

and (32) a lengthy but straightforward computation including the decomposition of the function $x \mapsto \frac{1-e^{-x}}{x} = \frac{\sinh x}{x} + \frac{1-\cosh x}{x}$ into even and odd parts shows that (31) is equivalent to

$$\begin{bmatrix} 0 & K^H \\ K & 0 \end{bmatrix} = \text{ad}_{\mathcal{K}}(\mathcal{I}) = \left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} (\text{ad}_{\mathcal{Z}'}) \right) (\mathcal{I}). \quad (34)$$

By assumption on P being close enough to \mathcal{I} , the selfadjoint operator $\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}}$ is invertible. Exploiting now the representation property of the ad-operator, i.e. $[\text{ad}_X, \text{ad}_Y] = \text{ad}_{[X,Y]}$, as well as linearity and anti-selfadjointness, i.e. $\text{tr}(A \text{ad}_B(C)) = -\text{tr}(C \text{ad}_B(A))$, we can conclude that (34) is equivalent to

$$\mathcal{Z}' = \left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} \right)^{-1} (\mathcal{K}). \quad (35)$$

In summary,

$$\begin{aligned} Df(P)H &= -2 \text{tr} \left(\mathcal{I} \text{ad}_{\mathcal{Z}} \left(\text{ad} \left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} \right)^{-1} (\mathcal{K}) \right) \right) \\ &= 2 \text{tr} \left(\left(\text{ad}_{\mathcal{I}} \mathcal{Z} \right) \left(\text{ad}_{\mathcal{I}} \left(\left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} \right)^{-1} (\mathcal{K}) \right) \right) \right) \\ &= -2 \text{tr} \left(\left(\underbrace{\text{ad}_{\mathcal{I}}^2 \mathcal{Z}}_{=\mathcal{Z}} \right) \left(\left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} \right)^{-1} (\mathcal{K}) \right) \right) \\ &= -2 \text{tr} \left(\mathcal{Z} \left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} \right)^{-1} (\mathcal{K}) \right) = -2 \text{tr}(\mathcal{Z}\mathcal{K}). \end{aligned} \quad (36)$$

The last equality in (36) is true by the self-adjointness of the operator $\left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} \right)^{-1}$ and $x \mapsto \frac{x}{\sinh x} = 1 - \frac{x^2}{6} + O(x^4)$ being an even function. In other words, $\left(\frac{\sinh \text{ad}_{\mathcal{Z}}}{\text{ad}_{\mathcal{Z}}} \right)^{-1}$ can be considered to act as the identity operator to the left onto \mathcal{Z} under the trace. Hence, critical points are characterized by

$$Df(P) = 0 \iff \Re \text{tr}(Z^H K) = 0 \forall K \in \mathbb{C}^{(n-m) \times m} \iff Z = 0. \quad (37)$$

Together with Eq. (30) this yields that the unique critical point of f is given by $P = \mathcal{I}$, as one would expect.⁵ Moreover, from (36) we can compute the Riemannian gradient of the function f by

$$\begin{aligned} Df(P)H &= -2 \text{tr}(\mathcal{Z}\mathcal{K}) = 2 \text{tr}([\mathcal{Z}, \mathcal{I}][\mathcal{K}, \mathcal{I}]) \\ &= 2 \text{tr}(e^{\mathcal{Z}}[\mathcal{Z}, \mathcal{I}]e^{-\mathcal{Z}}e^{\mathcal{Z}}[\mathcal{K}, \mathcal{I}]e^{-\mathcal{Z}}) = \text{tr}(2[\mathcal{Z}, P]H). \end{aligned} \quad (38)$$

Since $2[\mathcal{Z}, P] \in T_P \text{Gr}_{m,n}$ and since the trace is the Riemannian metric we can conclude, cf. (5), that

$$\text{grad } f(P) = 2[\mathcal{Z}, P]. \quad (39)$$

⁵ f is defined in a neighborhood of \mathcal{I} ensuring bijectivity of the Riemannian exponential.

Up to here all our computations were done sufficiently close to the standard projector \mathcal{I} . As the Grassmannian is a homogeneous space, meaning that every point $P \in \text{Gr}_{m,n}$ can be transformed to any other point on $Q \in \text{Gr}_{m,n}$ by a suitable unitary matrix transformation $P = \Theta Q \Theta^H$, $\Theta \in U_n$, we can now transfer all our computations to an arbitrary element of $\text{Gr}_{m,n}$. Let $Q \in \text{Gr}_{m,n}$ be arbitrary, and let P be sufficiently close to Q . Analogous to (30), we can then express $P = e^{[\xi(P), Q]} Q e^{-[\xi(P), Q]}$ for unique $\xi \in T_Q \text{Gr}_{m,n}$. Here, the tangent vector ξ plays the role which $\begin{bmatrix} 0 & Z^H \\ Z & 0 \end{bmatrix}$ played in (39). By a slight abuse of notation we consider the distance function between these *arbitrary* P and Q

$$f: \text{Gr}_{m,n} \rightarrow \mathbb{R}, \quad P \mapsto \text{tr } \xi \xi^H = \text{tr } \xi^2 = \text{dist}^2(P, Q). \quad (40)$$

Note, that by the above described invariance, the analogue of Eq. (37) yields the critical point condition

$$D f(P) = 0 \iff \xi = 0. \quad (41)$$

Theorem 3. *The Riemannian gradient, with respect to the Euclidean metric, of the function f , defined by (40), is given by*

$$\text{grad } f(P) = 2 \text{ad}_{[\xi, Q]} P. \quad (42)$$

Note that result (42) is in accordance with Proposition III 4.8 in [27].

Remark 1. *One can derive further explicit formulas for the distance, however, they are less well suited for gradient computations or numerics. Let $P, Q \in \text{Gr}_{m,n}$. For any given $\Theta \in U_n$ such that $P = \Theta^H \mathcal{I} \Theta$ we define*

$$\begin{bmatrix} Q_1 & Q_2 \\ Q_2^H & Q_3 \end{bmatrix} := \Theta Q \Theta^H. \quad (43)$$

Let $1 \geq \lambda_1 \geq \dots \geq \lambda_m \geq 0$ denote the eigenvalues of $Q_1 \in \text{Herm}_m$. Then

$$\text{dist}(P, Q) = \sqrt{2 \sum_{i=1}^m \arccos^2(\sqrt{\lambda_i})}. \quad (44)$$

Alternatively, let $1 \geq \mu_1 \geq \dots \geq \mu_{n-m} \geq 0$ denote the eigenvalues of Q_3 . Then

$$\text{dist}(P, Q) = \sqrt{2 \sum_{i=1}^{n-m} \arcsin^2(\sqrt{\mu_i})}. \quad (45)$$

In particular, if $P, Q \in \text{Gr}_{m,n}$ with $Q = Y Y^H$ and $Y^H Y = I_m$, then

$$\frac{1}{2} \text{dist}^2(P, Q) = \text{tr} \left(\arccos^2((Y^H P Y)^{\frac{1}{2}}) \right). \quad (46)$$

5.2. The Karcher mean

We now consider a geodesically convex open ball⁶ $\mathcal{B} \subset \text{Gr}_{m,n}$ containing all, say N data points Q_i . Note, that the Riemannian exponential map is bijective

⁶I.e. all points in \mathcal{B} can be connected by a unique shortest geodesic contained completely in \mathcal{B} . E.g. for a sphere, the maximal geodesically convex open balls are open hemispheres.

on \mathcal{B} and thus the results from the last section carry over to the subsequent analysis. Moreover, this assumption ensures that the Karcher mean is the unique minimizer of the function defined by (48), [5]. This seems to be a sensible assumption in many applications, where different data might be considered to be different measurements of one and the same observable. Let us assume that $P \in \mathcal{B}$ and thus, for each i there exists a unique $\xi_i \in T_{Q_i} \text{Gr}_{m,n}$ with

$$P = \exp_{Q_i}(\xi_i) = e^{[\xi_i, Q_i]} Q_i e^{-[\xi_i, Q_i]}. \quad (47)$$

Let the *Karcher mean* function now be defined as

$$F: \mathcal{B} \rightarrow \mathbb{R}, \quad P \mapsto \frac{1}{N} \sum_{i=1}^N \text{dist}^2(P, Q_i). \quad (48)$$

Adapting (28) and (26) accordingly, we get

$$F(P) = \frac{1}{N} \sum_{i=1}^N \text{tr} \xi_i^2 = -\frac{1}{N} \sum_{i=1}^N \text{tr} Q_i \text{ad}_{[\xi_i, Q_i]}^2 Q_i. \quad (49)$$

As a generalization of (41) we get the well known fact [5, 12] that

$$D F(P) = 0 \iff \sum_{i=1}^N e^{[\xi_i, Q_i]} \xi_i e^{-[\xi_i, Q_i]} = 0. \quad (50)$$

The interpretation of this condition is as follows. Let P be the unique critical point of F on \mathcal{B} . Attaching a suitable coordinate chart around P tells us that in this chart P is equal to the usual Euclidean geometric mean of the data points Q_i , expressed in exactly this chart.⁷ The Riemannian gradient of the Karcher mean now follows immediately from Theorem 3.

Theorem 4. *The Riemannian gradient of F , defined by (48), is as*

$$\text{grad } F(P) = \frac{2}{N} \sum_{i=1}^N \text{ad}_{[\xi_i, Q_i]} P. \quad (51)$$

5.3. Inverse of the Riemannian exponential

In the sequel we will present a procedure to explicitly compute the inverse of the Riemannian exponential.

To that end, let $Q \in \text{Gr}_{m,n}$ and $B \subset T_Q \text{Gr}_{m,n}$ be a neighborhood of 0 such that $\exp_Q: B \rightarrow \exp_Q(B)$ is a bijection. Assume further that $P \in \exp_Q(B)$. Thus, there exists a unique $\xi \in B$ such that

$$P = e^{[\xi, Q]} Q e^{-[\xi, Q]}. \quad (52)$$

From the previous section, it follows that $\text{tr} \xi^2 = \text{dist}^2(P, Q)$. A partial task in our optimization procedure will therefore be to compute ξ as a function of P for a given Q in (52). Let $\Theta \in \text{U}_n$ such that $Q = \Theta \mathcal{I} \Theta^H$. It follows that

$$\Theta^H P \Theta = e^{[\hat{\xi}, \mathcal{I}]} \cdot \mathcal{I} \cdot e^{-[\hat{\xi}, \mathcal{I}]} =: \hat{P} = \begin{bmatrix} \hat{P}_{11} & \hat{P}_{12} \\ \hat{P}_{12}^H & \hat{P}_{22} \end{bmatrix}, \quad (53)$$

⁷Such a chart is called Riemannian normal coordinate chart in the literature.

with $\widehat{\xi}$ of the form

$$\widehat{\xi} = \Theta^H \xi \Theta = \begin{bmatrix} 0 & -Z^H \\ Z & 0 \end{bmatrix}. \quad (54)$$

Using a singular value decomposition (SVD) as $Z^H = U \Sigma^T V^H$ we arrive at the representation

$$\begin{aligned} \begin{bmatrix} U^H & 0 \\ 0 & V^H \end{bmatrix} \widehat{P} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} &= e^{\begin{bmatrix} 0 & -\Sigma^T \\ \Sigma & 0 \end{bmatrix}} \cdot \mathcal{I} \cdot e^{-\begin{bmatrix} 0 & -\Sigma^T \\ \Sigma & 0 \end{bmatrix}} \\ &= \begin{bmatrix} \cos \sqrt{\Sigma^T \Sigma} & -\Sigma^T \frac{\sin \sqrt{\Sigma \Sigma^T}}{\sqrt{\Sigma \Sigma^T}} \\ \frac{\sin \sqrt{\Sigma \Sigma^T}}{\sqrt{\Sigma \Sigma^T}} \Sigma & \cos \sqrt{\Sigma \Sigma^T} \end{bmatrix} \cdot \mathcal{I} \cdot \begin{bmatrix} \cos \sqrt{\Sigma^T \Sigma} & \Sigma^T \frac{\sin \sqrt{\Sigma \Sigma^T}}{\sqrt{\Sigma \Sigma^T}} \\ -\frac{\sin \sqrt{\Sigma \Sigma^T}}{\sqrt{\Sigma \Sigma^T}} \Sigma & \cos \sqrt{\Sigma \Sigma^T} \end{bmatrix} \\ &= \begin{bmatrix} \cos^2 \sqrt{\Sigma^T \Sigma} & -\Sigma^T \frac{\cos \sqrt{\Sigma \Sigma^T} \sin \sqrt{\Sigma \Sigma^T}}{\sqrt{\Sigma \Sigma^T}} \\ \frac{\cos \sqrt{\Sigma \Sigma^T} \sin \sqrt{\Sigma \Sigma^T}}{\sqrt{\Sigma \Sigma^T}} \Sigma & \sin^2 \sqrt{\Sigma \Sigma^T} \end{bmatrix} \\ &= \begin{bmatrix} \cos^2 \sqrt{\Sigma^T \Sigma} & -\Sigma^T \frac{\sin 2\sqrt{\Sigma \Sigma^T}}{2\sqrt{\Sigma \Sigma^T}} \\ \frac{\sin 2\sqrt{\Sigma \Sigma^T}}{2\sqrt{\Sigma \Sigma^T}} \Sigma & \sin^2 \sqrt{\Sigma \Sigma^T} \end{bmatrix}. \end{aligned} \quad (55)$$

The above matrix valued trigonometric functions have to be interpreted via their series expansion. To be more precise, the matrix Σ is a rectangular diagonal matrix, and therefore $\Sigma \Sigma^T$, $\Sigma^T \Sigma$, and their corresponding square roots are diagonal (but square) as well. For $i \leq m$ the ii -th entry of $\cos^2 \sqrt{\Sigma^T \Sigma}$ equals the squared cosine of the i -th singular value of Σ , i.e. it equals $\cos^2 \sigma_i$, and equals 1 otherwise. The right-hand side of (55) can be efficiently computed via a CS-decomposition [28] of Y , where $\widehat{P} = Y Y^H$ with $Y^H Y = I_m$, or equivalently by an SVD of \widehat{P}_{12} . Using inverse trigonometric functions, ξ can now be constructed from (54). Ultimately, we have explicitly constructed the *inverse* of the Riemannian exponential map on $\text{Gr}_{m,n}$.

6. A conjugate gradient algorithm for computing the Karcher mean on the Grassmannian

In the last sections, all ingredients have been derived for a geometric CG algorithm as described in Section 3. Here, we focus on its implementation and provide an explicit pseudo code for computing the Karcher mean of a set of complex subspaces. Note that a projector P can be uniquely expressed as $P = X X^H$, where X is an element of the complex Stiefel manifold

$$\text{St}_{m,n} := \{X \in \mathbb{C}^{n \times m} \mid X^H X = I_m\}. \quad (56)$$

In order to compute the gradient of the Karcher mean function, the results in Section 5 require that after computing the tangent directions $\xi_i \in T_{Q_i} \text{Gr}_{m,n}$ in (47), one needs to parallel transport them back to $T_P \text{Gr}_{m,n}$, i.e. $e^{[\xi_i, Q_i]} \xi_i e^{-[\xi_i, Q_i]} \in T_P \text{Gr}_{m,n}$. The gradient of the Karcher mean function is simply the sum of all the parallel transported vectors in $T_P \text{Gr}_{m,n}$ according to (51).

Algorithm 2: Riemannian gradient of the Karcher mean on $\text{Gr}_{m,n}$

Input : $Y_i \in \text{St}_{m,n}$ for $i=1, \dots, N$ and $[X_1 \ X_2] \in \text{U}_n$ with $X_1 \in \text{St}_{m,n}$;
Step 1 : **for** $i = 1, \dots, N$ **do**;
 Compute the SVD of $X_1^H Y_i Y_i^H X_2 = U_i \Sigma_i V_i^H$;
 Compute $\Lambda_i := U_i^H X_1^H Y_i Y_i^H X_1 U_i$
Step 2 : Compute $\mathcal{Z} = \sum_{i=1}^N \begin{bmatrix} 0 & -Z_i^H \\ Z_i & 0 \end{bmatrix}$ with
 $-Z_i^H = U_i \begin{bmatrix} \arccos \sqrt{\Lambda_i} & \mathbf{0} \end{bmatrix} V_i^H \in \mathbb{C}^{m \times (n-m)}$;
Output: The Riemannian gradient
 $\text{grad } F(X_1 X_1^H) = -[X_1 \ X_2] \mathcal{Z} [X_1 \ X_2]^H$;

Now let N complex subspaces $\{Q_i\} \subset \text{Gr}_{m,n}$ be given, and let $Y_{i1} \in \text{St}_{m,n}$, for $i = 1, \dots, N$, be the respective set of N unitary bases. For a given initialization $P \in \text{Gr}_{m,n}$ with its representation $X \in \text{St}_{m,n}$, we summarize a CG algorithm for computing the Karcher mean of the Q_i 's in Algorithm 3.

Algorithm 3: A CG for computing the Karcher mean on $\text{Gr}_{m,n}$

Input : Stiefel matrices $Y_i \in \text{St}_{m,n}$ for $i = 1, \dots, N$;
Step 1 : Generate an initial guess $[X_1^{(1)} \ X_2^{(1)}] \in \text{U}_n$ and set $i = 1$;
Step 2 : Compute $H^{(1)} = -\text{grad } F(X^{(1)} X^{(1)H})$ using Algorithm 2 ;
Step 3 : Set $i = i + 1$;
Step 4 : Update $[X_1^{(i+1)} \ X_2^{(i+1)}] \leftarrow e^{a[H_i, X_1^{(i)} X_1^{(i)H}]} [X_1^{(i)} \ X_2^{(i)}]$, where a is computed via backtracking line search as in Algorithm 1;
Step 5 : Update $H^{(i+1)} \leftarrow -G^{(i+1)} + r \ G_{H_i}(a)$, where

$$G^{(i+1)} = \text{grad } F(X^{(i+1)} X^{(i+1)H}),$$
and r is chosen according to Eq. (11);
Step 6 : If $i \bmod (2m(n-m)-1) = 0$, set $H^{(i+1)} \leftarrow -G^{(i+1)}$;
Step 7 : If $\|G^{(i+1)}\|$ is small enough, stop. Otherwise, go to Step 3;

As mentioned in Section 3, instead of employing a backtracking line search for selecting an optimal step size at each conjugate direction in Step 4 in Algorithm 3, one can use a one dimensional Newton step instead. Applying this approach to a general Grassmannian requires the calculation of the first and second derivatives of eigenvalues and eigenvectors of a Hermitian matrix valued function $Y_i^H P(t) Y_i$, cf. (44). Unfortunately, this approach is not well defined when the corresponding eigenvalues are multiple.

In the rest of this section, we derive a Newton step size selection as in (7) for the special case where $m = 1$, i.e. the complex projective space $\mathbb{CP}^{n-1} = \text{Gr}_{1,n}$. Let $Q_i = y_i y_i^H \in \mathbb{CP}^{n-1}$, $i = 1, \dots, N$ be a given set of data points with $y_i \in \mathbb{C}^n$

satisfying $\|y_i\| = 1$. We define

$$\lambda_i(t) := y_i^H P(t) y_i, \quad (57)$$

where $P(t)$ is defined by (21), and by abuse of notation we set $\lambda_i = \lambda_i(0)$. The first and second derivatives of the Karcher mean F can be computed as

$$\frac{d}{d\varepsilon}(F \circ P)(\varepsilon)|_{\varepsilon=0} = \text{tr}(\text{grad } F(P(0))\dot{P}(0)), \quad (58)$$

and

$$\begin{aligned} \frac{d^2}{d\varepsilon^2}(F \circ P)(\varepsilon)|_{\varepsilon=0} &= \frac{d^2}{d\varepsilon^2} 2 \sum_{i=1}^n \arccos^2 \sqrt{\lambda_i(\varepsilon)} \Big|_{\varepsilon=0} \\ &= 2 \frac{d}{d\varepsilon} \sum_{i=1}^n \arccos \sqrt{\lambda_i(\varepsilon)} \frac{-\dot{\lambda}_i(\varepsilon)}{\sqrt{\lambda_i(\varepsilon) - \lambda_i(\varepsilon)^2}} \Big|_{\varepsilon=0} \\ &= 2 \sum_{i=1}^n \left(\frac{\dot{\lambda}_i^2}{\lambda_i - \lambda_i^2} - \frac{2(\lambda_i - \lambda_i^2)\ddot{\lambda}_i - \dot{\lambda}_i^2(1 - 2\lambda_i)}{2(\sqrt{\lambda_i - \lambda_i^2})^3} \arccos \sqrt{\lambda_i} \right), \end{aligned} \quad (59)$$

respectively. Finally, a one dimensional Newton step can be computed according to Eq. (7).

7. An application in blind identification

We outline how the above described computation of the Karcher mean can be applied to the problem of *Blind Identification (BI)*. A simple instantaneous BI model assumes that the observation is a linear combination of some unknown sources, i.e.

$$w(t) = As(t), \quad (60)$$

where $A \in \mathbb{C}^{n \times n}$ is the full rank system matrix and $w(t)$ presents n observed linear mixtures of n sources $s(t)$. Blind identification aims to estimate the system matrix A and various algorithms have been developed for this task, cf. [29, 30]. Recall, that the system matrix A can be estimated only up to an arbitrary complex scaling and permutation of the columns, cf. [30]. Thus it is reasonable to consider estimates of columns of A as elements in the complex projective space $\mathbb{CP}^{n-1} := \text{Gr}_{1,n}$. We refer to [19] for an approach on how to include the full rank constraint of A into this setting.

It is known that performance of BI methods are sensitive to the distribution of noise, cf. [31, 32]. In other words, different distributions for the noise might lead to different optimal estimation of the system matrix. In particular, when system noise is present that varies over time, the matrix A can no longer be guaranteed to be estimated correctly by a single process. To overcome this difficulty, an intuitive idea is to simply average over sub-optimal estimations of the system. A similar approach has been investigated in [33], where a Karcher mean based method is proposed for solving the real and whitened ICA problem. Unfortunately, its applications are limited to the cases with stationary signals and noise.

Our experiments employ the following noisy model

$$w_i(t) = (A + \epsilon_i Z_i) s_i(t), \quad i = 1, \dots, N, \quad (61)$$

where A is the ground truth system matrix, $Z_i \in \mathbb{C}^{n \times n}$ models system noise, $\epsilon_i > 0$ represents the noise level, and $s_i(t)$ denotes the unknown signals. Both, real and imaginary parts of all entries of the ground truth system matrix A are drawn from a normal distribution. Perturbations Z_i are applied to the system matrix, where the real and imaginary part of each entry of Z_i are drawn from a uniform distribution on the interval $[-0.5, 0.5]$.

A popular BI algorithm is the so-called Strong Uncorrelated Transform (SUT), cf. [34]. It uses the assumption that the source signals are uncorrelated and non-circular with distinct circularity coefficients. A joint diagonalizer of both the covariance and the pseudo-covariance matrix of the observations serves as an estimation of the inverse of the system matrix A . We employ the SUT for each of the N subproblems to get N estimates of A . After applying a suitable preprocess, we may assume that corresponding columns of the estimates are aligned, cf. [35] for more details, and thus we neglect the permutation ambiguity in our experiments. The overall estimate is given by column wise computing the Karcher mean of the solutions of the sub-problems. Identification performance of the proposed method is measured by the normalized Amari error, cf. [36], defined as

$$J(\hat{A}, A) := \frac{1}{m} \left(\sum_{i=1}^m \frac{\sum_{j=1}^m |b_{ij}|}{\max_j |b_{ij}|} + \sum_{j=1}^m \frac{\sum_{i=1}^m |b_{ij}|}{\max_i |b_{ij}|} \right) - 2, \quad (62)$$

where \hat{A} is an estimation of A , and $B = (b_{ij})_{i,j=1}^m = \hat{A}^{-1}A$. In general, the smaller the Amari error, the better the identification of the system matrix.

In our experiments, we compare the proposed Karcher mean method to a simple standard approach, referred to as the Euclidean mean approach. Thereby, all solutions produced by SUT are summed up and the columns of the obtained matrix are normalized to have unit norm. First of all, we investigate the performance of both methods against the number of estimations N . We fix $n = 5$ and run $N = 100$ experiments per number of estimations. As the box plots of Amari errors in Figure 3 and Figure 4 suggest, both Karcher and Euclidean subspace averaging methods admit a consistently increasing performance with an increasing number of estimations. In our second experiment, we choose a various number of noise levels $\epsilon_i \in \{1, 0.5, 0.2, 0.1, 0.01\}$ and fix the number of estimations $N = 10$. As shown in Figure 5, the Karcher mean approach outperforms the Euclidean counterpart consistently and its advantage is considerably higher, the more noise is present.

8. Conclusion

This work focuses on the problem of averaging complex subspaces of equal dimension by computing their Karcher mean, which is under mild assumptions the

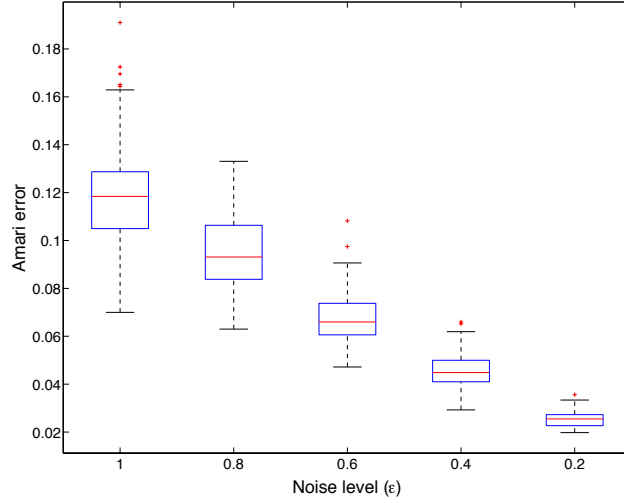


Figure 3: Performance of subspace averaging via Karcher mean.

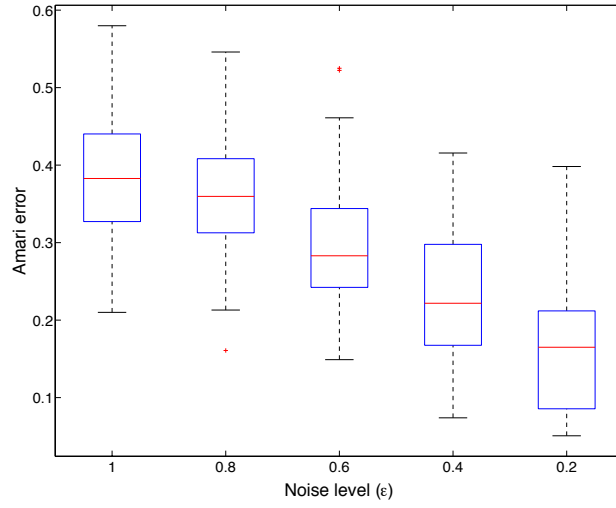


Figure 4: Performance of subspace averaging via Euclidean mean.

unique minimum of a well defined smooth function on the complex Grassmanian. An accessible introduction to the geometric structure of the Grassmanian is provided by its identification with the set of Hermitian projectors. In particular, explicit formulas for geodesics, parallel transport, and the Riemannian gradient of the Karcher mean function are given, which, in contrast to other formulas available in the literature, are well suited for implementation.

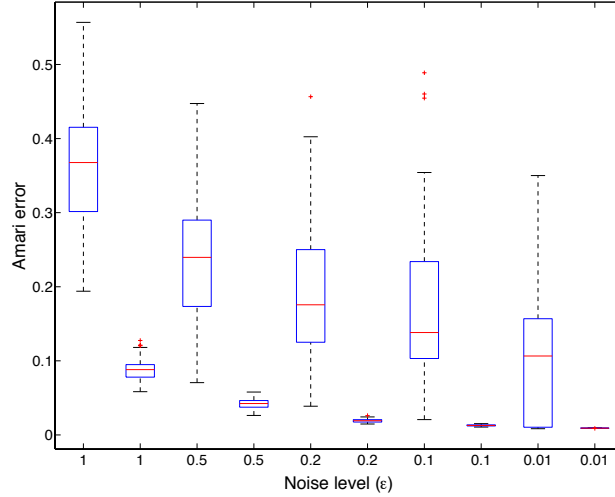


Figure 5: Comparison of performance (left = Euclidean, right = Karcher).

These results are used to propose an intrinsic conjugate gradient algorithm on the Grassmannian for computing the Karcher mean. We present experiments and outline the usability of such an approach in blind identification.

References

- [1] A. Srivastava, E. Klassen, Bayesian and geometric subspace tracking, *Advances in Applied Probability* 36 (2004) 43–56.
- [2] T. Adalı, S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*, Adaptive and Learning Systems for Signal Processing, Communications and Control, Wiley-IEEE Press, 2010.
- [3] P. Turaga, A. Veeraraghavan, A. Srivastava, R. Chellappa, Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (2011) 2273–2286.
- [4] L. Simonsson, L. Eldén, Grassmann algorithms for low rank approximation of matrices with missing values, *BIT* 50 (2010) 173–191.
- [5] H. Karcher, Riemannian center of mass and mollifier smoothing, *Commun. Pure Appl. Math.* 30 (1977) 509–541.
- [6] W. S. Kendall, Probability, convexity, and harmonic maps with small image. I: Uniqueness and fine existence, *Proc. Lond. Math. Soc., III. Ser.* 61 (1990) 371–406.

- [7] H. Le, Locating Fréchet means with application to shape spaces, *Adv. Appl. Probab.* 33 (2001) 324–338.
- [8] C. Gramkow, On averaging rotations, *J. Math. Im. Vis.* 15 (2001) 7–16.
- [9] T. Ando, C.-K. Li, R. Mathias, Geometric means, *Linear Algebra Appl.* 385 (2004) 305–334.
- [10] S. R. Buss, J. P. Fillmore, Spherical averages and applications to spherical splines and interpolation, *ACM Trans. Graph.* 20 (2001) 95–126.
- [11] W. S. Kendall, Convexity and the hemisphere, *J. Lond. Math. Soc., II. Ser.* 43 (1991) 567–576.
- [12] J. M. Corcuera, W. S. Kendall, Riemannian barycentres and geodesic convexity, *Math. Proc. Camb. Philos. Soc.* 127 (1999) 253–269.
- [13] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, Princeton, NJ, 2008.
- [14] U. Helmke, K. Hüper, J. Trumpf, Newton’s method on Grassmann manifolds, 2007. ArXiv:0709.2205v2.
- [15] D. W. Dreisigmeyer, Direct Search Algorithms over Riemannian Manifolds, Technical Report, Los Alamos National Laboratory, USA, 2006.
- [16] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, NJ, 2008.
- [17] M. Kleinstenber, K. Hüper, An intrinsic CG algorithm for computing dominant subspaces, in: *Proc. 32nd IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Hawaii, USA, pp. IV1405–IV1408.
- [18] S. Mittal, P. Meer, Conjugate gradient on Grassmann manifolds for robust subspace estimation, *Image and Vision Computing* (2012).
- [19] H. Shen, M. Kleinstenber, Complex blind source separation via simultaneous strong uncorrelating transform, in: *Latent Variable Analysis and Signal Separation*, volume 6365 of *LNCS*, Springer, 2010, pp. 287–294.
- [20] M. P. do Carmo, Riemannian geometry, *Mathematics: Theory & Applications*. Boston, MA: Birkhäuser, 1992.
- [21] M. Kleinstenber, A. Seghouane, A CG-type method for computing the dominant subspace of a symmetric matrix, 2010. *Proc. of the 9th Portuguese Conf. on Automatic Control (CONTROLO’2010)*. CD-ROM.
- [22] J. Nocedal, S. J. Wright, Numerical optimization, Springer, 1999.
- [23] S. Smith, Optimization techniques on Riemannian manifolds, in: A. Bloch (Ed.), *Hamiltonian and gradient flows, algorithms and control*, AMS, Providence, 1994, pp. 113–136.

- [24] D. H. Husemoller, *Fibre bundles*. 3rd ed., Berlin: Springer-Verlag, 1993.
- [25] A. L. Onishchik, *Topology of transitive transformation groups*, Leipzig: Johann Ambrosius Barth, 1994.
- [26] K. Hüper, F. Silva Leite, On the geometry of rolling and interpolation curves on S^n , SO_n , and Grassmann manifolds., *J. Dyn. Control Syst.* 13 (2007) 467–502.
- [27] T. Sakai, *Riemannian geometry*, Translations of Mathematical Monographs. 149. Providence, RI: American Mathematical Society, 1996.
- [28] G. Golub, C. F. Van Loan, *Matrix computations*. 3rd ed., Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
- [29] A. Cichocki, S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, Wiley, 2002.
- [30] P. Comon, C. Jutten (Eds.), *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic, 2010.
- [31] J.-F. Cardoso, A. Souloumiac, Blind beamforming for non Gaussian signals, *The IEE Proceedings of F* 140 (1993) 363–370.
- [32] M. Wax, J. Sheinvald, A least-squares approach to joint diagonalization, *IEEE Signal Processing Letters* 4 (1997) 52–53.
- [33] J. Manton, A centroid (Karcher mean) approach to the joint approximate diagonalisation problem: The real symmetric case, *Digital Signal Processing* 16 (2006) 468–478.
- [34] J. Eriksson, V. Koivunen, Complex-valued ICA using second order statistics, in: *Proceedings of the 14th IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2004)*, pp. 183–191.
- [35] S. Makino, T.-W. Lee, H. Sawada (Eds.), *Blind Speech Separation*, Signals and Communication Technology, Springer Netherlands, 2007.
- [36] S. Amari, A. Cichocki, H. H. Yang, A new learning algorithm for blind signal separation, *Advances in Neural Information Processing Systems* 8 (1996) 757–763.