

Riemannian geometry of Grassmann manifolds with a view on algorithmic computation

P.-A. Absil*

R. Mahony†

R. Sepulchre‡

Last revised: 14 Dec 2003

PREPRINT

Abstract. We give simple formulas for the canonical metric, gradient, Lie derivative, Riemannian connection, parallel translation, geodesics and distance on the Grassmann manifold of p -planes in \mathbb{R}^n . In these formulas, p -planes are represented as the column space of $n \times p$ matrices. The Newton method on abstract Riemannian manifolds proposed by S. T. Smith is made explicit on the Grassmann manifold. Two applications –computing an invariant subspace of a matrix and the mean of subspaces– are worked out.

Key words. Grassmann manifold, noncompact Stiefel manifold, principal fiber bundle, Levi-Civita connection, parallel transportation, geodesic, Newton method, invariant subspace, mean of subspaces.

AMS subject classification: 65J05 General theory (Numerical analysis in abstract spaces); 53C05 Connections, general theory; 14M15 Grassmannians, Schubert varieties, flag manifolds.

1 Introduction

The majority of available numerical techniques for optimization and nonlinear equations assume an underlying Euclidean space. Yet many computational problems are posed on non-Euclidean spaces. Several authors [Gab82, Smi94, Udr94, Mah96, MM02] have proposed abstract algorithms that exploit the underlying geometry (e.g. symmetric, homogeneous, Riemannian) of manifolds on which problems are cast, but the conversion of these abstract geometric algorithms into numerical procedures in practical situations is often a nontrivial task that critically relies on an adequate representation of the manifold.

The present paper contributes to addressing this issue in the case where the relevant non-Euclidean space is the set of fixed dimensional subspaces of a given Euclidean space. This non-Euclidean space is commonly called the Grassmann manifold. Our motivation for

*School of Computational Science and Information Technology, Florida State University, Tallahassee FL 32306-4120. Part of this work was done while the author was a Research Fellow with the Belgian National Fund for Scientific Research (Aspirant du F.N.R.S.) at the University of Liège. URL: www.montefiore.ulg.ac.be/~absil

†Department of Engineering, Australian National University, ACT, 0200, Australia.

‡Department of Electrical Engineering and Computer Science, Université de Liège, Bât. B28 Systèmes, Grande Traverse 10, B-4000 Liège, Belgium. URL: www.montefiore.ulg.ac.be/systems

considering the Grassmann manifold comes from the number of applications that can be formulated as finding zeros of fields defined on the Grassmann manifold. Examples include invariant subspace computation and subspace tracking; see e.g. [Dem87, CG90] and references therein.

A simple and robust manner of representing a subspace in computer memory is in the form of a matrix array of double precision data whose columns span the subspace. Using this representation technique, we produce formulas for fundamental Riemannian-geometric objects on the Grassmann manifold endowed with its canonical metric: gradient, Riemannian connection, parallel translation, geodesics and distance. The formulas for the Riemannian connection and geodesics directly yield a matrix expression for a Newton method on Grassmann, and we illustrate the applicability of this Newton method on two computational problems cast on the Grassmann manifold.

The classical Newton method for computing a zero of a function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ can be formulated as follows [DS83, Lue69]: Solve the Newton equation

$$DF(x)[\eta] = -F(x) \quad (1)$$

for the unknown $\eta \in \mathbb{R}^n$ and compute the update

$$x_+ := x + \eta. \quad (2)$$

When F is defined on a non-Euclidean manifold, a possible approach is to choose local coordinates and use the Newton method as in (1)-(2). However, the successive iterates on the manifold will depend on the chosen coordinate system. Smith [Smi93, Smi94] proposes a coordinate-independent Newton method for computing a zero of a C^∞ one-form μ on an abstract complete Riemannian manifold M . He suggests to solve the Newton equation

$$\nabla_\eta \mu = -\mu_x \quad (3)$$

for the unknown $\eta \in T_x M$, where ∇ denotes the Riemannian connection (also called **Levi-Civita connection**) on M , and update along the geodesic as $x_+ := \text{Exp}_x \eta$. It can be proven that if x is chosen suitably close to a point \hat{x} in M such that $\mu_{\hat{x}} = 0$ and $T_{\hat{x}} M \ni \eta \mapsto \nabla_\eta \mu$ is nondegenerate, then the algorithm converges quadratically to \hat{x} . We will refer to this iteration as the *Riemann-Newton method*.

In practical cases it may not be obvious to particularize the Riemann-Newton method into a concrete algorithm. Given a Riemannian manifold M and an initial point x on M , one may pick a coordinate system containing x , compute the metric tensor in these coordinates, deduce the Christoffel symbols and obtain a tensorial equation for (3), but this procedure is often exceedingly complicated and computationally inefficient. One can also recognize that the Riemann-Newton method is equivalent to the classical Newton method in normal coordinates at x [MM02], but obtaining a tractable expression for these coordinates is often elusive.

On the Grassmann manifold, a formula for the Riemannian connection was given by Machado and Salavessa in [MS85]. They identify the Grassmann manifold with the set of projectors into subspaces of \mathbb{R}^n , embed the set of projectors in the set of linear maps from \mathbb{R}^n to \mathbb{R}^n (which is an Euclidean space), and endow this set with the Hilbert-Schmidt inner product. The induced metric on the Grassmann manifold is then the essentially unique O_n -invariant metric mentioned above. The embedding of the Grassmann manifold in an Euclidean space

allows the authors to compute the Riemannian connection by taking the derivative in the Euclidean space and projecting the result into the tangent space of the embedded manifold. They obtain a formula for the Riemannian connection in terms of projectors.

Edelman, Arias and Smith [EAS98] have proposed an expression of the Riemann-Newton method on the Grassmann manifold in the particular case where μ is the differential df of a real function f on M . Their approach avoids the derivation of a formula for the Riemannian connection on Grassmann. Instead, they obtain a formula for the Hessian $(\nabla_{\Delta_1} df)\Delta_2$ by polarizing the second derivative of f along the geodesics.

In the present paper, we derive an easy-to-use formula for the Riemannian connection $\nabla_\eta \xi$ where η and ξ are *arbitrary* smooth vector fields on the Grassmann manifold of p -dimensional subspaces of \mathbb{R}^n . This formula, expressed in terms of $n \times p$ matrices, intuitively relates to the geometry of the Grassmann manifold expressed as a set of equivalence classes of $n \times p$ matrices. Once the formula for Riemannian connection is available, expressions for parallel transport and geodesics directly follow. Expressing the Riemann-Newton method on the Grassmann manifold for concrete vector fields ξ reduces to a directional derivative in \mathbb{R}^n followed by a projection.

We work out an example where the zeros of ξ are the p -dimensional right invariant subspaces of an arbitrary $n \times n$ matrix A . This generalizes an application considered in [EAS98] where ξ was the gradient of a generalized scalar Rayleigh quotient of a matrix $A = A^T$. The Newton method for our ξ converges locally quadratically to the nondegenerate zeros of ξ . We show that the rate of convergence is cubic if and only if the targeted zero of ξ is also a left invariant subspace of A . In a second example, the zero of ξ is the mean of a collection of p -dimensional subspaces of \mathbb{R}^n . We illustrate on a numerical experiment the fast convergence of the Newton algorithm to the mean subspace.

The present paper only requires from the reader an elementary background in Riemannian geometry (tangent vectors, gradient, parallel transport, geodesics, distance), which can be read e.g. from Boothby [Boo75], do Carmo [dC92] or the introductory chapter of [Cha93]. The relevant definitions are summarily recalled in the text. Concepts of reductive homogeneous space and symmetric spaces (see [Boo75, Nom54, KN63, Hel78] and particularly sections II.4, IV.3, IV.A and X.2 in the latter) are not needed, but they can help to get insight into the problem. Although some elementary concepts of principal fiber bundle theory [KN63] are used, no specific background is needed.

The paper is organized as follows. In Section 2, the linear subspaces of \mathbb{R}^n are identified with equivalent classes of matrices and the manifold structure of Grassmann is defined. Section 3 defines a Riemannian structure on Grassmann. Formulas are given for Lie brackets, Riemannian connection, parallel transport, geodesics and distance between subspaces. The Grassmann-Newton algorithm is made explicit in Section 4 and practical applications are worked out in details in Section 5.

2 The Grassmann manifold

The goal of this section is to recall relevant facts about the Grassmann manifolds. More details can be read from [Won67, Boo75, DM90, HM94, FGP94].

Let n be a positive integer and let p be a positive integer not greater than n . The set of p -dimensional linear subspaces of \mathbb{R}^n (“linear” will be omitted in the sequel) is termed the *Grassmann manifold*, denoted here by $\text{Grass}(p, n)$.

An element \mathcal{Y} of $\text{Grass}(p, n)$, i.e. a p -dimensional subspace of \mathbb{R}^n , can be specified by a basis, i.e. a set of p vectors y_1, \dots, y_p such that \mathcal{Y} is the set of all their linear combinations. When the y 's are ordered as the columns of an n -by- p matrix Y , then Y is said to *span* \mathcal{Y} and \mathcal{Y} is said to be the *column space* (or range, or image, or span) of Y , and we write $\mathcal{Y} = \text{span}(Y)$. The span of an n -by- p matrix Y is an element of $\text{Grass}(p, n)$ if and only if Y has full rank. The set of such matrices is termed the *noncompact Stiefel manifold*¹

$$\text{ST}(p, n) := \{Y \in \mathbb{R}^{n \times p} : \text{rank}(Y) = p\}.$$

Given $\mathcal{Y} \in \text{Grass}(p, n)$, the choice of a Y in $\text{ST}(p, n)$ such that Y spans \mathcal{Y} is not unique. There are infinitely many possibilities. Given a matrix Y in $\text{ST}(p, n)$, the set of the matrices in $\text{ST}(p, n)$ that have the same span as Y is

$$Y \text{GL}_p := \{YM : M \in \text{GL}_p\} \quad (4)$$

where GL_p denotes the set of the p -by- p invertible matrices. This identifies $\text{Grass}(p, n)$ with the quotient space $\text{ST}(p, n)/\text{GL}_p := \{Y \text{GL}_p : Y \in \text{ST}(p, n)\}$. In fiber bundle theory, the quadruple $(\text{GL}_p, \text{ST}(p, n), \pi, \text{Grass}(p, n))$ is called a *principal GL_p fiber bundle*, with total space $\text{ST}(p, n)$, base space $\text{Grass}(p, n) = \text{ST}(p, n)/\text{GL}_p$, group action

$$\text{ST}(p, n) \times \text{GL}_p \ni (Y, M) \mapsto YM \in \text{ST}(p, n)$$

and projection map

$$\pi : \text{ST}(p, n) \ni Y \mapsto \text{span}(Y) \in \text{Grass}(p, n).$$

See e.g. [KN63] for the general theory of principal *fiber bundles* and [FGP94] for a detailed treatment of the Grassmann case. In this paper, we use the notation $\text{span}(Y)$ and $\pi(Y)$ to denote the column space of Y .

To each subspace \mathcal{Y} corresponds an equivalence class (4) of n -by- p matrices that span \mathcal{Y} , and each equivalence class contains infinitely many elements. It is however possible to locally single out a unique matrix in (almost) each equivalence class, by means of *cross sections*. Here we will consider affine cross sections, which are defined as follows (see illustration on Figure 1). Let $W \in \text{ST}(p, n)$. The matrix W defines an affine cross section

$$S_W := \{Y \in \text{ST}(p, n) : W^T(Y - W) = 0\} \quad (5)$$

orthogonal to the fiber $W \text{GL}_p$. Let $Y \in \text{ST}(p, n)$. If $W^T Y$ is invertible, then the equivalence class $Y \text{GL}_p$ (i.e. the set of matrices with the same span as Y) intersects the cross section S_W at the single point $Y(W^T Y)^{-1} W^T W$. If $W^T Y$ is not invertible, which means that the span of W contains an orthogonal direction to the span of Y , then the intersection between the fiber $W \text{GL}_p$ and the section S_W is empty. Let

$$\mathcal{U}_W := \{\text{span}(Y) : W^T Y \text{ is invertible}\} \quad (6)$$

be the set of subspaces whose representing fiber $Y \text{GL}_p$ intersects the section S_W . The mapping

$$\sigma_W : \mathcal{U}_W \ni \text{span}(Y) \mapsto Y(W^T Y)^{-1} W^T W \in S_W, \quad (7)$$

¹The (compact) Stiefel manifold is the set of *orthonormal* $n \times p$ matrices.

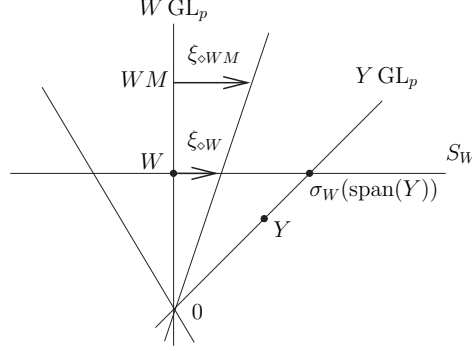


Figure 1: This is an illustration of $\text{Grass}(p, n)$ as the quotient $\text{ST}(p, n)/\text{GL}_p$ for the case $p = 1$, $n = 2$. Each point, the origin excepted, is an element of $\text{ST}(p, n) = \mathbb{R}^2 - \{0\}$. Each line is an equivalence class of elements of $\text{ST}(p, n)$ that have the same span. So each line corresponds to an element of $\text{Grass}(p, n)$. The affine subspace S_W is an affine cross section as defined in (5). The relation (10) satisfied by the horizontal lift ξ_{\diamond} of a tangent vector $\xi \in T_W \text{Grass}(p, n)$ is also illustrated. This picture can help to get insight into the general case. One has nonetheless to be careful when drawing conclusions from this picture. For example, in general there does not exist a submanifold of $\mathbb{R}^{n \times p}$ that is orthogonal to the fibers $Y \text{GL}_p$ at each point, although it is obviously the case when $p = 1$ (any centered sphere in \mathbb{R}^n will do).

which we will call *cross section mapping*, realizes a bijection between the subset \mathcal{U}_W of $\text{Grass}(p, n)$ and the affine subspace S_W of $\text{ST}(p, n)$. The classical manifold structure of $\text{Grass}(p, n)$ is the one that, for all $W \in \text{ST}(p, n)$, makes σ_W a diffeomorphism between \mathcal{U}_W and S_W (embedded in the Euclidean space $\mathbb{R}^{n \times p}$) [FGP94]. Parameterizations of $\text{Grass}(p, n)$ are then given by

$$\mathbb{R}^{(n-p) \times p} \ni K \mapsto \pi(W + W_{\perp} K) = \text{span}(W + W_{\perp} K) \in \mathcal{U}_W,$$

where W_{\perp} is any element of $\text{ST}(n - p, n)$ such that $W^T W_{\perp} = 0$.

3 Riemannian structure on $\text{Grass}(p, n) = \text{ST}(p, n)/\text{GL}_p$

The goal of this section is to define a Riemannian metric on $\text{Grass}(p, n)$ and then derive formulas for the associated gradient, connection and geodesics. For an introduction to Riemannian geometry, see e.g. [Boo75], [dC92] or the introductory chapter of [Cha93].

Tangent vectors

A tangent vector ξ of $\text{Grass}(p, n)$ at \mathcal{W} can be thought of as an elementary variation of the p -dimensional subspace \mathcal{W} (see [Boo75, dC92] for a more formal definition of a tangent vector). Here we give a way to represent ξ by a matrix. The principle is to decompose variations of a basis W of \mathcal{W} into a component that does not modify the span and a component that does modify the span. The latter represents a tangent vector of $\text{Grass}(p, n)$ at \mathcal{W} .

Let $W \in \text{ST}(p, n)$. The tangent space to $\text{ST}(p, n)$ at W , denoted $T_W \text{ST}(p, n)$, is trivial: $\text{ST}(p, n)$ is an open subset of $\mathbb{R}^{n \times p}$, so $\text{ST}(p, n)$ and $\mathbb{R}^{n \times p}$ are identical in a neighbourhood

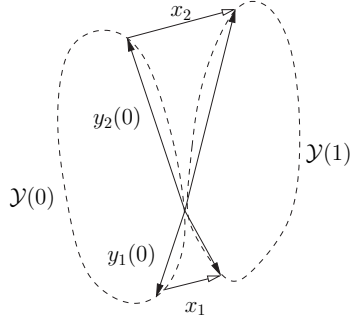


Figure 2: This picture illustrates for the case $p = 2$, $n = 3$ how $\xi_{\diamond Y}$ represents an “elementary variation” ξ of the subspace \mathcal{Y} spanned by the columns of Y . Consider $Y(0) = [y_1(0)|y_2(0)]$ and $\xi_{\diamond Y(0)} = [x_1|x_2]$. By the horizontality condition $Y(0)^T \xi_{\diamond Y(0)} = 0$, both x_1 and x_2 are normal to the space $\mathcal{Y}(0)$ spanned by $Y(0)$. Let $y_1(t) = y_1(0) + tx_1$, $y_2(t) = y_2(0) + tx_2$ and let $\mathcal{Y}(t)$ be the subspace spanned by $y_1(t)$ and $y_2(t)$. Then we have $\xi = \dot{\mathcal{Y}}(0)$.

of W , and therefore $T_W \text{ST}(p, n) = T_W \mathbb{R}^{n \times p}$ which is just a copy of $\mathbb{R}^{n \times p}$. The *vertical space* V_W is by definition the tangent space to the fiber $W \text{GL}_p$, namely

$$V_W = W \mathbb{R}^{p \times p} = \{Wm : m \in \mathbb{R}^{p \times p}\}.$$

Its elements are the elementary variations of W that do not modify its span. We define the *horizontal space* H_W as

$$H_W := T_W S_W = \{W_{\perp} K : K \in \mathbb{R}^{(n-p) \times p}\}. \quad (8)$$

One readily verifies that H_W verifies the characteristic properties of horizontal spaces in principal fiber bundles [KN63, FGP94]. In particular, $T_W \text{ST}(p, n) = V_W \oplus H_W$. Note that with our choice of H_W , $\Delta_V^T \Delta_H = 0$ for all $\Delta_V \in V_W$ and $\Delta_H \in H_W$.

Let ξ be a tangent vector to $\text{Grass}(p, n)$ at \mathcal{W} and let W span \mathcal{W} . According to the theory of principal fiber bundles [KN63], there exists one and only one horizontal vector $\xi_{\diamond W}$ that represents ξ in the sense that $\xi_{\diamond W}$ projects to ξ via the span operation, i.e. $d\pi(W) \xi_{\diamond W} = \xi$. See Figure 1 for a graphical interpretation. It is easy to check that

$$\xi_{\diamond W} = d\sigma_W(\mathcal{W}) \xi \quad (9)$$

where σ_W is the cross section mapping defined in (7). Indeed, it is horizontal and projects to ξ via π since $\pi \circ \sigma_W$ is locally the identity. The representation $\xi_{\diamond W}$ is called the *horizontal lift* of $\xi \in T_{\mathcal{W}} \text{Grass}(p, n)$ at W . The next proposition characterizes how the horizontal lift varies along the equivalence class $W \text{GL}_p$.

Proposition 3.1 *Let $\mathcal{W} \in \text{Grass}(p, n)$, let W span \mathcal{W} and let $\xi \in T_{\mathcal{W}} \text{Grass}(p, n)$. Let $\xi_{\diamond W}$ denote the horizontal lift of ξ at W . Then for all $M \in \text{GL}_p$,*

$$\xi_{\diamond WM} = \xi_{\diamond W} M. \quad (10)$$

Proof. This comes from (9) and the property $\sigma_{WM}(\mathcal{Y}) = \sigma_W(\mathcal{Y}) M$. \square

The homogeneity property (10) and the horizontality of $\xi_{\diamond W}$ are characteristic of horizontal lifts.

We now introduce notation for derivatives. Let f be a smooth function between two linear spaces. We denote by

$$Df(x)[y] := \frac{d}{dt}f(x + ty)|_{t=0}$$

the directional derivative of f at x in the direction of y . Let f be a smooth real-valued function defined on $\text{Grass}(p, n)$ in a neighbourhood of \mathcal{W} . We will use the notation $f_{\diamond}(W)$ to denote $f(\text{span}(W))$. The derivative of f in the direction of the tangent vector ξ at \mathcal{W} , denoted by ξf , can be computed as

$$\xi f = Df_{\diamond}(W)[\xi_{\diamond}W]$$

where W spans \mathcal{W} .

Lie derivative

A tangent vector field ξ on $\text{Grass}(p, n)$ assigns to each $\mathcal{Y} \in \text{Grass}(p, n)$ an element $\xi_{\mathcal{Y}} \in T_{\mathcal{Y}}\text{Grass}(p, n)$.

Proposition 3.2 (Lie bracket) *Let η and ξ be smooth tangent vector fields on $\text{Grass}(p, n)$. Let $\xi_{\diamond}W$ denote the horizontal lift of ξ at W as defined in (9). Then*

$$[\eta, \xi]_{\diamond}W = \Pi_{W^{\perp}}[\eta_{\diamond}W, \xi_{\diamond}W] \quad (11)$$

where

$$\Pi_{W^{\perp}} := I - W(W^TW)^{-1}W^T \quad (12)$$

denotes the projection into the orthogonal complement of the span of W and

$$[\eta_{\diamond}W, \xi_{\diamond}W] = D\xi_{\diamond}(W)[\eta_{\diamond}W] - D\eta_{\diamond}(W)[\xi_{\diamond}W]$$

denotes the Lie bracket in $\mathbb{R}^{n \times p}$.

That is, the horizontal lift of the Lie bracket of two tangents vector fields on the Grassmann manifold is equal to the horizontal projection of the Lie bracket of the horizontal lifts of the two tangent vector fields.

Proof. Let $W \in \text{ST}(p, n)$ be fixed. We prove formula (11) by making computations in the coordinate chart $(\mathcal{U}_W, \sigma_W)$. In order to simplify notations, let $\hat{Y} := \sigma_W \mathcal{Y}$ and $\hat{\xi}_{\hat{Y}} := \sigma_{W*\mathcal{Y}}\xi_{\mathcal{Y}}$. Note that $\hat{W} = W$ and $\hat{\xi}_W = \xi_{\diamond}W$. One has

$$[\eta, \xi]_{\diamond}W = D\hat{\xi}(W)[\hat{\eta}_W] - D\hat{\eta}(W)[\hat{\xi}_W].$$

After some manipulations using (5) and (7), it comes

$$\hat{\xi}_{\hat{Y}} = \frac{d}{dt}\sigma_W[\hat{Y} + \xi_{\diamond\hat{Y}}t]|_{t=0} = \xi_{\diamond\hat{Y}} - \hat{Y}(W^TW)^{-1}W^T\xi_{\diamond\hat{Y}}.$$

Then, using $W^T\xi_{\diamond}W = 0$,

$$D\hat{\xi}(W)[\hat{\eta}_W] = D\xi_{\diamond}(W)[\hat{\eta}_W] - W(W^TW)^{-1}W^TD\xi_{\diamond}(W)[\hat{\eta}_W].$$

The term $D\hat{\eta}(W)[\hat{\xi}_W]$ is directly deduced by interchanging ξ and η , and the result is proved.

□

Metric

We consider the following metric on $\text{Grass}(p, n)$:

$$\langle \xi, \eta \rangle_{\mathcal{Y}} := \text{trace} \left((Y^T Y)^{-1} \xi_{\diamond Y}^T \eta_{\diamond Y} \right) \quad (13)$$

where Y spans \mathcal{Y} . It is easily checked that the expression (13) does not depend on the choice of the basis Y that spans \mathcal{Y} . This metric is the only one (up to multiplications by a constant) to be invariant under the action of O_n on \mathbb{R}^n . Indeed

$$\text{trace} \left(((QY)^T QY)^{-1} (Q\xi_{\diamond Y})^T Q\eta_{\diamond Y} \right) = \text{trace} \left((Y^T Y)^{-1} \xi_{\diamond Y}^T \eta_{\diamond Y} \right)$$

for all $Q \in O_n$, and uniqueness is proved in [Lei61]. We will see later that the definition (13) induces a natural notion of *distance* between subspaces.

Gradient

On an abstract Riemannian manifold M , the gradient of a smooth real function f at a point x of M , denoted by $\text{grad}f(x)$, is roughly speaking the steepest ascent vector of f in the sense of the Riemannian metric. More rigorously, $\text{grad}f(x)$ is the element of $T_x M$ satisfying $\langle \text{grad}f(x), \xi \rangle = \xi f$ for all $\xi \in T_x M$. On the Grassmann manifold $\text{Grass}(p, n)$ endowed with the metric (13), one checks that

$$\boxed{(\text{grad}f)_{\diamond Y} = \Pi_{Y^\perp} \text{grad}f_{\diamond}(Y) Y^T Y} \quad (14)$$

where Π_{Y^\perp} is the orthogonal projection (12) into the orthogonal complement of Y , $f_{\diamond}(Y) = f(\text{span}(Y))$ and $\text{grad}f_{\diamond}(Y)$ is the Euclidean gradient of f_{\diamond} at Y , given by $(\text{grad}f_{\diamond}(Y))_{ij} = \frac{\partial f_{\diamond}(Y)}{\partial Y_{ij}}(Y)$. The Euclidean gradient is characterized by

$$Df_{\diamond}(Y)[Z] = \text{trace}(Z^T \text{grad}f_{\diamond}(Y)), \quad \forall Z \in \mathbb{R}^{n \times p}, \quad (15)$$

which can ease its computation in some cases.

Riemannian connection

Let ξ, η be two tangent vector fields on $\text{Grass}(p, n)$. There is no predefined way of computing the derivative of ξ in the direction of η because there is no predefined way of comparing the different tangent spaces $T_{\mathcal{Y}} \text{Grass}(p, n)$ as \mathcal{Y} varies. However, there is a preferred definition for the directional derivative, called the *Riemannian connection* (or *Levi-Civita connection*), defined as follows [Boo75, dC92].

Definition 3.3 (Riemannian connection) *Let M be a Riemannian manifold and let its metric be denoted by $\langle \cdot, \cdot \rangle$. Let $x \in M$. The Riemannian connection ∇ on M has the following properties: For all smooth real functions f, g on M , all η, η' in $T_x M$ and all smooth vectors fields ξ, ξ' :*

1. $\nabla_{f\eta + g\eta'} \xi = f \nabla_{\eta} \xi + g \nabla_{\eta'} \xi$
2. $\nabla_{\eta}(f\xi + g\xi') = f \nabla_{\eta} \xi + g \nabla_{\eta} \xi' + (\eta f) \xi + (\eta g) \xi'$
3. $[\xi, \xi'] = \nabla_{\xi} \xi' - \nabla_{\xi'} \xi$.
4. $\eta \langle \xi, \xi' \rangle = \langle \nabla_{\eta} \xi, \xi' \rangle + \langle \xi, \nabla_{\eta} \xi' \rangle$.

Properties 1 and 2 define connections in general. Property 3 states that the connection is torsion-free, and property 4 specifies that the metric tensor is invariant by the connection. A famous theorem of Riemannian geometry states that there is one and only one connection verifying these four properties. If M is a submanifold of an Euclidean space, then the Riemannian connection $\nabla_\eta \xi$ consists in taking the derivative of ξ in the ambient Euclidean space in the direction of η and projecting the result into the tangent space of the manifold. As we show in the next theorem, the Riemannian connection on the Grassmann manifold, expressed in terms of horizontal lifts, works in a similar way.

Theorem 3.4 (Riemannian connection) *Let $\mathcal{Y} \in \text{Grass}(p, n)$, let $Y \in \text{ST}(p, n)$ span \mathcal{Y} . Let $\eta \in T_Y \text{Grass}(p, n)$, and let ξ be a smooth tangent vector field defined in a neighbourhood of \mathcal{Y} . Let $\xi_\diamond : W \mapsto \xi_{\diamond W}$ be the horizontal lift of ξ as defined in (9). Let $\text{Grass}(p, n)$ be endowed with the O_n -invariant Riemannian metric (13) and let ∇ denote the associated Riemannian connection. Then*

$$\boxed{(\nabla_\eta \xi)_{\diamond Y} = \Pi_{Y_\perp} \nabla_{\eta_{\diamond Y}} \xi_\diamond} \quad (16)$$

where Π_{Y_\perp} is the projection (12) into the orthogonal complement of \mathcal{Y} and

$$\nabla_{\eta_{\diamond Y}} \xi_\diamond := D\xi_\diamond(Y)[\eta_{\diamond Y}] := \left. \frac{d}{dt} \xi_\diamond(Y + \eta_{\diamond Y} t) \right|_{t=0}$$

is the directional derivative of ξ_\diamond in the direction of $\eta_{\diamond Y}$ in the Euclidean space $\mathbb{R}^{n \times p}$.

This theorem says that the horizontal lift of the covariant derivative of a vector field ξ on the Grassmannian in the direction of η is equal to the horizontal projection of the derivative of the horizontal lift of ξ in the direction of the horizontal lift of η .

Proof. One has to prove that (16) satisfies the four characteristic properties of the Riemannian connection. The two first properties concern linearity in η and ξ and are easily checked. The torsion-free property is direct from (16) and (11). The fourth property, invariance of the metric, holds for (16) since

$$\begin{aligned} \eta \langle \mu, \nu \rangle &= D_Y \text{trace}((Y^T Y)^{-1} \mu_{\diamond Y}^T \nu_{\diamond Y})(W)[\eta_{\diamond W}] \\ &= \text{trace}((W^T W)^{-1} D\mu_{\diamond \cdot}^T(W)[\eta_{\diamond W}] \nu_{\diamond W} + \mu_{\diamond W}^T D\nu_{\diamond \cdot}(W)[\eta_{\diamond W}]) \\ &= \langle \nabla_\eta \mu, \nu \rangle + \langle \mu, \nabla_\eta \nu \rangle \end{aligned}$$

□

Parallel transport

Let $t \mapsto \mathcal{Y}(t)$ be a smooth curve on $\text{Grass}(p, n)$. Let ξ be a tangent vector defined along the curve $\mathcal{Y}(\cdot)$. Then ξ is said to be *parallel transported* along $\mathcal{Y}(\cdot)$ if

$$\nabla_{\dot{\mathcal{Y}}(t)} \xi = 0 \quad (17)$$

for all t , where $\dot{\mathcal{Y}}(t)$ denotes the tangent vector to $\mathcal{Y}(\cdot)$ at t .

We will need the following classical result of fiber bundle theory [KN63]. A curve $t \mapsto Y(t)$ on $\text{ST}(p, n)$ is termed *horizontal* if $\dot{Y}(t)$ is horizontal for all t , i.e. $\dot{Y}(t) \in H_{Y(t)}$. Let $t \mapsto \mathcal{Y}(t)$ be a smooth curve on $\text{Grass}(p, n)$ and let $Y_0 \in \text{ST}(p, n)$ span $\mathcal{Y}(0)$. Then there exists a unique horizontal curve $t \mapsto Y(t)$ on $\text{ST}(p, n)$ such that $Y(0) = Y_0$ and $\mathcal{Y}(t) = \text{span}(Y(t))$. The curve $Y(0)$ is called the *horizontal lift* of $\mathcal{Y}(0)$ through Y_0 .

Proposition 3.5 (parallel transport) *Let $t \mapsto \mathcal{Y}(t)$ be a smooth curve on $\text{Grass}(p, n)$. Let ξ be a tangent vector field on $\text{Grass}(p, n)$ defined along $\mathcal{Y}(\cdot)$. Let $t \mapsto Y(t)$ be a horizontal lift of $t \mapsto \mathcal{Y}(t)$. Let $\xi_{\diamond Y(t)}$ denote the horizontal lift of ξ at $Y(t)$ as defined in (9). Then ξ is parallel transported along the curve $\mathcal{Y}(\cdot)$ if and only if*

$$\boxed{\dot{\xi}_{\diamond Y(t)} + Y(t)(Y(t)^T Y(t))^{-1} \dot{Y}(t)^T \xi_{\diamond Y(t)} = 0} \quad (18)$$

where $\dot{\xi}_{\diamond Y(t)} := \frac{d}{d\tau} \xi_{\diamond Y(\tau)}|_{\tau=t}$.

In other words, the parallel transport of ξ along $\mathcal{Y}(\cdot)$ is obtained by infinitesimally removing the vertical component (the second term in the left-hand side of (18) is vertical) of the horizontal lift of ξ along a horizontal lift of $\mathcal{Y}(\cdot)$.

Proof. Let $t \mapsto \mathcal{Y}(t)$, ξ and $t \mapsto Y(t)$ be as in the statement of the proposition. Then $\dot{Y}(t)$ is the horizontal lift of $\dot{\mathcal{Y}}(t)$ at $Y(t)$ and

$$\nabla_{\dot{\mathcal{Y}}(t)} \xi = \Pi_{Y_{\perp}} \dot{\xi}_{\diamond Y(t)}$$

by (16), where $\dot{\xi}_{\diamond Y(t)} := \frac{d}{d\tau} \xi_{\diamond Y(\tau)}|_{\tau=t}$. So $\nabla_{\dot{\mathcal{Y}}(t)} \xi = 0$ if and only if $\Pi_{Y_{\perp}} \dot{\xi}_{\diamond Y(t)} = 0$, i.e. $\dot{\xi}_{\diamond Y(t)} \in V_{Y(t)}$, i.e. $\dot{\xi}_{\diamond Y(t)} = Y(t)M(t)$ for some $M(t)$. Since ξ_{\diamond} is horizontal, one has $Y^T \xi_{\diamond Y} = 0$, thus $\dot{Y}^T \xi_{\diamond Y} + Y^T \dot{\xi}_{\diamond Y} = 0$ and therefore $M = -(Y^T Y)^{-1} \dot{Y}^T \xi_{\diamond Y}$. \square

It is interesting to notice that (18) is not symmetric in \dot{Y} and ξ_{\diamond} . This is apparently in contradiction with the symmetry of the Riemannian connection, but one should bear in mind that Y and ξ_{\diamond} are *not* expressions of \mathcal{Y} and ξ in a fixed coordinate chart, so (18) need not be symmetric.

Geodesics

We now give a formula for the geodesic $t \mapsto \mathcal{Y}(t)$ with initial point $\mathcal{Y}(0) = \mathcal{Y}_0$ and initial “velocity” $\dot{\mathcal{Y}}_0 \in T_{\mathcal{Y}_0} \text{Grass}(p, n)$. The geodesic is characterized by $\nabla_{\dot{\mathcal{Y}}} \dot{\mathcal{Y}} = 0$, which says that the tangent vector to $\mathcal{Y}(\cdot)$ is parallel transported along $\mathcal{Y}(\cdot)$. This expresses the idea that $\mathcal{Y}(t)$ goes “straight on at constant pace”.

Theorem 3.6 (geodesics) *Let $t \mapsto \mathcal{Y}(t)$ be a geodesic on $\text{Grass}(p, n)$ with Riemannian metric (13) from \mathcal{Y}_0 with initial velocity $\dot{\mathcal{Y}}_0 \in T_{\mathcal{Y}_0} \text{Grass}(p, n)$. Let Y_0 span \mathcal{Y}_0 , let $(\dot{\mathcal{Y}}_0)_{\diamond Y_0}$ be the horizontal lift of $\dot{\mathcal{Y}}_0$, and let $(\dot{\mathcal{Y}}_0)_{\diamond Y_0} (Y_0^T Y_0)^{-1/2} = U \Sigma V^T$ be a thin singular value decomposition, i.e. U is $n \times p$ orthonormal, V is $p \times p$ orthonormal and Σ is $p \times p$ diagonal with nonnegative elements.*

Then

$$\boxed{\mathcal{Y}(t) = \text{span}(Y_0(Y_0^T Y_0)^{-1/2} V \cos \Sigma t + U \sin \Sigma t).} \quad (19)$$

This expression obviously simplifies when Y_0 is chosen orthonormal.

The exponential of $\dot{\mathcal{Y}}_0$, denoted by $\text{Exp}(\dot{\mathcal{Y}}_0)$, is by definition $\mathcal{Y}(t=1)$.

Note that this formula is not new except for the fact that a nonorthonormal Y_0 is allowed. In practice, however, one will prefer to orthonormalize Y_0 and use the simplified expression. Edelman, Arias and Smith [EAS98] obtained the orthonormal version of the geodesic formula using the symmetric space structure of $\text{Grass}(p, n) = O_n / (O_p \times O_{n-p})$.

Proof. Let $\mathcal{Y}(t)$, Y_0 , $U\Sigma V^T$ be as in the statement of the theorem. Let $t \mapsto Y(t)$ be the unique horizontal lift of $\mathcal{Y}(\cdot)$ through Y_0 , so that $\dot{\mathcal{Y}}_{\diamond Y(t)} = \dot{Y}(t)$. Then the formula for parallel transport (18), applied to $\xi := \dot{\mathcal{Y}}$, yields

$$\ddot{Y} + Y(Y^T Y)^{-1} \dot{Y}^T \dot{Y} = 0. \quad (20)$$

Since $Y(\cdot)$ is horizontal, one has

$$Y^T(t) \dot{Y}(t) = 0 \quad (21)$$

which is compatible with (20). This implies that $Y(t)^T Y(t)$ is constant. Moreover, equations (20) and (21) imply that $\frac{d}{dt} \dot{Y}(t)^T \dot{Y}(t) = 0$, so $\dot{Y}(t)^T \dot{Y}(t)$ is constant. Consider the thin SVD $\dot{Y}(0)(Y^T Y)^{-1/2} = U\Sigma V^T$. From (20), one obtains

$$\begin{aligned} \ddot{Y}(t)(Y^T Y)^{-1/2} + Y(t)(Y^T Y)^{-1/2} (Y^T Y)^{-1/2} (\dot{Y}^T \dot{Y})(Y^T Y)^{-1/2} &= 0 \\ \ddot{Y}(t)(Y^T Y)^{-1/2} + Y(t)(Y^T Y)^{-1/2} V \Sigma^2 V^T &= 0 \\ \ddot{Y}(t)(Y^T Y)^{-1/2} V + Y(t)(Y^T Y)^{-1/2} V \Sigma^2 &= 0 \end{aligned}$$

which yields

$$Y(t)(Y^T Y)^{-1/2} V = Y_0(Y_0^T Y_0)^{-1/2} V \cos \Sigma t + \dot{Y}_0(Y_0^T Y_0)^{-1/2} V \Sigma^{-1} \sin \Sigma t$$

and the result follows. \square

As an aside, Theorem 3.6 shows that the Grassmann manifold is *complete*, i.e. the geodesics can be extended indefinitely [Boo75].

Distance between subspaces

The geodesics can be locally interpreted as curves of shortest length [Boo75]. This motivates the following notion of distance between two subspaces.

Let \mathcal{X} and \mathcal{Y} belong to $\text{Grass}(p, n)$ and let X, Y be orthonormal bases for \mathcal{X}, \mathcal{Y} , respectively. Let $\mathcal{Y} \in \mathcal{U}_X$ (6), i.e. $X^T Y$ is invertible. Let $\Pi_{X^\perp} Y (X^T Y)^{-1} = U\Sigma V^T$ be an SVD. Let $\Theta = \text{atan}\Sigma$. Then the geodesic

$$t \mapsto \text{Expt}\xi = \text{span}(XV \cos \Theta t + U \sin \Theta t),$$

where $\xi_{\diamond X} = U\Theta V^T$, is the shortest curve on $\text{Grass}(p, n)$ from \mathcal{X} to \mathcal{Y} . The elements θ_i of Θ are called the principal angles between \mathcal{X} and \mathcal{Y} . The columns of XV and those of $(XV \cos \Theta + U \sin \Theta)$ are the corresponding principal vectors. The geodesic distance on $\text{Grass}(p, n)$ induced by the metric (13) is

$$\text{dist}(\mathcal{X}, \mathcal{Y}) = \sqrt{\langle \xi, \xi \rangle} = \sqrt{\theta_1^2 + \dots + \theta_p^2}.$$

Other definitions of distance on Grassmann are given in [EAS98, 4.3]. A classical one is the projection 2-norm $\|\Pi_X - \Pi_Y\|_2 = \sin \theta_{\max}$ where θ_{\max} is the largest principal angle [Ste73, GV96]. An algorithm for computing the principal angles and vectors is given in [BG73, GV96].

Discussion

This completes our study of the Riemannian structure of the Grassmann manifold $\text{Grass}(p, n)$ using bases, i.e. elements of $\text{ST}(p, n)$, to represent its elements. We are now ready to give in the next section a formulation of the Riemann-Newton method on the Grassmann manifold. Following Smith [Smi94], the function F in (1) becomes a tangent vector field ξ (Smith works with one-forms, but this is equivalent because the Riemannian connection leaves the metric invariant [MM02]). The directional derivative D in (1) is replaced by the Riemannian connection, for which we have given a formula in Theorem 3.4. As far as we know, this formula has never been published, and as we shall see it makes the derivation of the Newton algorithm very simple for some vector fields ξ . The update (2) is performed along the geodesic (Theorem 3.6) generated by the Newton vector. Convergence of the algorithms can be assessed using the notion of distance defined above.

4 Newton iteration on the Grassmann manifold

A number of authors have proposed and developed a general theory of Newton iteration on Riemannian manifolds [Gab82, Smi93, Smi94, Udr94, MM02]. In particular, Smith [Smi94] proposes an algorithm for abstract Riemannian manifolds which amounts to the following.

Algorithm 4.1 (Riemann-Newton) *Let M be a Riemannian manifold, ∇ be the Levi-Civita connection on M , and ξ be a smooth vector field on M . The Newton iteration on M for computing a zero of ξ consists in iterating the mapping $x \mapsto x_+$ defined by*

1. Solve the Newton equation

$$\nabla_\eta \xi = -\xi(x) \quad (22)$$

for $\eta \in T_x M$.

2. Compute the update $x_+ := \text{Exp } \eta$, where Exp denotes the Riemannian exponential mapping.

The Riemann-Newton iteration, expressed in the so-called normal coordinates at x (normal coordinates use the inverse exponential as a coordinate chart [Boo75]), reduces to the classical Newton method (1)-(2) [MM02]. It converges locally quadratically to the *nondegenerate* zeros of ξ , i.e. the points x such that $\xi(x) = 0$ and $T_x \text{Grass}(p, n) \ni \eta \mapsto \nabla_\eta \xi$ is invertible (see proof in Section A).

On the Grassmann manifold, the Riemann-Newton iteration yields the following algorithm.

Theorem 4.2 (Grassmann-Newton) *Let the Grassmann manifold $\text{Grass}(p, n)$ be endowed with the O_n -invariant metric (13). Let ξ be a smooth vector field on $\text{Grass}(p, n)$. Let ξ_\diamond denote the horizontal lift of ξ as defined in (9). Then the Riemann-Newton method (Algorithm 4.1) on $\text{Grass}(p, n)$ for ξ consists in iterating the mapping $\mathcal{Y} \mapsto \mathcal{Y}_+$ defined by*

1. Pick a basis Y that spans \mathcal{Y} and solve the equation

$$\Pi_{Y^\perp} D\xi_\diamond(Y)[\eta_{\diamond Y}] = -\xi_{\diamond Y} \quad (23)$$

for the unknown $\eta_{\diamond Y}$ in the horizontal space $H_Y = \{Y^\perp K : K \in \mathbb{R}^{(n-p) \times p}\}$.

2. Compute an SVD $\eta_{\diamond Y} = U\Sigma V^T$ and perform the update

$$\mathcal{Y}_+ := \text{span}(YV \cos \Sigma + U \sin \Sigma). \quad (24)$$

Proof. Equation (23) is the horizontal lift of equation (22) where the formula (16) for the Riemannian connection has been used. Equation (24) is the exponential update given in formula (19). \square

It often happens that ξ is the gradient (14) of a cost function f , $\xi = \text{grad } f$, in which case the Newton iteration searches a stationary point of f . In this case, the Newton equation (23) reads

$$\Pi_{Y^\perp} D(\Pi_{Y^\perp} \text{grad } f_\diamond(\cdot))(Y)[\eta_{\diamond Y}] = -\Pi_{Y^\perp} \text{grad } f_\diamond(Y)$$

where the formula (14) has been used for the Grassmann gradient. This equation can be interpreted as the Newton equation in $\mathbb{R}^{n \times p}$

$$D(\text{grad } f_\diamond(\cdot))(Y)[\Delta] = -\text{grad } f_\diamond(Y)$$

projected onto the horizontal space (8). The projection operation cancels out the directions along the equivalence class $Y \text{GL}_p$, which intuitively makes sense since they do not generate variations of the span of Y .

It is often the case that ξ admits the expression

$$\xi_{\diamond Y} = \Pi_{Y^\perp} F(Y) \quad (25)$$

where F is a homogeneous function, i.e. $F(YM) = F(Y)M$. In this case, the Newton equation (23) becomes

$$\Pi_{Y^\perp} DF(Y)[\eta_{\diamond Y}] - \eta_{\diamond Y}(Y^T Y)^{-1} Y^T F(Y) = -\Pi_{Y^\perp} F(Y) \quad (26)$$

where we have taken into account that $Y^T \eta_{\diamond Y} = 0$ since $\eta_{\diamond Y}$ is horizontal.

5 Practical applications of the Newton method

In this section, we illustrate the applicability of the Grassmann-Newton method (Theorem 4.2) on two problems that can be cast as the computing a zero of a tangent vector field on the Grassmann manifold.

Invariant subspace computation

Let A be an $n \times n$ matrix and let

$$\xi_{\diamond Y} := \Pi_{Y^\perp} AY \quad (27)$$

where Π_{Y^\perp} denotes the projector (12) into the orthogonal complement of the span of Y . This expression is homogeneous and horizontal, therefore it is a well-defined horizontal lift and defines a tangent vector field on Grassmann. Moreover, $\xi(\mathcal{Y}) = 0$ if and only if \mathcal{Y} is an invariant subspace of A . Obtaining the Newton equation (23) for ξ defined in (27) is now extremely simple: the simplification (25) holds with $F(Y) = AY$, and (26) immediately yields

$$\Pi_{Y^\perp} (A\eta_{\diamond Y} - \eta_{\diamond Y}(Y^T Y)^{-1} Y^T AY) = -\Pi_{Y^\perp} AY \quad (28)$$

which has to be solved for $\eta_{\diamond Y}$ in the horizontal space H_Y (8). The resulting iteration, (28)-(24), converges locally to the nondegenerate zeros of ξ , which are the spectral² right invariant

²A right invariant subspace \mathcal{Y} of A is termed *spectral* if, given $[Y|Y^\perp]$ orthogonal such that Y spans \mathcal{Y} , $Y^T AY$ and $Y_\perp^T AY_\perp$ have no eigenvalue in common [RR02].

subspaces of A ; see [Abs03] for details. The rate of convergence is quadratic. It is cubic if and only if the zero of ξ is also a left invariant subspace of A (see Section B). This happens in particular when $A = A^T$.

Edelman, Arias and Smith [EAS98] consider the Newton method on Grassmann for the Rayleigh quotient cost function $f_\diamond(Y) := \text{trace}((Y^T Y)^{-1} Y^T A Y)$, assuming $A = A^T$. They obtain the same equation (28), which is not surprising since it can be shown, using (14), that our ξ is the gradient of their f .

Equation (28) also connects with a method proposed by Chatelin [Cha84] for refining invariant subspace estimates. She considers the equation $AY = YB$ whose solutions $Y \in \text{ST}(p, n)$ span invariant subspaces of A and imposes a normalization condition $Z^T Y = I$ on Y , where Z is a given $n \times p$ matrix. This normalization condition can be interpreted as restricting Y to a cross section (5). Then she applies the classical Newton method for finding solutions of $AY = YB$ in the cross section and obtains an equation similar to (28). The equations are in fact identical if the matrix Z is chosen to span the current iterate. Following Chatelin's approach, the projective update $\mathcal{Y}_+ = \text{span}(Y + \eta_{\diamond} Y)$ is used instead of the geodesic update (24).

The algorithm with projective update is also related to the Grassmannian Rayleigh quotient iteration (GRQI) proposed in [AMSV02]. The two methods are identical when $p = 1$ [SE02]. They differ when $p > 1$, but they both compute eigenspaces of $A = A^T$ with cubic rate of convergence. For A arbitrary, a two-sided version of GRQI is proposed in [AV02] that also computes the eigenspaces with cubic rate of convergence.

Methods for solving (28) are given in [Dem87] and [LE02]. Lundström and Eldén [LE02] give an algorithm that allows to solve the equation without explicitly computing the interaction matrix $\Pi_{Y_\perp} A \Pi_{Y_\perp}$. The global behaviour of the iteration is studied in [ASVM04] and heuristics are proposed that enlarge the basins of attraction of the invariant subspaces.

Mean of subspaces

Let \mathcal{Y}^i , $i = 1, \dots, m$, be a collection of p -dimensional subspaces of \mathbb{R}^n . We consider the problem of computing the mean of the subspaces \mathcal{Y}^i . Since $\text{Grass}(p, n)$ is complete, if the subspaces \mathcal{Y}^i are clustered sufficiently close together then there is a unique \mathcal{X} that minimizes $V(\mathcal{X}) := \sum_{i=1}^m \text{dist}^2(\mathcal{X}, \mathcal{Y}^i)$. This \mathcal{X} is called the Karcher mean of the m subspaces [Kar77, Ken90].

A steepest descent algorithm is proposed in [Woo02] for computing the Karcher mean of a cluster of point on a Riemannian manifold. Since it is a steepest descent algorithm, its convergence rate is only linear.

The Karcher mean verifies $\sum_{i=1}^m \delta^i = 0$ where $\delta^i := \text{Exp}_{\mathcal{X}}^{-1} \mathcal{Y}^i$. This suggests to take $\xi(\mathcal{X}) := \sum_{i=1}^m \text{Exp}_{\mathcal{X}}^{-1} \mathcal{Y}^i$ and apply the Riemann-Newton algorithm. On the Grassmann manifold, however, this idea does not work well because of the complexity of the relation between \mathcal{Y}^i and δ^i , see Section 3. Therefore, we use another definition of the mean in which $\delta_{\diamond X}^i = \Pi_{X_\perp} \Pi_{Y^i} X$, where X spans \mathcal{X} , Y^i spans \mathcal{Y}^i , $\Pi_{X_\perp} = I - X(X^T X)^{-1} X^T$ is the orthogonal projector into the orthogonal complement of the span of X and $\Pi_Y = Y(Y^T Y)^{-1} Y^T$ is the orthogonal projector into the span of \mathcal{Y} . While the Karcher mean minimizes $\sum_{i=1}^m \sum_{j=1}^p \theta_{i,j}^2$ where $\theta_{i,j}$ is the j th canonical angle between \mathcal{X} and \mathcal{Y}^i , our modified mean minimizes $\sum_{i=1}^m \sum_{j=1}^p \sin^2 \theta_{i,j}$. Both definitions are asymptotically equivalent for small principal an-

gles. Our definition yields

$$\xi_{\circ X} = \sum_{i=1}^m \Pi_{X_{\perp}} \Pi_{Y^i} X$$

and one readily obtains, using (26), the following expression for the Newton equation

$$\sum_{i=1}^m (\Pi_{X_{\perp}} \Pi_{Y^i} \eta_{\circ X} - \eta_{\circ X} (X^T X)^{-1} X^T \Pi_{Y^i} X) = - \sum_{i=1}^m \Pi_{X_{\perp}} \Pi_{Y^i} X$$

which has to be solved for $\eta_{\circ X}$ in the horizontal space $H_X = \{X_{\perp} K : K \in \mathbb{R}^{(n-p) \times p}\}$.

We have tested the resulting Newton iteration in the following situation. We draw m samples $K_i \in \mathbb{R}^{(n-p) \times p}$ where the elements of each K are i.i.d. normal random variables with mean zero and standard deviation $1e^{-2}$ and define $\mathcal{Y}^i = \text{span}(\begin{bmatrix} I_p \\ K_i \end{bmatrix})$. The initial iterate is $\mathcal{X} := \mathcal{Y}^1$ and we define $\mathcal{Y}^0 = \text{span}(\begin{bmatrix} I_p \\ 0 \end{bmatrix})$. Experimental results are shown below.

	Newton		$X_+ = X - \text{grad}V/m$	
Iterate number	$\ \sum_{i=1}^m \delta^i\ $	$\text{dist}(\mathcal{X}, \mathcal{Y}^0)$	$\ \sum_{i=1}^m \delta^i\ $	$\text{dist}(\mathcal{X}, \mathcal{Y}^0)$
0	2.4561e+01	2.9322e-01	2.4561e+01	2.9322e-01
1	1.6710e+00	3.1707e-02	1.9783e+01	2.1867e-01
2	5.7656e-04	2.0594e-02	1.6803e+01	1.6953e-01
3	2.4207e-14	2.0596e-02	1.4544e+01	1.4911e-01
4	8.1182e-16	2.0596e-02	1.2718e+01	1.2154e-01
300			5.6525e-13	2.0596e-02

6 Conclusion

We have considered the Grassmann manifold $\text{Grass}(p, n)$ of p -planes in \mathbb{R}^n as the base space of a GL_p -principal fiber bundle with the noncompact Stiefel manifold $\text{ST}(p, n)$ as total space. Using the essentially unique O_n -invariant metric on $\text{Grass}(p, n)$, we have derived a formula for the Levi-Civita connection in terms of horizontal lifts. Moreover, formulas have been given for the Lie bracket, parallel translation, geodesics and distance between p -planes. Finally, these results have been applied to a detailed derivation of the Newton method on the Grassmann manifold. The Grassmann-Newton method has been illustrated on two examples.

A Quadratic convergence of Riemann-Newton

For completeness we include a proof of quadratic convergence of the Riemann-Newton iteration (Algorithm 4.1). Our proof significantly differs from the proof previously reported in the literature [Smi94]. This proof also prepares the discussion on cubic convergence cases in Section B.

Let ξ be a smooth vector field on a Riemannian manifold M and let ∇ denote the Riemannian connection. Let $z \in M$ be a nondegenerate zero of the smooth vector field ξ (i.e. $\xi_z = 0$ and the linear operator $T_z M \ni \eta \mapsto \nabla_{\eta} \xi \in T_z M$ is invertible). Let N_z be a normal neighbourhood of z , sufficiently small so that any two points of N_z can be joined by a unique geodesic [Boo75]. Let τ_{xy} denote the parallel transport along the unique geodesic between x and y . Let the tangent vector $\zeta \in T_x M$ be defined by $\text{Exp}_x \zeta = z$. Define the vector field $\tilde{\zeta}$

on N_z adapted to the tangent vector $\zeta \in T_x M$ by $\tilde{\zeta}_y = \tau_{xy}\zeta$. Applying Taylor's formula to the function $\lambda \mapsto \xi_{\text{Exp}_x \lambda \zeta}$ yields [Smi94]

$$0 = \xi_z = \xi_x + \nabla_{\zeta} \xi + \frac{1}{2} \nabla_{\zeta}^2 \xi + O(\zeta^3) \quad (29)$$

where $\nabla_{\zeta}^2 \xi := \nabla_{\zeta}(\nabla_{\zeta} \xi)$. Subtracting Newton equation (22) from Taylor's formula (29) yields

$$\nabla_{(\eta - \zeta)} \xi = \nabla_{\zeta}^2 \xi + O(\zeta^3). \quad (30)$$

Since ξ is a smooth vector field and z is a nondegenerate zero of ξ , and reducing the size of N_z if necessary, one has

$$\begin{aligned} \|\nabla_{\alpha} \xi\| &\geq c_1 \|\alpha\| \\ \|\nabla_{\alpha}^2 \xi\| &\leq c_3 \|\alpha\|^2 \end{aligned}$$

for all $y \in N_z$ and $\alpha \in T_y M$. Using these results in (30) yields

$$c_1 \|\eta - \zeta\| \leq c_2 \|\zeta\|^2 + O(\zeta^3). \quad (31)$$

From now on the proof significantly differs from the one in [Smi94]. We will show next that, reducing again the size of N_z if necessary, there exists a constant c_4 such that

$$\text{dist}(\text{Exp}_y \alpha, \text{Exp}_y \beta) \leq c_4 \|\alpha - \beta\| \quad (32)$$

for all $y \in N_z$ and all $\alpha, \beta \in T_y M$ small enough for $\text{Exp}_y \alpha$ and $\text{Exp}_y \beta$ to be in N_z . Then it follows immediately from (31) and (32) that

$$\text{dist}(x_+, z) = \text{dist}(\text{Exp}_y \eta, \text{Exp}_y \zeta) \leq \|\eta - \zeta\| \leq \frac{c_2}{c_1} \|\zeta\|^2 + O(\zeta^3) = O(\text{dist}(x, z)^2)$$

and this is quadratic convergence.

To show (32), we work in local coordinates covering N_z and use tensorial notations (see e.g. [Boo75]), so e.g. u^i denotes the coordinates of $u \in M$. Consider the geodesic equation $\ddot{u}_i + \Gamma_{jk}^i(u) \dot{u}^j \dot{u}^k = 0$ where Γ stands for the (smooth) Christoffel symbol, and denote the solution by $\phi^i[t, u(0), \dot{u}(0)]$. Then $(\text{Exp}_y \alpha)^i = \phi^i[1, y, \alpha]$, $(\text{Exp}_y \beta)^i = \phi^i[1, y, \beta]$, and the curve $\gamma^i : \tau \mapsto \phi^i[1, y, \alpha + \tau(\beta - \alpha)]$ verifies $\gamma^i(0) = (\text{Exp}_y \alpha)^i$ and $\gamma^i(1) = (\text{Exp}_y \beta)^i$. Then

$$\begin{aligned} &\text{dist}(\text{Exp}_y \alpha, \text{Exp}_y \beta) \\ &\leq \int_0^1 \sqrt{g_{ij}[\gamma(\tau)] \dot{\gamma}^i(\tau) \dot{\gamma}^j(\tau)} d\tau \end{aligned} \quad (33)$$

$$\begin{aligned} &= \int_0^1 \sqrt{g_{ij}[\gamma(\tau)] \frac{\partial \phi^i}{\partial \dot{u}^k}[1, y, \alpha + \tau(\beta - \alpha)] \frac{\partial \phi^j}{\partial \dot{u}^\ell}[1, y, \alpha + \tau(\beta - \alpha)] (\beta^k - \alpha^k)(\beta^\ell - \alpha^\ell)} d\tau \\ &\leq c' \sqrt{\delta_{k\ell}(\beta^k - \alpha^k)(\beta^\ell - \alpha^\ell)} \end{aligned} \quad (34)$$

$$\begin{aligned} &\leq c_4 \sqrt{g_{k\ell}[y](\beta^k - \alpha^k)(\beta^\ell - \alpha^\ell)} \\ &= c_4 \|\beta - \alpha\|. \end{aligned} \quad (35)$$

Equation (33) gives the length of the curve $\gamma(0, 1)$, for which $\text{dist}(\text{Exp}_y \alpha, \text{Exp}_y \beta)$ is a lower bound. Equation (34) comes from the fact that the metric tensor g_{ij} and the derivatives of ϕ are smooth functions defined on a compact set, thus bounded. Equation (35) comes because g_{ij} is nondegenerate and smooth on a compact set.

B Cubic convergence of Riemann-Newton

We use the notations of the previous section.

If $\nabla_\alpha^2 \xi = 0$ for all tangent vector $\alpha \in T_z M$, then the rate of convergence of the Riemann-Newton method (Algorithm 4.1) is cubic. Indeed, by the smoothness of ξ , and defining ζ_x such that $\text{Exp}_x \zeta = z$, one has $\nabla_\zeta^2 \xi = O(\zeta^3)$, and replacing this into (30) gives the result.

For the sake of illustration, we consider the particular case where M is the Grassmann manifold $\text{Grass}(p, n)$ of p -planes in \mathbb{R}^n , A is an $n \times n$ real matrix and the tangent vector field ξ is defined by the horizontal lift (27)

$$\xi_{\diamond Y} = \Pi_{Y^\perp} AY$$

where $\Pi_{Y^\perp} := (I - YY^T)$. Let $\mathcal{Z} \in \text{Grass}(p, n)$ verify $\xi_{\mathcal{Z}} = 0$, which happens if and only if \mathcal{Z} is a right invariant subspace of A . We show that $\nabla_\alpha^2 \xi = 0$ for all $\alpha \in T_{\mathcal{Z}} \text{Grass}(p, n)$ if and only if \mathcal{Z} is a left invariant subspace of A (which happens e.g. when $A = A^T$).

Let Z be an orthonormal basis for \mathcal{Z} , i.e. $\xi_{\diamond Z} = 0$ and $Z^T Z = I$. Let $\alpha_{\diamond Z} = U \Sigma V^T$ be a thin singular value decomposition of $\alpha \in T_{\mathcal{Z}} \text{Grass}(p, n)$. Then the curve

$$Y(t) = ZV \cos \Sigma t + U \sin \Sigma t$$

is horizontal and projects through the “span” operation to the Grassmann geodesic $\text{Exp}_{\mathcal{Z}} t \alpha$. Since by definition the tangent vector of a geodesic is parallel transported along the geodesic, the adaptation $\tilde{\alpha}$ of α verifies

$$\tilde{\alpha}_{\diamond Y(t)} = \dot{Y}(t) = U \Sigma \cos \Sigma t - ZV \Sigma \cos \Sigma t.$$

Then one obtains successively

$$\begin{aligned} (\nabla_{\tilde{\alpha}} \xi)_{\diamond Y(t)} &= \Pi_{Y(t)^\perp} D\xi_{\diamond}(Y(t))[\tilde{\alpha}_{\diamond Y(t)}] \\ &= \Pi_{Y(t)^\perp} \frac{d}{dt} \Pi_{Y(t)^\perp} AY(t) \\ &= \Pi_{Y(t)^\perp} A \dot{Y}(t) - \Pi_{Y(t)^\perp} \dot{Y}(t) Y(t)^T AY(t) \end{aligned}$$

and

$$\begin{aligned} \nabla_\alpha^2 \xi &= (\nabla_\alpha \nabla_{\tilde{\alpha}} \xi)_{\diamond Z} \\ &= \Pi_{Z^\perp} \frac{d}{dt} (\nabla_{\tilde{\alpha}} \xi)_{\diamond Y(t)}|_{t=0} \\ &= -\Pi_{Z^\perp} A \ddot{Y}(0) - \Pi_{Z^\perp} \dot{Y}(0) Y(0)^T A \dot{Y}(0) - \Pi_{Z^\perp} \dot{Y}(0) (\dot{Y}(0)^T AY(0) + Y(0)^T A \dot{Y}(0)) \\ &= -2U \Sigma V^T Z^T A U \Sigma \end{aligned}$$

where we have used $\Pi_{Y^\perp} Y = 0$, $Z^T U = 0$, $U^T A Z = \Pi_{Z^\perp} A Z = 0$. This last expression vanishes for all $\alpha \in T_{\mathcal{Z}} \text{Grass}(p, n)$ if and only if $U^T A^T Z = 0$ for all U such that $U^T Z = 0$, i.e. \mathcal{Z} is a left invariant subspace of A .

Acknowledgements

The first author thanks P. Lecomte (Université de Liège) and U. Helmke and S. Yoshizawa (Universität Würzburg) for useful discussions.

This paper presents research partially supported by the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. Part of this work was performed while the first author was a guest at the Mathematisches Institut der Universität Würzburg under a grant from the European Nonlinear Control Network. The hospitality of the members of the institute is gratefully acknowledged. The work was completed while the first and last authors were visiting the department of Mechanical and Aerospace Engineering at Princeton university. The hospitality of the members of the department, especially Prof. N. Leonard, is gratefully acknowledged. The last author thanks N. Leonard and E. Sontag for partial financial support under US Air Force Grants F49620-01-1-0063 and F49620-01-1-0382.

References

- [Abs03] P.-A. Absil, *Invariant subspace computation: A geometric approach*, Ph.D. thesis, Faculté des Sciences Appliquées, Université de Liège, Secrétariat de la FSA, Chemin des Chevreuils 1 (Bât. B52), 4000 Liège, Belgium, 2003.
- [AMSV02] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, *A Grassmann-Rayleigh quotient iteration for computing invariant subspaces*, SIAM Review **44** (2002), no. 1, 57–73.
- [ASVM04] P.-A. Absil, R. Sepulchre, P. Van Dooren, and R. Mahony, *Cubically convergent iterations for invariant subspace computation*, to appear in SIAM J. Matrix Anal. Appl., 2004.
- [AV02] P.-A. Absil and P. Van Dooren, *Two-sided Grassmann-Rayleigh quotient iteration*, submitted to SIAM J. Matrix Anal. Appl., October 2002.
- [BG73] Å. Björk and G. H. Golub, *Numerical methods for computing angles between linear subspaces*, Math. Comp. **27** (1973), 579–594.
- [Boo75] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, 1975.
- [CG90] P. Common and G. H. Golub, *Tracking a few extreme singular values and vectors in signal processing*, Proceedings of the IEEE **78** (1990), no. 8, 1327–1343.
- [Cha84] F. Chatelin, *Simultaneous Newton's iteration for the eigenproblem*, Computing, Suppl. **5** (1984), 67–74.
- [Cha93] I. Chavel, *Riemannian geometry - A modern introduction*, Cambridge University Press, 1993.
- [dC92] M. P. do Carmo, *Riemannian geometry*, Birkhäuser, 1992.
- [Dem87] J. W. Demmel, *Three methods for refining estimates of invariant subspaces*, Computing **38** (1987), 43–57.
- [DM90] B. F. Doolin and C. F. Martin, *Introduction to differential geometry for engineers*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 136, Marcel Dekker, Inc., New York, 1990.

- [DS83] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice Hall Series in Computational Mathematics, Prentice Hall, Englewood Cliffs, NJ, 1983.
- [EAS98] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.
- [FGP94] J. Ferrer, M. I. García, and F. Puerta, *Differentiable families of subspaces*, Linear Algebra Appl. **199** (1994), 229–252.
- [Gab82] D. Gabay, *Minimizing a differentiable function over a differential manifold*, Journal of Optimization Theory and Applications **37** (1982), no. 2, 177–219.
- [GV96] G. H. Golub and C. F. Van Loan, *Matrix computations, third edition*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 1996.
- [Hel78] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Pure and Applied Mathematics, no. 80, Academic Press, Oxford, 1978.
- [HM94] U. Helmke and J. B. Moore, *Optimization and dynamical systems*, Springer, 1994.
- [Kar77] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Comm. Pure Appl. Math. **30** (1977), no. 5, 509–541.
- [Ken90] W. S. Kendall, *Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence*, Proc. London Math. Soc. **61** (1990), no. 2, 371–406.
- [KN63] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, volumes 1 and 2*, John Wiley & Sons, 1963.
- [LE02] E. Lundström and L. Eldén, *Adaptive eigenvalue computations using Newton’s method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl. **23** (2002), no. 3, 819–839.
- [Lei61] K. Leichtweiss, *Zur riemannschen Geometrie in grassmannschen Mannigfaltigkeiten*, Math. Z. **76** (1961), 334–366.
- [Lue69] D. G. Luenberger, *Optimization by vector space methods*, John Wiley & Sons, Inc., 1969.
- [Mah96] R. E. Mahony, *The constrained Newton method on a Lie group and the symmetric eigenvalue problem*, Linear Algebra Appl. **248** (1996), 67–89.
- [MM02] R. Mahony and J. H. Manton, *The geometry of the Newton method on non-compact Lie groups*, J. Global Optim. **23** (2002), no. 3, 309–327.
- [MS85] A. Machado and I. Salavessa, *Grassmannian manifolds as subsets of Euclidean spaces*, Res. Notes in Math. **131** (1985), 85–102.
- [Nom54] K. Nomizu, *Invariant affine connections on homogeneous spaces*, Amer. J. Math. **76** (1954), 33–65.

- [RR02] A. C. M. Ran and L. Rodman, *A class of robustness problems in matrix analysis*, Interpolation Theory, Systems Theory and Related Topics, The Harry Dym Anniversary Volume (D. Alpay, I. Gohberg, and V. Vinnikov, eds.), Operator Theory: Advances and Applications, vol. 134, Birkhäuser, 2002, pp. 337–383.
- [SE02] V. Simoncini and L. Eldén, *Inexact Rayleigh quotient-type methods for eigenvalue computations*, BIT **42** (2002), no. 1, 159–182.
- [Smi93] S. T. Smith, *Geometric optimization methods for adaptive filtering*, Ph.D. thesis, Division of Applied Sciences, Harvard University, Cambridge, Massachusetts, 1993.
- [Smi94] ———, *Optimization techniques on Riemannian manifolds*, Hamiltonian and gradient flows, algorithms and control (Anthony Bloch, ed.), Fields Institute Communications, vol. 3, American Mathematical Society, 1994, pp. 113–136.
- [Ste73] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review **15** (1973), no. 4, 727–764.
- [Udr94] C. Udriște, *Convex functions and optimization methods on Riemannian manifolds*, Kluwer Academic Publishers, 1994.
- [Won67] Y.-C. Wong, *Differential geometry of Grassmann manifolds*, Proc. Nat. Acad. Sci. U.S.A **57** (1967), 589–594.
- [Woo02] R. P. Woods, *Characterizing volume and surface deformations in an atlas framework*, Submitted to NeuroImage, 2002.