

# A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank

BART VANDEREYCKEN\*

Department of Computer Science, Katholieke Universiteit Leuven,  
Celestijnenlaan 200A, 3001 Leuven, Belgium

P.-A. ABSIL

Department of Mathematical Engineering, Université catholique de Louvain,  
Bâtiment Euler 4–6, avenue Georges Lemaître, 1348 Louvain-la-Neuve, Belgium

AND

STEFAN VANDEWALLE

Department of Computer Science, Katholieke Universiteit Leuven,  
Celestijnenlaan 200A, 3001 Leuven, Belgium

September 2, 2010

## Abstract

We present a homogeneous space geometry for the manifold of symmetric positive semidefinite matrices of fixed rank. The total space is the general linear group endowed with its natural right-invariant metric, and the metric on the homogeneous space is chosen such that the quotient space is the image of a Riemannian submersion from the total space. As a result, we obtain complete geodesics that are the image of certain geodesics on the general linear group. We derive in addition an efficient closed-form expression for these geodesics. Furthermore, we isometrically embed the abstract quotient space into the space of real matrices. This allows us to interpret the vector fields, the metric and the geodesics in terms of concrete matrices. Finally, we compare our geometry with some other geometries in the literature. In particular, we show how other geodesics and their approximations relate to ours.

## 1 Introduction

Let  $p \leq n$  be two positive integers. The focus of this paper is the set of all real  $n \times n$  symmetric positive semidefinite (s.p.s.d.) matrices of fixed rank  $p$ . We denote this set by  $S_+(p, n)$ .

Central will be the description of  $S_+(p, n)$  as a Riemannian manifold. It is widely accepted that positive semidefinite matrices are fundamental objects in many areas of applied mathematics. They have their use in modelling as well as in computation. We mention only their application as covariance matrices in statistics (Huber, 1981), as optimization variables in semidefinite programming (Boyd & Vandenberghe, 2004) and as kernels in machine learning (Lanckriet *et al.*, 2004). In many of these applications, the link with the Riemannian geometry of  $S_+(n, n)$  turns out to be crucial; see, e.g, Smith (2005) for intrinsic Cramér–Rao bound analysis on  $S_+(n, n)$  and Pennec *et al.* (2006) for a Riemannian framework to image processing with positive definite tensors. Sometimes this link is established in a later phase, like in Nesterov & Todd (2008) where the optimisation paths in short-steps SDP methods were related to the geodesics on  $S_+(n, n)$ .

---

\*Corresponding author. Email: bart.vandereycken@cs.kuleuven.be

Most of the aforementioned applications deal with the full-rank matrices of  $S_+(n, n)$ . Since the typical matrix algorithms on  $S_+(n, n)$ , like the eigenvalue decomposition, have an  $O(n^3)$  complexity, they do not scale well for large  $n$ . A much-used remedy in matrix computation is to work with low-rank matrices. By exploiting this low-rank structure, one can usually reduce the cubic complexity to  $O(np^c)$ , with  $c$  small, say two. This way, large-scale matrix problems involving positive semidefinite matrices become tractable. Due to the importance of the Riemannian geometry of  $S_+(n, n)$ , it is sensible to generalise these geometrical implications to the fixed-rank case.

This paper can be seen as a contribution to the first step of this generalisation, namely to introduce a well-chosen geometry. There already exists quite a body of recent literature that makes use of  $S_+(p, n)$  as a Riemannian manifold. In contrast to the full rank case, we will show that there is no longer a natural geometry for  $S_+(p, n)$  when  $p < n$ . This freedom has resulted in several different geometries for  $S_+(p, n)$ : a submanifold embedded in  $\mathbb{R}^{n \times n}$  as in Helmke & Moore (1994), Helmke & Shayman (1995), Orsi *et al.* (2004), Orsi *et al.* (2006), Vandereycken *et al.* (2009), Vandereycken & Vandewalle (2010); a quotient manifold of  $\mathbb{R}^{n \times p}$  as in Absil *et al.* (2009), Journée *et al.* (2010), Meyer *et al.* (2010), Bonnabel *et al.* (2010); and a quotient manifold of the product of the Stiefel manifold and  $S_+(p, p)$  as in Meyer *et al.* (2009), Meyer *et al.* (2010), Bonnabel *et al.* (2010), Bonnabel & Sepulchre (2010), Sepulchre *et al.* (2010). All of these choices have their merits, but none of them succeeds in equipping  $S_+(p, n)$  with closed-form geodesics that can be extended indefinitely.

The existence of geodesics that can be extended indefinitely has a profound global implication: it guarantees that any two points on the manifold can be connected by a geodesic. The minimal length among all such connecting geodesics coincides with the usual distance function on the manifold. Furthermore, by virtue of the Hopf–Rinow theorem (Boothby, 1986, Th. VII.7.7), the manifold will be a complete metric space w.r.t. this distance function. Complete spaces are attractive in many ways and this is also the case when applied to manifolds. Here, we mention only one application area that benefits greatly from these complete geodesics, namely, that of optimisation and Newton algorithms on Riemannian manifolds; see Absil *et al.* (2008) for a recent overview. The prototype example of an optimisation algorithm performs a line-search along some search direction. In case of a manifold, this search can be done along a geodesic. The fact that the geodesics are complete and available in an efficient closed-form makes this line-search straightforward and well-defined. Alternatively, trust-region methods on manifolds also benefit since every modification to the trust-region will be algorithmically possible.

We do not wish to claim that geodesics are necessary for optimising on a manifold. Retraction-based Riemannian optimisation, like in Adler *et al.* (2002), dispenses with geodesics in favour of their first-order approximations, called retractions. Retractions that can be extended indefinitely surely share the same practical advantages as complete geodesics. However, most global convergence theory for optimisation on Riemannian manifolds assumes completeness; see, e.g., Ferreira & Svaiter (2002); Dedie *et al.* (2003); Li & Wang (2006); Absil *et al.* (2007); Alvarez *et al.* (2008)—although Yang (2006) does not.

In this paper, we therefore introduce another geometry which leads to complete geodesics, namely, that of a homogeneous space of the general linear group  $GL_n$ . In comparison to the existing approaches, homogeneous spaces have more structure and a richer theory; see, e.g., Kobayashi & Nomizu (1963). Together with the machinery of Riemannian submersions as in Cheeger & Ebin (1975), O’Neill (1983) and Gallot *et al.* (2004), we can derive the complete geodesics based on the geodesics of  $GL_n$ . Due to the importance of scalability in the applications of low-rank matrices, we give much attention to deriving efficient expressions for the typical objects of differential geometry, like these geodesics.

The plan of the paper is as follows. After some preliminaries to fix notation, we start in §3 with the description of  $S_+(p, n)$  as a homogeneous space. We outline why we choose the right-invariant metric on this space and derive efficient expressions for some typical objects from differential geometry. Then, in §4 we derive the complete geodesics for this metric. Most of the attention is paid to a closed-form expression. The next section, §5, is devoted to the isometrical embedding of the previous geometry into the space of real matrices. This was done to compare our geometry with other descriptions and to make interpreting the quotient geometry more evident. Since the

complete geodesics are the crux of the paper, in §6, we explain some special solutions to these geodesics and derive so-called quasi-geodesics. They allow for a lean expression of a connecting curve between any two matrices of  $S_+(p, n)$ . Finally, in §7 we compare our description with some other existing Riemannian geometries in the literature. Conclusions are drawn in §8.

## 2 Notational preliminaries

We will use the following notations:  $\mathbb{R}^{n \times p}$  is the space of real  $n \times p$  matrices,  $\mathbb{R}_*^{n \times p}$  is the restriction to full rank matrices. The identity matrix on  $\mathbb{R}^{n \times n}$  is denoted by  $I_n$ . Lie groups:  $GL_n := \mathbb{R}_*^{n \times n}$  is the general linear group and  $O_n := \{Q \in GL_n : Q^T Q = I_n\}$  is the orthogonal group. The symmetric and skew-symmetric  $n \times n$  matrices are denoted by  $\mathcal{S}_n^{\text{sym}}$  and  $\mathcal{S}_n^{\text{skew}}$ , respectively. If  $S \in \mathcal{S}_n^{\text{sym}}$ , then  $S \succ (\succeq) 0$  means that  $S$  is positive (semi)-definite.

Suppose  $Y \in \mathbb{R}_*^{n \times p}$  and  $n > p$ , then  $Y_\perp \in \mathbb{R}_*^{n \times (n-p)}$  is any orthonormal basis for the orthogonal complement of  $Y$  in  $GL_n$ , i.e.,  $Y^T Y_\perp = 0$  and  $Y_\perp^T Y_\perp = I_{n-p}$ . Moreover,  $P_Y := Y(Y^T Y)^{-1} Y^T$  is the orthogonal projector onto the column span of  $Y$  and  $P_Y^\perp := I_n - P_Y$ . In addition, two matrices  $Y$  and  $Z$  are perpendicular, denoted  $Y \perp Z$ , when  $Y^T Z = 0$ .

The directional derivative of a function  $f$  at  $x$  along  $\eta$  is denoted by  $Df(x)[\eta]$ . The tangent space of a manifold  $\mathcal{M}$  at  $x \in \mathcal{M}$  is  $T_x \mathcal{M}$ . The Lie bracket is denoted by  $[\eta, \nu]$ , which for a matrix Lie group equals  $[\eta, \nu] = \eta\nu - \nu\eta$ . The couple  $(\mathcal{M}, g)$  denotes the Riemannian manifold  $\mathcal{M}$  equipped with metric  $g$ . The existence of a diffeomorphism between two manifolds is denoted by  $\simeq$ . The matrix exponential of a matrix  $X \in \mathbb{R}^{n \times n}$  is given by  $\exp_m(X) = I_n + X + X^2/(2!) + X^3/(3!) + \dots$ .

## 3 Manifold $S_+(p, n)$ as a homogeneous space

We have the following characterization of  $S_+(p, n)$ .

**PROPOSITION 3.1** The set

$$S_+(p, n) = \{S : S \in \mathcal{S}_n^{\text{sym}}, S \succeq 0, \text{rank}(S) = p\} = \{YY^T : Y \in \mathbb{R}_*^{n \times p}\}$$

is a  $C^\infty$  smooth embedded submanifold in  $\mathbb{R}^{n \times n}$  of dimension  $pn - (p(p-1))/2$ . It has one connected component.

*Proof.* The fact that  $S_+(p, n)$  is a smooth embedded submanifold with one connected component follows from Helmke & Shayman (1995, Prop. 2.1). We proof the equivalence between the two definitions of  $S_+(p, n)$ . The inclusion  $\{YY^T : Y \in \mathbb{R}_*^{n \times p}\} \subseteq S_+(p, n)$  is obvious. The inclusion  $S_+(p, n) \subseteq \{YY^T : Y \in \mathbb{R}_*^{n \times p}\}$  can be shown using the eigenvalue decomposition: if  $S \in S_+(p, n)$ , then  $S$  can be written as  $UDU^T$ , where  $U$  is an  $n \times p$  orthonormal matrix and  $D = \text{diag}(d_i)$  is a  $p \times p$  diagonal matrix with strictly positive elements  $d_i$  on the diagonal. Hence  $S = UD^{1/2}(UD^{1/2})^T$  with  $D^{1/2} = \text{diag}(\sqrt{d_i})$ .  $\square$

### 3.1 Matrix congruence as a transitive group action

Two matrixes  $A, B \in \mathbb{R}^{n \times n}$  are called congruent if there exists a matrix  $X \in GL_n$  such that  $A = XBX^T$ . For symmetric matrices, Sylvester's law of inertia (Stewart, 2001, Th. 3.1) states that congruent matrices have the same number of positive, negative and zero eigenvalues. Hence, congruence defines a  $GL_n$ -action on  $S_+(p, n)$  that is also transitive. This allows us to describe  $S_+(p, n)$  as a so-called *homogeneous space*. We will assume a basic knowledge of homogeneous spaces which are a fundamental topic in Riemannian geometry, see, e.g., Boothby (1986) for an introduction.

**PROPOSITION 3.2** Manifold  $S_+(p, n)$  is a homogeneous space with transitive  $GL_n$ -action

$$\theta : GL_n \times S_+(p, n) \rightarrow S_+(p, n), (A, X) \mapsto AXA^T. \quad (1)$$

*Proof.* Transitivity of  $\theta$  means that there exists an  $A \in \mathrm{GL}_n$  for every  $X_1, X_2 \in \mathrm{S}_+(p, n)$  such that  $\theta(A, X_1) = X_2$ . Let  $X_1 = Y_1 Y_1^T$  and  $X_2 = Y_2 Y_2^T$  with  $Y_1, Y_2 \in \mathbb{R}_{*}^{n \times p}$ , then such an  $A$  is given by

$$A = [Y_2 \quad Z_2] [Y_1 \quad Z_1]^{-1}.$$

for any  $Z_1, Z_2 \in \mathbb{R}_{*}^{n \times (n-p)}$  with  $Z_1 \perp Y_1$  and  $Z_2 \perp Y_2$ .  $\square$

Take the following matrix

$$E := \begin{bmatrix} I_p & 0 \\ 0 & 0_{(n-p) \times (n-p)} \end{bmatrix} \in \mathrm{S}_+(p, n).$$

If we fix  $E$  as argument of  $\mathrm{S}_+(p, n)$  in  $\theta$ , then map

$$\theta_E : \mathrm{GL}_n \rightarrow \mathrm{S}_+(p, n), A \mapsto \theta(A, E)$$

is surjective. Its image is called an orbit and this orbit is exactly the set  $\mathrm{S}_+(p, n)$ .

### 3.2 Quotient manifold $\mathrm{GL}_n/\mathrm{Stab}_E$

Map  $\theta_E$  is surjective but not injective. In this section, we will construct a bijection based on  $\theta_E$  that gives us an alternative description of manifold  $\mathrm{S}_+(p, n)$ , namely as a quotient manifold.

An elegant way to express the many-to-one relation of  $\theta_E$  is with the so-called *stability group*, denoted by  $\mathrm{Stab}_E$ . This is the maximal subgroup of  $\mathrm{GL}_n$  that leaves the action of  $\theta_E$  fixed, that is  $\theta_E(L) = E$  for all  $L \in \mathrm{Stab}_E$ . It is not difficult to see that this group is given by

$$\mathrm{Stab}_E = \begin{bmatrix} \mathrm{O}_p & \mathbb{R}^{p \times (n-p)} \\ 0 & \mathrm{GL}_{n-p} \end{bmatrix}.$$

The many-to-one relation can now be factored out by means of the following equivalence relation for  $A, B \in \mathrm{GL}_n$ :

$$A \sim B \text{ if and only if } B = AL \text{ for some } L \in \mathrm{Stab}_E.$$

The equivalence class containing  $A$  is denoted by  $[A] := \{B \in \mathrm{GL}_n : A \sim B\}$ . Now, let

$$\mathrm{GL}_n/\mathrm{Stab}_E := \{[A] : A \in \mathrm{GL}_n\} \tag{2}$$

denote the quotient of  $\mathrm{GL}_n$  by this equivalence relation. It will be the set of all the equivalence classes of  $\sim$ . Map

$$\pi : \mathrm{GL}_n \rightarrow \mathrm{GL}_n/\mathrm{Stab}_E, \quad A \mapsto [A] \tag{3}$$

is the canonical projection or *quotient map*. In order to avoid ambiguity regarding to which space  $[A]$  belongs, we will use  $\pi(A)$  to denote  $[A]$  viewed as an abstract element of  $\mathrm{GL}_n/\mathrm{Stab}_E$ , and  $\pi^{-1}(\pi(A))$  for  $[A]$  as a subset of  $\mathrm{GL}_n$ .

It is a standard result for homogeneous spaces that  $\mathrm{GL}_n/\mathrm{Stab}_E$  is a quotient manifold diffeomorphic to  $\mathrm{S}_+(p, n)$ .

PROPOSITION 3.3 Map

$$\Theta_E : \mathrm{GL}_n/\mathrm{Stab}_E \rightarrow \mathrm{S}_+(p, n), \pi(A) \mapsto \theta_E(A)$$

is a diffeomorphism with  $\theta_E = \Theta_E \circ \pi$ . In other words,  $\mathrm{GL}_n/\mathrm{Stab}_E \simeq \mathrm{S}_+(p, n)$ .

*Proof.* Since  $\mathrm{S}_+(p, n)$  is a smooth manifold, we can use e.g. Boothby (1986, Th. IV.9.3). Alternatively, the fact that  $\mathrm{Stab}_E$  is a closed Lie subgroup of  $\mathrm{GL}_n$  makes the set  $\mathrm{S}_+(p, n)$  a smooth manifold, see Boothby (1986, Th. IV.9.6). The dimension of  $\mathrm{GL}_n/\mathrm{Stab}_E$  equals  $\dim(\mathrm{GL}_n) - \dim(\mathrm{Stab}_E)$  and is the same as  $\dim(\mathrm{S}_+(p, n))$ .  $\square$

### 3.3 Representatives for equivalence classes

Manifold  $\mathrm{GL}_n/\mathrm{Stab}_E$  is a quotient space that contains abstract equivalence classes as elements. Since we prefer to work with concrete matrices as elements, we will use a matrix  $A \in \mathrm{GL}_n$  to represent the equivalence class  $[A]$ . These representatives are obviously not unique and so they are not a replacement for the equivalence classes. However, with the right mathematical tools (like horizontal lifts) this will pose no problems. In fact, this non-uniqueness will be advantageous later on. Let us see how these representatives look like.

First note that throughout the paper we will use the following partitioning for an  $A \in \mathrm{GL}_n$ :

$$A = [Y \ Z], \quad Y \in \mathbb{R}_*^{n \times p}, Z \in \mathbb{R}_*^{n \times (n-p)}. \quad (4)$$

By using the notation  $Y$  and  $Z$  we will implicitly always mean matrices that satisfy the form of (4). For such an  $A$ , the equivalence class  $\pi^{-1}(\pi(A))$  can be written as

$$\begin{aligned} \pi^{-1}(\pi(A)) &= A \mathrm{Stab}_E, \\ &= [Y \mathrm{O}_p \quad Y \mathbb{R}^{(n-p) \times p} + Z \mathrm{GL}_{n-p}], \\ &= [Y \mathrm{O}_p \quad Y \mathbb{R}^{(n-p) \times p} + Y_\perp \mathrm{GL}_{n-p}]. \end{aligned}$$

As long as  $A \in \mathrm{GL}_n$ , the last  $n - p$  columns of  $A$  are all equivalent. So only the first  $p$  columns have an influence on  $\pi$  and these columns belong to the equivalence class  $Y \mathrm{O}_p$ .

### 3.4 Reductive space $\mathrm{GL}_n/\mathrm{O}_n$

When  $p = n$ , the homogeneous space geometry in Prop. 3.2 has an important additional property, namely that of a reductive space; see Smith (2005). There is now a natural choice for a metric on  $\mathrm{S}_+(n, n) \simeq \mathrm{GL}_n/\mathrm{O}_n$  and this metric will be invariant to the  $\mathrm{GL}_n$ -action. This invariance is a powerful property to exploit (e.g., when deriving geodesics) and has been used in most of the literature on the Riemannian geometry of  $\mathrm{S}_+(n, n)$ .

When  $p < n$ , it is sensible to try to find a metric on  $\mathrm{GL}_n/\mathrm{Stab}_E$  which is also invariant to the  $\mathrm{GL}_n$ -action. This is however not possible, as was shown in Bonnabel & Sepulchre (2009) for  $\mathrm{S}_+(2, 1)$  by a continuity argument on this metric. In fact, we will show that  $\mathrm{GL}_n/\mathrm{Stab}_E$  is never a reductive space for all rank deficient cases, i.e.,  $p < n$ . As such, we can no longer derive a  $\mathrm{GL}_n$ -invariant metric based on this reductive property.

Let  $\mathfrak{g} = \mathbb{R}^{n \times n}$  denote the Lie algebra of  $\mathrm{GL}_n$  and let

$$\mathfrak{h} = \begin{bmatrix} \mathcal{S}_p^{\text{skew}} & \mathbb{R}^{p \times (n-p)} \\ 0 & \mathbb{R}^{(n-p) \times (n-p)} \end{bmatrix}$$

be the Lie algebra of  $\mathrm{Stab}_E$ . The homogeneous space  $\mathrm{GL}_n/\mathrm{Stab}_E$  is reductive if and only if there exists a subspace  $\mathfrak{p}$  of  $\mathfrak{gl}$ , complementary to  $\mathfrak{h}$ , such that  $H\mathfrak{p}H^{-1} \subseteq \mathfrak{p}$  for all  $H \in \mathrm{Stab}_E$  (Kobayashi & Nomizu, 1963, Chap. X). Differentiating, this implies that

$$[\mathfrak{h}, \mathfrak{p}] \subseteq \mathfrak{p}. \quad (5)$$

All subspaces of  $\mathfrak{gl}$  complementary to  $\mathfrak{h}$  are of the form

$$\mathfrak{p} = \left\{ \begin{bmatrix} S + \mathcal{W}(B, S) & \mathcal{D}(B, S) \\ B & \mathcal{E}(B, S) \end{bmatrix} : B \in \mathbb{R}^{(n-p) \times p}, S \in \mathcal{S}_p^{\text{sym}} \right\},$$

where

$$\begin{aligned} \mathcal{W} : \mathbb{R}^{(n-p) \times p} \times \mathcal{S}_p^{\text{sym}} &\rightarrow \mathcal{S}_p^{\text{skew}} \\ \mathcal{D} : \mathbb{R}^{(n-p) \times p} \times \mathcal{S}_p^{\text{sym}} &\rightarrow \mathbb{R}^{p \times (n-p)} \\ \mathcal{E} : \mathbb{R}^{(n-p) \times p} \times \mathcal{S}_p^{\text{sym}} &\rightarrow \mathbb{R}^{(n-p) \times (n-p)} \end{aligned}$$

are linear maps. Since map  $\mathcal{E}$  is linear, we can decompose it as  $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$  with  $\mathcal{E}_1 : \mathbb{R}^{(n-p) \times p} \rightarrow \mathbb{R}^{(n-p) \times (n-p)}$ ,  $\mathcal{E}_2 : \mathcal{S}_p^{\text{sym}} \rightarrow \mathbb{R}^{(n-p) \times (n-p)}$ , and similarly for  $\mathcal{W}$  and  $\mathcal{D}$ .

Observe that the l.h.s. of condition (5) reduces to

$$\begin{aligned} [\mathfrak{h}, \mathfrak{p}] &= \begin{bmatrix} \Sigma & K \\ 0 & L \end{bmatrix} \begin{bmatrix} S + \mathcal{W} & \mathcal{D} \\ B & \mathcal{E} \end{bmatrix} - \begin{bmatrix} S + \mathcal{W} & \mathcal{D} \\ B & \mathcal{E} \end{bmatrix} \begin{bmatrix} \Sigma & K \\ 0 & L \end{bmatrix} \\ &= \begin{bmatrix} \Sigma S + \Sigma \mathcal{W} + KB - S\Sigma - \mathcal{W}\Sigma & \Sigma \mathcal{D} + K\mathcal{E} - SK - \mathcal{W}K - \mathcal{D}L \\ LB - B\Sigma & L\mathcal{E} - BK - \mathcal{E}L \end{bmatrix}, \end{aligned} \quad (6)$$

with  $\Sigma \in \mathcal{S}_p^{\text{skew}}$ ,  $K \in \mathbb{R}^{p \times (n-p)}$  and  $L \in \mathbb{R}^{(n-p) \times (n-p)}$ . The proof is now as follows.

**PROPOSITION 3.4** The homogeneous space  $\mathrm{GL}_n / \mathrm{Stab}_E \simeq \mathrm{S}_+(p, n)$  is reductive only for  $p = n$ .

*Proof.* We will show that for  $p < n$ , there are no linear maps  $\mathcal{W}, \mathcal{D}, \mathcal{E}$  in (6) such that (5) is satisfied for all  $\Sigma, S, K, L, B$  arbitrary matrices of suitable form.

First, if  $p = n$ , linear maps  $\mathcal{D}, \mathcal{E}$  and matrices  $K, L, B$  are void. We immediately get that  $\mathcal{W} = 0$  satisfies (5), as is known from Smith (2005).

Next, we prove the case  $p < n$  by contradiction. Take  $K = 0$  and  $\Sigma = 0$ , then condition (5) with (6) implies

$$\begin{bmatrix} 0 & -\mathcal{D}(B, S)L \\ LB & L\mathcal{E}(B, S) - \mathcal{E}(B, S)L \end{bmatrix} = \begin{bmatrix} \mathcal{W}(LB, 0) & \mathcal{D}(LB, 0) \\ LB & \mathcal{E}(LB, 0) \end{bmatrix}. \quad (7)$$

So one of the conditions is  $\mathcal{E}(LB, 0) = L\mathcal{E}(B, S) - \mathcal{E}(B, S)L$ , or  $\mathcal{E}_1(LB) = L\mathcal{E}_1(B) + L\mathcal{E}_2(S) - \mathcal{E}_1(B)L - \mathcal{E}_2(S)L$  for all  $L, B$  and  $S$ . Hence when  $B = 0$ , we get that  $L\mathcal{E}_2(S) - \mathcal{E}_2(S)L = 0$ , or, in other words, that  $\mathcal{E}_2(S)$  commutes with  $L$ . For general  $L$  and  $S$ , this can only be if  $\mathcal{E}_2(S) = \alpha(S)I_{n-p}$  where  $\alpha$  is a scalar-valued function. Now take  $L = 0$  and  $\Sigma = 0$ , then condition (5) with (6) satisfies

$$\begin{bmatrix} KB & K\mathcal{E}(B, S) - SK - \mathcal{W}(B, S)K \\ 0 & -BK \end{bmatrix} = \begin{bmatrix} X + \mathcal{W}(0, X) & \mathcal{D}(0, X) \\ 0 & \mathcal{E}(0, X) \end{bmatrix}, \quad (8)$$

where

$$X = (KB + B^T K^T)/2.$$

We get the condition  $\mathcal{E}(0, (KB + B^T K^T)/2) = -BK$ . From above, we already know that  $\mathcal{E}(0, (KB + B^T K^T)/2) = \alpha((KB + B^T K^T)/2)I_{n-p}$ . This leads to a contradiction if  $BK$  is not a scalar, i.e.,  $n - p > 1$ . If  $n - p = 1$ , then  $\alpha(S) = -\mathrm{tr}(S)$  satisfies the condition.

We continue with proving the case  $n - p = 1$ . Take again (7), since the lower-right corner of the l.h.s. is now zero, it implies  $\mathcal{E}(LB, 0) = \mathcal{E}_1(LB) = 0$  for all  $L, B$ . Hence  $\mathcal{E}_1 = 0$  and  $\mathcal{E}(B, S) = -\mathrm{tr}(S)$ . In addition (7) also implies  $\mathcal{D}(LB, 0) = -\mathcal{D}(B, S)L$ , and, since  $L$  is a scalar, also  $L\mathcal{D}_1(B) = -L\mathcal{D}_1(B) - L\mathcal{D}_2(S)$ . From this we have  $\mathcal{D} = 0$ . From (7) we have also  $\mathcal{W}(LB, 0) = 0$ , so  $\mathcal{W}_1 = 0$ . Finally, from (8) again, we get that  $-\mathrm{tr}(S)K - SK - \mathcal{W}_2(S)K = 0$  for all  $K$ . Since  $\mathcal{W}_2(S)$  is always skew-symmetric and  $\mathrm{tr}(S)I_p + S$  symmetric, their sum can never be zero and we have a contradiction.  $\square$

### 3.5 Riemannian metric

For  $p < n$ , manifold  $\mathrm{S}_+(p, n)$  is no longer reductive and we lose the existence of a natural metric. The typical choice to equip  $\mathrm{GL}_n$  with the Euclidean metric

$$\bar{g}_A^{\text{Eucl}}(\eta_A, \nu_A) := \mathrm{tr}(\eta_A^T \nu_A), \quad \text{for all } \eta_A, \nu_A \in T_A \mathrm{GL}_n, A \in \mathrm{GL}_n,$$

turns out to be less than ideal when one is concerned about geodesics. Take e.g. curve  $t \mapsto I_n - tI_n$ . It is obviously a length minimising geodesic, but it is not complete since at  $t = 1$  its image is zero. Complete geodesics are important however, since, by virtue of the Hopf–Rinow theorem (Boothby, 1986, Th. VII.7.7), they allow one to define a distance function on  $\mathrm{S}_+(p, n)$  as the minimal length of all connecting geodesics between two points.

Since all left- and right-invariant vector fields on  $\mathrm{GL}_n$  are complete (Boothby, 1986, Cor. V.5.8), it is reasonable to have a left- or right-invariant metric also. (Note that both left- and right-invariant is not possible on  $\mathrm{GL}_n$ .) In this paper, we therefore choose the right-invariant metric  $\bar{g}$ , defined as

$$\bar{g}_A(\eta_A, \nu_A) := \bar{g}_I(\eta_A A^{-1}, \nu_A A^{-1}) = \mathrm{tr}(A^{-T} \eta_A^T \nu_A A^{-1}), \quad \text{for all } \eta_A, \nu_A \in T_A \mathrm{GL}_n, A \in \mathrm{GL}_n. \quad (9)$$

In §4.1, we will explicitly show that, although the geodesics of  $(\mathrm{GL}_n, \bar{g})$  are not always right-invariant, they are still complete.

At first sight, an equally logical choice would be the left-invariant metric for which all left-invariant vector fields are complete. We will later see in §7.3 however that the left-invariant metric does not allow us to reuse the geodesics of  $\mathrm{GL}_n$  in the same way as the right-invariant does.

### 3.6 Tangent space

Similar to the representatives of §3.3, we want to use the tangent space of  $\mathrm{GL}_n$  to represent tangent vectors of  $\mathrm{GL}_n/\mathrm{Stab}_E$ . This requires the notion of the vertical and horizontal space: a specific decomposition of  $T_A \mathrm{GL}_n \simeq \mathbb{R}^{n \times n}$ .

Since  $\mathrm{Stab}_E$  is closed,  $\pi^{-1}(\pi(A))$  is a submanifold embedded in  $\mathrm{GL}_n$  (Boothby, 1986, Lemma IV.9.7). This means that its tangent space is a subspace of the embedding space  $\mathbb{R}^{n \times n}$  and it is called the *vertical space*  $\mathcal{V}_A$ . In  $A = [Y \ Z]$  it is given by

$$\mathcal{V}_A = [Y \ \mathcal{S}_p^{\text{skew}} \ \mathbb{R}^{n \times (n-p)}] = A \begin{bmatrix} \mathcal{S}_p^{\text{skew}} & \mathbb{R}^{p \times (n-p)} \\ 0 & \mathbb{R}^{(n-p) \times (n-p)} \end{bmatrix}. \quad (10)$$

The *horizontal space*  $\mathcal{H}_A$  is any complementary subspace of  $\mathcal{V}_A$  in  $\mathbb{R}^{n \times n}$  and we will take the orthogonal complement of  $\mathcal{V}_A$  w.r.t. the right-invariant metric. So, using the right-invariant metric  $\bar{g}$  of (9), the horizontal space becomes

$$\mathcal{H}_A = A^{-T} \begin{bmatrix} \mathcal{S}_p^{\text{sym}} & 0 \\ \mathbb{R}^{(n-p) \times p} & 0 \end{bmatrix} A^T A. \quad (11)$$

The tangent space of  $\mathrm{GL}_n/\mathrm{Stab}_E$  can now be represented uniquely by tangent vectors from the horizontal space. These are called the *horizontal lifts* and we will consistently denote the (horizontal) lift of  $\eta_{\pi(A)} \in T_{\pi(A)} \mathrm{GL}_n/\mathrm{Stab}_E$  by  $\bar{\eta}_A \in T_A \mathrm{GL}_n$ . Let  $\bar{\eta}$  denote such a unique horizontal lift on  $\mathrm{GL}_n$  of a vector field  $\eta$  on  $\mathrm{GL}_n/\mathrm{Stab}_E$ . It satisfies

$$D\pi(A)[\bar{\eta}_A] = \eta_{\pi(A)}, \quad \bar{\eta}_A \in \mathcal{H}_A. \quad (12)$$

We have in addition that the lifts are related along the equivalence class  $\pi^{-1}(\pi(A))$  in the following way.

**PROPOSITION 3.5** A horizontal vector field  $\bar{\eta}$  of  $\mathrm{GL}_n$  is the horizontal lift of a vector field  $\eta$  on  $\mathrm{GL}_n/\mathrm{Stab}_E$  if and only if, for all  $A \in \mathrm{GL}_n$ , it holds that

$$\bar{\eta}_{AL} = \bar{\eta}_A L, \quad \text{for all } L \in \mathrm{Stab}_E.$$

*Proof.* First observe that  $\mathcal{H}_I = L^{-T} \mathcal{H}_I L^T$  for all  $L \in \mathrm{Stab}_E$ , from which it follows that  $\mathcal{H}_{AL} = \mathcal{H}_A L$ . Thus, we have constructed a horizontal space that satisfies a connection Kobayashi & Nomizu (1963, Ch. II) on the principal bundle  $\mathrm{GL}_n(\mathrm{GL}_n/\mathrm{Stab}_E, \mathrm{Stab}_E)$  and the proof follows by Kobayashi & Nomizu (1963, Prop. II.1.2).  $\square$

We construct the space  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  by restricting the right-invariant metric of  $(\mathrm{GL}_n, \bar{g})$  to the horizontal space.

PROPOSITION 3.6 Let  $\bar{g}$  be the right-invariant metric (9). Then the relation

$$g_{\pi(A)}(\eta_{\pi(A)}, \nu_{\pi(A)}) = \bar{g}_A(\bar{\eta}_A, \bar{\nu}_A) = \text{tr}(A^{-T} \bar{\eta}_A^T \bar{\nu}_A A^{-1}) \quad (13)$$

defines a Riemannian metric  $g$  on  $\text{GL}_n/\text{Stab}_E$ . The metric  $g$  turns the quotient map

$$\pi : \text{GL}_n \rightarrow \text{GL}_n/\text{Stab}_E$$

into a Riemannian submersion and  $(\text{GL}_n/\text{Stab}_E, g)$  is a Riemannian quotient manifold of  $(\text{GL}_n, \bar{g})$ .

*Proof.* The lifted metric is invariant on each fibre

$$\bar{g}_{AL}(\bar{\eta}_{AL}, \bar{\nu}_{AL}) = \bar{g}_I(\bar{\eta}_{AL} L^{-1} A^{-1}, \bar{\nu}_{AL} L^{-1} A^{-1}) = \bar{g}_I(\bar{\eta}_A A^{-1}, \bar{\nu}_A A^{-1}) = \bar{g}_A(\bar{\eta}_A, \bar{\nu}_A),$$

hence  $D\pi(A)|_{\mathcal{H}_A}$  is an isometry for each  $A$ . This makes  $\pi$  a Riemannian submersion (O'Neill, 1983, Def. 7.44).  $\square$

### 3.7 Some useful expressions

We will derive some relations regarding the horizontal space which will be convenient later on. Take  $A = [Y \ Z]$  as in (4). We have the identity

$$A^{-T} = [\text{P}_Z^\perp Y(Y^T \text{P}_Z^\perp Y)^{-1} \quad \text{P}_Y^\perp Z(Z^T \text{P}_Y^\perp Z)^{-1}]. \quad (14)$$

The choice  $Z = Y_\perp$ , i.e.,

$$A = [Y \ Y_\perp], \quad Y \in \mathbb{R}_*^{n \times p}, Y_\perp \in \mathbb{R}_*^{n \times (n-p)}, Y^T Y_\perp = 0, Y_\perp^T Y_\perp = I_{n-p}, \quad (15)$$

allows to simplify the horizontal space to

$$\mathcal{H}_A = [Y(Y^T Y)^{-1} \mathcal{S}_p^{\text{sym}}(Y^T Y) + Y_\perp \mathbb{R}^{(n-p) \times p} \quad 0_{n \times (n-p)}]. \quad (16)$$

Still using  $Z = Y_\perp$ , the inner product of two tangent vectors  $\bar{\eta}_1, \bar{\eta}_2 \in \mathcal{H}_A$  can also be written succinctly. Suppose  $\bar{\eta}_1 = [Y(Y^T Y)^{-1} H_1(Y^T Y) + Y_\perp K_1 \ 0]$  and analogously for  $\bar{\eta}_2$ , then

$$\bar{g}_A(\bar{\eta}_1, \bar{\eta}_2) = \text{tr}((Y^T Y)^{-1} (H_1(Y^T Y) H_2 + K_1^T K_2)). \quad (17)$$

### 3.8 Orthogonal projection onto the horizontal space

Now that we have defined the metric, we can specify the projection onto the horizontal space orthogonal w.r.t. this metric. Since  $T_A \text{GL}_N \simeq \mathbb{R}^{n \times n} = \mathcal{V}_A \oplus \mathcal{H}_A$  for all  $A \in \text{GL}_n$ , we can define for every  $A \in \text{GL}_n$  the following orthogonal projections:

$$\text{P}^h : \mathbb{R}^{n \times n} \rightarrow \mathcal{H}_A, \quad \text{with } \text{P}^h(V) = 0 \text{ for all } V \in \mathcal{V}_A \text{ and } \text{P}^h(H) = H \text{ for all } H \in \mathcal{H}_A. \quad (18)$$

$$\text{P}^v : \mathbb{R}^{n \times n} \rightarrow \mathcal{V}_A, \quad \text{with } \text{P}^v(H) = 0 \text{ for all } H \in \mathcal{H}_A \text{ and } \text{P}^v(V) = V \text{ for all } V \in \mathcal{V}_A, \quad (19)$$

So we can decompose every  $Z \in \mathbb{R}^{n \times n}$  into a horizontal term  $\text{P}^h(Z)$  and a vertical term  $\text{P}^v(Z)$  for which  $\bar{g}_A(\text{P}^h(Z), \text{P}^v(Z)) = 0$  and  $Z = \text{P}^h(Z) + \text{P}^v(Z)$ .

These projections can be computed for all  $A$  and  $Z$  as oblique projections onto  $\mathcal{H}_A$  along  $\mathcal{V}_A$ , and vice versa. We will derive an efficient expression for these projections by exploiting the equivalence along the fibre  $[A]$ . First, we need a technical Lemma regarding the so-called stable and symmetric *generalized Lyapunov equation*. This equation is well-known in control theory, but we reformulate it for convenience.

LEMMA 3.7 Let  $S_1, S_2 \in \mathcal{S}_n^{\text{sym}}$  with  $S_1, S_2 \succ 0$  be given and define the generalised Lyapunov operator

$$\mathcal{L}_{S_1, S_2} : \mathcal{S}_n^{\text{sym}} \rightarrow \mathcal{S}_n^{\text{sym}}, \quad X \mapsto S_1 X S_2 + S_2 X S_1. \quad (20)$$

Then operator  $\mathcal{L}_{S_1, S_2}$  is linear and bijective. Furthermore, equation  $\mathcal{L}_{S_1, S_2}(X) = B$ , can be solved for  $X$  in  $O(n^3)$  time and memory.

*Proof.* See Penzl (1998, Th. 3).  $\square$

The choice  $A = [Y \ Y_\perp] \in \mathrm{GL}_n$  allows for a direct computation of the projections.

LEMMA 3.8 Let  $A = [Y \ Y_\perp] \in \mathrm{GL}_n$  be as in (15). The horizontal projection  $P^h$  from (18) of a tangent vector  $\eta_A = [\eta_1 \ \eta_2]$ , with the same partitioning as  $A$ , is given by

$$P^h(\eta_A) = [Y(Y^T Y)^{-1} H(Y^T Y) + P_Y^\perp \eta_1 \ 0_{n \times (n-p)}]$$

with  $H \in \mathcal{S}_p^{\text{sym}}$  the solution of the Lyapunov equation

$$(Y^T Y)^{-1} H(Y^T Y) + (Y^T Y) H(Y^T Y)^{-1} = (Y^T Y)^{-1} (Y^T \eta_1) + (\eta_1^T Y) (Y^T Y)^{-1}. \quad (21)$$

The vertical projection  $P^v$  from (19) is given by

$$P^v(\eta_A) = [P_Y \eta_1 - Y(Y^T Y)^{-1} H(Y^T Y) \ \eta_2].$$

*Proof.* We will give a constructive proof. The vertical and horizontal space are given by (10) and (11) respectively:

$$\mathcal{V}_A = [Y \ Y_\perp] \begin{bmatrix} \Omega & M \\ 0 & N \end{bmatrix} \quad \text{and} \quad \mathcal{H}_A = [Y \ Y_\perp] \begin{bmatrix} (Y^T Y)^{-1} H(Y^T Y) & 0 \\ K(Y^T Y) & 0 \end{bmatrix},$$

with  $\Omega, M, N, H$  and  $K$  coefficient matrices of suitable form. The horizontal projection of  $\eta_A$  must satisfy  $P^h(\eta_A) = \eta_A - v$ ,  $v \in \mathcal{V}_A$ . Writing the tangent vector as

$$\eta_A = [\eta_1 \ \eta_2] = [Y \ Y_\perp] \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix},$$

this condition reduces to

$$\begin{bmatrix} (Y^T Y)^{-1} H(Y^T Y) & 0 \\ K(Y^T Y) & 0 \end{bmatrix} = \begin{bmatrix} X_{11} - \Omega & X_{12} - M \\ X_{21} & X_{22} - N \end{bmatrix},$$

which immediately gives that  $K = X_{21}(Y^T Y)^{-1}$ ,  $M = X_{12}$  and  $N = X_{22}$ . The condition

$$(Y^T Y)^{-1} H(Y^T Y) = X_{11} - \Omega \quad \text{with } H = H^T \text{ and } \Omega = -\Omega^T$$

is solved by adding it to its transpose. This gives  $(Y^T Y)^{-1} H(Y^T Y) + (Y^T Y) H(Y^T Y)^{-1} = X_{11} + X_{11}^T$ . Since  $Y^T Y \succ 0$ , Lemma 3.7 guarantees a unique and symmetric solution  $H$ . This way, we have determined all the coefficients  $H, K$  of the projected matrix  $P^h(\eta_A)$ . Since  $\eta_1 = YX_{11} + Y_\perp X_{21}$ , it is straightforward to express the obtained matrix for  $P^h(\eta_A)$  into the form of the Lemma.  $\square$

Using the previous Lemma, we can compute the horizontal projection of a tangent vector  $\eta_A$  at an arbitrary  $A = [Y \ Z] \in \mathrm{GL}_n$  by transporting  $\eta_A$  along the fibre  $[A]$  to a point  $B = [Y \ Y_\perp]$ . Since  $B = AL$  for some  $L \in \mathrm{Stab}_E$ , it suffices to compute the projection of the tangent vector  $\eta_B = \eta_A L$  and transporting  $P^h(\eta_B)$  back to  $A$  by  $P^h(\eta_B)L^{-1}$ .

PROPOSITION 3.9 Let  $A = [Y \ Z] \in \mathrm{GL}_n$  be as in (4). The horizontal projection  $P^h$  from (18) of the tangent vector  $\eta_A = [\eta_1 \ \eta_2]$ , with the same partitioning as  $A$ , is given by

$$P^h(\eta_A) = [\eta_1^h \ \eta_1^h(Y^T Y)^{-1} Y^T Z]$$

with  $\eta_1^h = Y(Y^T Y)^{-1} H(Y^T Y) + P_Y^\perp \eta_1$  and  $H$  the solution of the Lyapunov equation (21).

*Proof.* Suppose we fix  $B = [Y \ Y_\perp]$  for some orthonormal  $Y_\perp$ , then we have  $BL = A$  with

$$L = \begin{bmatrix} I_k & (Y^T Y)^{-1} Y^T Z \\ 0 & Y_\perp^T Z \end{bmatrix}. \quad (22)$$

The transported tangent vector  $\eta_B = \eta_A L^{-1}$  has the same first  $p$  columns as  $\eta_A$ . By Lemma 3.8, we have as horizontal projection  $P^h(\eta_B) = [\eta_1^h \ 0]$ . Transporting this back to  $A$  gives the desired projection  $P^h(\eta_A) = P^h(\eta_B)L$   $\square$

Notice that we can compute the projections based only on  $Y$ , the first  $p$  columns of  $A$ , and we do not need to construct nor use  $Y_\perp$ .

### 3.9 Levi-Civita connection

We will use the Levi-Civita connection on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$ , denoted by  $\nabla$ . Since this connection can be related to the one on  $(\mathrm{GL}_n, \bar{g})$ , denoted by  $\bar{\nabla}$ , we will first derive this connection for two arbitrary vector fields.

**PROPOSITION 3.10** Let  $\eta, \nu$  be two vector fields on  $\mathrm{GL}_n$ , then the Levi-Civita connection of  $(\mathrm{GL}_n, \bar{g})$  in  $A \in \mathrm{GL}_n$  satisfies

$$(\bar{\nabla}_\nu \eta)(A) = D\eta(A)[\nu] + \frac{1}{2} \{ [A^{-T} \eta^T, \nu A^{-1}] A + [A^{-T} \nu^T, \eta A^{-1}] A - \eta A^{-1} \nu - \nu A^{-1} \eta \}. \quad (23)$$

*Proof.* Let  $\eta, \nu, \lambda$  be vector fields on  $\mathrm{GL}_n$ . Notice that since  $\mathrm{GL}_n$  is a vector space, one has  $[\nu, \eta] = D\eta[\nu] - D\nu[\eta]$ , and likewise for all permutations between  $\eta, \nu$  and  $\lambda$ . Furthermore, we have the identity

$$\begin{aligned} D\bar{g}_A(\eta, \lambda)[\nu] &= \mathrm{tr}(A^{-T} \eta^T D\lambda[\nu] A^{-1}) + \mathrm{tr}(A^{-T} \lambda^T D\eta[\nu] A^{-1}) \\ &\quad - \mathrm{tr}(A^{-T} \eta^T \lambda A^{-1} \nu A^{-1}) - \mathrm{tr}(A^{-T} \lambda^T \eta A^{-1} \nu A^{-1}) \end{aligned}$$

and again for all permutations. Substituting these identities in Koszul's formula (Lee, 1997, (5.1)),

$$\begin{aligned} 2\bar{g}_A(\bar{\nabla}_\nu \eta, \lambda) &= D\bar{g}_A(\eta, \lambda)[\nu] + D\bar{g}_A(\nu, \lambda)[\eta] - D\bar{g}_A(\eta, \nu)[\lambda] \\ &\quad + \bar{g}_A(\lambda, [\nu, \eta]) + \bar{g}_A(\eta, [\lambda, \nu]) - \bar{g}_A(\nu, [\eta, \lambda]), \end{aligned}$$

we obtain

$$\begin{aligned} 2\mathrm{tr}(A^{-T} \lambda^T \nabla_\nu \eta A^{-1}) &= 2\mathrm{tr}(A^{-T} \lambda^T D\eta[\nu] A^{-1} - A^{-T} \eta^T \lambda A^{-1} \nu A^{-1} - A^{-T} \lambda^T \eta A^{-1} \nu A^{-1} \\ &\quad - A^{-T} \nu^T \lambda A^{-1} \eta A^{-1} - A^{-T} \lambda^T \nu A^{-1} \eta A^{-1} + A^{-T} \eta^T \nu A^{-1} \lambda A^{-1} + A^{-T} \nu^T \eta A^{-1} \lambda A^{-1}). \end{aligned}$$

This identity holds for all  $\lambda$ , hence we recover (23).  $\square$

**REMARK 1** Despite the widespread use of  $\mathrm{GL}_n$ , to the best of our knowledge, the expression of the connection for arbitrary vector fields on  $(\mathrm{GL}_n, \bar{g})$  is new. For left- or right-invariant vector fields however, the connection simplifies considerably since it is again left- or right-invariant. This has been observed by many authors; see, e.g., Cheeger & Ebin (1975, Prop. 3.18) and Mahony (1994, Ex. 5.8.3).

Since  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  is a Riemannian quotient manifold, the connection on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  can be expressed in terms of horizontal lifts. Recall that the unique horizontal lift of a vector field  $\eta$  is denoted by  $\bar{\eta}$ .

**PROPOSITION 3.11** Let  $\nabla$  and  $\bar{\nabla}$  denote the connections on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  and  $(\mathrm{GL}_n, \bar{g})$  respectively. Then for all vector fields  $\nu, \eta$  on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$ , the horizontal lift of  $\nabla_\nu \eta$  satisfies

$$\bar{\nabla}_\nu \bar{\eta} = \bar{\nabla}_{\bar{\nu}} \bar{\eta} - \frac{1}{2} P^v[\bar{\nu}, \bar{\eta}] = P^h(\bar{\nabla}_{\bar{\nu}} \bar{\eta}),$$

where  $P^v$  and  $P^h$  denote the orthogonal projections (19)–(18).

*Proof.* The first identity follows from (3.25) in Cheeger & Ebin (1975), while the second is stated in O'Neill (1983, Lemma 7.45).  $\square$

## 4 Geodesics

According to Prop. 3.6, the quotient map  $\pi : \mathrm{GL}_n \rightarrow \mathrm{GL}_n/\mathrm{Stab}_E$  is a Riemannian submersion. This means that we can identify all the geodesics on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  as the geodesics on  $\mathrm{GL}_n$  for the right-invariant metric  $\bar{g}$  that stay horizontal; see, e.g., O'Neill (1983, Cor. 7.46) or Gallot *et al.* (2004, Prop. 2.109). In this section, we will therefore first derive the geodesics of  $(\mathrm{GL}_n, \bar{g})$ , restrict them to the horizontal space  $\mathcal{H}_A$  and finally formulate and solve them in a closed form.

### 4.1 Geodesics of $(\mathrm{GL}_n, \bar{g})$

The geodesics of  $(\mathrm{GL}_n, \bar{g})$  satisfy the following initial value problem (IVP).

LEMMA 4.1 Let  $A_0 \in \mathrm{GL}_n$  and  $U_0 \in \mathbb{R}^{n \times n}$  be given. Let  $\bar{g}$  be the metric (9). Then the geodesic in  $(\mathrm{GL}_n, \bar{g})$  through  $A_0$  along  $U_0 A_0$  is the solution  $A(t)$  of the IVP

$$\dot{A}(t) = U(t)A(t), \quad A(0) = A_0, \quad (24)$$

$$\dot{U}(t) = U(t)U(t)^T - U(t)^T U(t), \quad U(0) = U_0, \quad (25)$$

with  $U(t) \in \mathbb{R}^{n \times n}$ . Furthermore,  $A(t)$  is defined for all  $t \in \mathbb{R}$ .

*Proof.* We apply the Euler–Lagrange formalism to the length functional (or strictly speaking, the energy Lagrangian). Since we can (formally) write  $A(t)$  as the solution of the initial value problem  $\dot{A}(t) = U(t)A(t)$  with  $A(0) = A_0$  for some  $U(t) \in \mathbb{R}^{n \times n}$ , this functional is given by

$$S(U) := \int_0^1 \bar{g}_{A(t)}(\dot{A}(t), \dot{A}(t)) dt = \int_0^1 \bar{g}_I(U(t), U(t)) dt.$$

The extremals of  $S$  will be exactly the geodesics  $A(t)$  (Postnikov, 2001, Prop. 11.1). A calculus of variations then gives that  $U(t)$  has to satisfy (25) in order for  $S(U)$  to be stationary. Since  $d(\mathrm{tr}(UU^T))/dt = 2\mathrm{tr}(\dot{U}U) = 0$ , matrix  $U(t)$  has constant Frobenius norm. The differential equations are thus Lipschitz on  $\mathbb{R}$  and the solution of the IVP exists and is unique on the whole real line for all initial conditions (Stoer & Bulirsch, 1992, Th. 7.1.1.).  $\square$

*Alternative proof.* Curve  $A(t)$  is a geodesic if and only if  $(\bar{\nabla}_{\dot{A}} \dot{A})(A(t)) = 0$  for all  $t$ . Since

$$D\dot{A}(A(t))[\dot{A}] = d^2 A/dt^2,$$

formula (23) for the connection becomes

$$(\bar{\nabla}_{\dot{A}} \dot{A})(A(t)) = d^2 A/dt^2 + [A^{-T} \dot{A}^T, \dot{A} A^{-1}]A - \dot{A} A^{-1} \dot{A}.$$

Taking the derivative of (24) and using (25), we obtain  $(\bar{\nabla}_{\dot{A}} \dot{A})(A(t)) = 0$ .  $\square$

REMARK 2 In the same way geodesics for the left-invariant metric have been proved in Lee *et al.* (2007). Another proof based on a more general framework can be found in Miller *et al.* (2003).

We derive the closed-form solution of the IVP in Lemma. 4.1. First observe that IVP (25) is actually a Lax pair with solution

$$U(t) := Q(t)U_0Q(t)^T, \quad \text{with } Q(t) := \exp(t(U_0 - U_0^T)) \in \mathrm{O}_n.$$

This is easily verified using the relation  $\dot{U}(t) = Q(U_0 - U_0^T)U_0Q^T + QU_0(U_0^T - U_0)Q^T$ . Hence,  $A(t)$  is the solution of the IVP

$$\dot{A}(t) = Q(t)U_0Q(t)^T A(t), \quad A(0) := A_0 \in \mathrm{GL}_n.$$

Take  $B(t) := Q(t)^T A(t)$ , then

$$\dot{B}(t) = (U_0^T - U_0)Q(t)^T A(t) + U_0 Q(t)^T A(t) = U_0^T B(t).$$

Together with  $B(0) := A_0$ , this gives  $B(t) = \expm(tU_0^T)A_0$ , and so we have basically proved the following result.

**PROPOSITION 4.2** The geodesics of  $(\mathrm{GL}_n, \bar{g})$  are given by

$$A(t) = \expm(t(U_0 - U_0^T)) \expm(tU_0^T)A_0,$$

for all  $A_0 \in \mathrm{GL}_n$  and  $U_0 \in \mathbb{R}^{n \times n}$ . They are complete.

*Proof.* The fact that  $A(t)$  is a geodesic follows by the construction above. Since the image of the matrix exponential is always a full-rank matrix (Boothby, 1986, Th. IV.6.2), matrix  $A(t)$  is well-defined and will be in  $\mathrm{GL}_n$  for all  $t$ .  $\square$

Let  $\gamma_I(t)$  be a geodesic of Prop. 4.2 with foot  $\gamma_I(0) = I$ . Then for any  $A \in \mathrm{GL}_n$ , the curve  $\gamma_A(t) := \gamma_I(t)A$  is also a geodesic. Furthermore, their velocity vector fields are related by right-translation  $R_A : \mathrm{GL}_n \rightarrow \mathrm{GL}_n$ ,  $X \mapsto XA$  since

$$dR_A(\dot{\gamma}_I(t)) = \dot{\gamma}_I(t)A \stackrel{!}{=} \dot{\gamma}_A(t),$$

with  $dR_A := R_A$  the differential of  $R_A$ . Such vector fields are called *right-related*. They are however not always *right-invariant*, or in other words, their flow does not need to be same after right translation. Take again  $\gamma_I(t)$  at  $t = 0$ . Now we right-translate  $\gamma_I(t)$  to itself, i.e.,  $R_A := R_{\gamma_I(T)}$  for some  $t = T$ . Right-invariance would mean equality for each  $T$  in

$$dR_A(\dot{\gamma}_I(0)) = \dot{\gamma}_I(0)\gamma_I(T) = U_0\gamma_I(T) \stackrel{?}{=} \dot{\gamma}_I(T) = U(T)\gamma_I(T).$$

We have equality when  $U(t) = U_0$  stays constant, e.g., when  $U_0$  is a normal matrix. In this case the geodesic is simply a matrix exponential  $t \mapsto \expm(tU_0)A_0$ , which is obviously right-invariant.

There is another property which depends on the normality of  $U_0$ . Let  $n > 1$  and let  $\gamma_I(t)$  again be a geodesic with  $\gamma_I(0) = I$ . Since there are initial conditions  $U_0$  for which

$$\gamma_I(t+s) \neq \gamma_I(t)\gamma_I(s) \neq \gamma_I(s)\gamma_I(t),$$

the geodesics do not form a one-parameter subgroup in general. However, when  $U_0$  is a normal matrix, the matrix exponentials in Prop. 4.2 commute and the geodesics can be written as

$$t \mapsto \expm(tU_0)A_0. \tag{26}$$

This is e.g. the case for (skew-)symmetric and orthogonal matrices  $U_0$ . In fact, one can show, together with the formula for the Riemannian connection, that the normality condition  $[U_0, U_0^T] = 0$  captures all the cases for which the geodesics are one-parameter subgroups (Cheeger & Ebin, 1975, Prop. 3.18). Since there is a one-to-one correspondence between one-parameters subgroups in  $\mathrm{GL}_n$  and  $\expm$  (Boothby, 1986, Cor. IV.6.3), the curves (26) will be geodesics if and only if  $U_0$  is normal. We will later see in Prop. 4.3 that the case for general  $U_0$  is necessary to have meaningful horizontal geodesics on  $\mathrm{GL}_n/\mathrm{Stab}_E$ .

**REMARK 3** There are other affine connections for which the curves (26) do describe all the geodesics, most notably the Cartan connections (Postnikov, 2001, §6.4). However, these connections do not share some important properties of the Levi–Civita connection like geodesics that are length-minimising. For these reasons, we prefer to work with the Levi–Civita connection.

**REMARK 4** Geodesics on Lie groups for the right-invariant metric are widespread in Lagrangian and symplectic dynamics; see, e.g., Marsden & Tudor (1999). In this field, the geodesics are usually derived based on the Euler–Arnold formalism: the Euler–Poincaré equations lead to an IVP for the intrinsic velocity  $U(t)$  as the Lax pair (25); the IVP for the geodesic follows directly by the definition of  $U(t) = \dot{A}(t)A(t)^{-1}$ . Since these derivations require some more involved differential geometry, we prefer our more constructive approach.

## 4.2 Horizontal geodesics of $(\mathrm{GL}_n, \bar{g})$

By restricting a geodesic on  $(\mathrm{GL}_n, \bar{g})$  to stay horizontal we obtain a representative of a geodesic on the abstract manifold  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$ . At first sight, these geodesics on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  appear to be easily found, since the geodesics of  $(\mathrm{GL}_n, \bar{g})$  are right-related and available in closed form. This is however not as useful as it seems, since the horizontal space does not share the same right relation and the matrix exponentials involve large  $n \times n$  matrices. We will therefore derive, in this and the next two sections, an alternative IVP for the horizontal geodesics that allows us to solve the first  $p$  columns of a geodesic efficiently.

**PROPOSITION 4.3** Let  $A_0 \in \mathrm{GL}_n$ ,  $H_0 \in \mathcal{S}_p^{\text{sym}}$  and  $K_0 \in \mathbb{R}^{(n-p) \times p}$  be given and let  $A(t)$  be the solution of the initial value problem

$$\dot{A}(t) = A(t)^{-T} \begin{bmatrix} H_0 & 0 \\ K_0 & 0 \end{bmatrix} A(t)^T A(t), \quad A(0) = A_0. \quad (27)$$

Then the complete geodesic on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  through  $\pi(A_0)$  along  $\eta_{\pi(A_0)}$  with horizontal lift  $\bar{\eta}_{A_0} := \dot{A}(0)$  is given by  $\pi(A(t))$  for all  $t \in \mathbb{R}$ .

*Proof.* First observe that since  $\dot{A}(t) \in \mathcal{H}_{A(t)}$ , the curve  $A(t)$  stays horizontal in  $(\mathrm{GL}_n, \bar{g})$ . By Gallot *et al.* (2004, Prop. 2.109),  $\pi(A(t))$  will be a complete geodesic on  $(\mathrm{GL}_n/\mathrm{Stab}_E, g)$  if  $A(t)$  is a complete geodesic in  $(\mathrm{GL}_n, \bar{g})$ . Identifying  $A(t)$  in (27) as the curve  $A(t)$  of (24) we see that  $U(t)$  has to satisfy

$$U(t) = A(t)^{-T} \begin{bmatrix} H_0 & 0 \\ K_0 & 0 \end{bmatrix} A(t)^T$$

in order that  $A(t)$  satisfies the ODE of Lemma 4.1. Taking the derivative,

$$\dot{U}(t) = -A(t)^{-T} \dot{A}(t)^T A(t)^{-T} \begin{bmatrix} H_0 & 0 \\ K_0 & 0 \end{bmatrix} A(t)^T + A(t)^{-T} \begin{bmatrix} H_0 & 0 \\ K_0 & 0 \end{bmatrix} \dot{A}(t)^T,$$

we see that  $\dot{U}(t)$  also satisfies (25). Thus,  $A(t)$  is a geodesic in  $(\mathrm{GL}_n, \bar{g})$ .  $\square$

The following corollary is a simple consequence of the partitioning in (4) and identity (14).

**COROLLARY 4.4** Let  $A(t) := [Y(t) \ Z(t)]$  be a geodesic as defined in Prop. 4.3 and be partitioned like (4). Then the matrices  $Y(t)$  and  $Z(t)$  are solutions of the IVP

$$\begin{aligned} \dot{Y} &= P_Z^\perp Y(Y^T P_Z^\perp Y)^{-1} H_0 (Y^T Y) + P_Y^\perp Z(Z^T P_Y^\perp Z)^{-1} K_0 (Y^T Y), & Y(0) &= Y_0, \\ \dot{Z} &= P_Z^\perp Y(Y^T P_Z^\perp Y)^{-1} H_0 (Y^T Z) + P_Y^\perp Z(Z^T P_Y^\perp Z)^{-1} K_0 (Y^T Z), & Z(0) &= Z_0. \end{aligned}$$

## 4.3 Moving the geodesics along the fibre

Take the usual partitioning  $A(t) = [Y(t) \ Z(t)]$ . Since in the end we are only interested in  $Y$ , we are free to pick another representative of  $\pi(A)$  such that matrix  $Z$  is of a more suitable form. We choose  $Z$  to be orthogonal to  $Y$ . This can always be done, since  $A$  is of full rank. The aim of this section is to transform the IVP of Cor. 4.4 into a more suited “triangular” form by moving  $Z(t)$  along the equivalence class  $[A(t)]$ . This will introduce a new variable, called  $W(t)$ .

At  $t = 0$ , we can simply take an  $A(0) := [Y_0 \ Z_0]$  such that  $Y_0 \perp Z_0$ . For  $t > 0$  however,  $Z(t)$  will not stay orthogonal to  $Y(t)$ . We will therefore introduce a new variable

$$W := P_Y^\perp Z = Z - Y(Y^T Y)^{-1} Y^T Z \quad (28)$$

and derive equations of motion in terms of only  $Y$  and  $W$ . Since  $W \perp Y$ , it will turn out that the projectors in Cor. 4.4 disappear and that the new ODE is of more suited form.

Let us first introduce the operator  $E : (\mathcal{S}_p^{\text{sym}}, \mathbb{R}) \rightarrow \mathcal{S}_p^{\text{sym}}$  defined as

$$E(S, t) = Q \operatorname{diag}(d_i) Q^T, \quad \text{with} \quad d_i = \begin{cases} t, & \lambda_i = 0, \\ (e^{t\lambda_i} - 1)/\lambda_i, & \lambda_i \neq 0, \end{cases}$$

where  $S = Q \operatorname{diag}(\lambda_i) Q^T$  is an eigenvalue decomposition. Observe that  $\lambda_i = 0$  is a removable singularity and  $dE(S, t)/dt = \expm(tS)$ . The following identities can be proved directly from Cor. 4.4.

LEMMA 4.5 Suppose  $Y_0 \perp Z_0$ , then  $Y(t)$  and  $Z(t)$  of Cor. 4.4 satisfy

$$Y(t)^T Y(t) = \expm(tH_0) Y_0^T Y_0 \expm(tH_0), \quad (29)$$

$$Y(t)^T Z(t) = \expm(tH_0) Y_0^T Y_0 E(H_0, t) K_0^T. \quad (30)$$

*Proof.* First observe that from Cor. 4.4, we get that  $Y^T \dot{Y} = H_0 Y^T Y$ ,  $Y^T \dot{Z} = H_0 Y^T Z$  and  $\dot{Y}^T Z = Y^T Y K_0^T$ . Next, we have

$$\frac{d(Y^T Y)}{dt} = \dot{Y}^T Y + Y^T \dot{Y} = (Y^T Y) H_0 + H_0 (Y^T Y), \quad Y(0)^T Y(0) = Y_0^T Y_0.$$

So (29) is indeed the solution of this initial value problem. In addition, we get

$$\frac{d(Y^T Z)}{dt} = \dot{Y}^T Z + Y^T \dot{Z} = \expm(tH_0) Y_0^T Y_0 \expm(tH_0) K_0^T + H_0 (Y^T Z), \quad Y(0)^T Z(0) = 0.$$

After substituting  $X := \expm(-tH_0) Y^T Z$  we get  $\dot{X} = Y_0^T Y_0 \expm(tH_0) K_0^T$  with  $X(0) = 0$ . Its solution is  $X(t) = Y_0^T Y_0 E(H_0, t) K_0^T$  and we recover (30) as solution.  $\square$

The equation of motion for  $W$  can be found by taking the derivative of (28). This gives

$$\dot{W} = \dot{Z} - \dot{Y}(Y^T Y)^{-1} Y^T Z - Y \frac{d}{dt}((Y^T Y)^{-1} Y^T Z).$$

According to Cor. 4.4, we have that  $\dot{Y}(Y^T Y)^{-1} (Y^T Z) = \dot{Z}$ . Together with Lemma 4.5, we can simplify the expression for  $\dot{W}$  to

$$\begin{aligned} \dot{W} &= -Y \frac{d}{dt}(\expm(-tH_0) E(H_0, t) K_0^T) = Y(H_0 \expm(-tH_0) E(H_0, t) - I_p) K_0^T \\ &= -Y \expm(-tH_0) K_0^T. \end{aligned} \quad (31)$$

We have in addition the identity

$$Z = W + Y \expm(-tH_0) E(H_0, t) K_0^T. \quad (32)$$

Next, we rewrite  $\dot{Y}$  in terms of  $W$  instead of  $Z$ . For this we need the following technical result.

LEMMA 4.6 Suppose  $Y_0 \perp Z_0$ , then  $Y(t)$  and  $Z(t)$  of Cor. 4.4 satisfy

$$P_Z^\perp Y(Y^T P_Z^\perp Y)^{-1} H_0 + P_Y^\perp Z(Z^T P_Y^\perp Z)^{-1} K_0 = Y(Y^T Y)^{-1} H_0 + W(W^T W)^{-1} K_0 \expm(-tH_0)$$

with  $W := P_Y^\perp Z$ .

*Proof.* Matrix  $A := [Y \ Z]$  will be full rank for any  $Y$  and  $Z$  of Cor. 4.4, so it suffices to verify that

$$\begin{aligned} A^T \{P_Z^\perp Y(Y^T P_Z^\perp Y)^{-1} H_0 + P_Y^\perp Z(Z^T P_Y^\perp Z)^{-1} K_0\} \\ \stackrel{?}{=} A^T \{Y(Y^T Y)^{-1} H_0 + W(W^T W)^{-1} K_0 \expm(-tH_0)\}. \end{aligned} \quad (33)$$

Working out  $A^T$  in partitioned form and using that  $Y \perp W$ , this gives

$$\begin{aligned} Y^T P_Z^\perp Y (Y^T P_Z^\perp Y)^{-1} H_0 &\stackrel{!}{=} Y^T Y (Y^T Y)^{-1} H_0, \\ Z^T P_Y^\perp Z (Z^T P_Y^\perp Z)^{-1} K_0 &\stackrel{?}{=} Z^T Y (Y^T Y)^{-1} H_0 + Z^T W (W^T W)^{-1} K_0 \exp(-tH_0). \end{aligned}$$

The first equality is trivially satisfied. Plugging in  $Z$  as given by (32), the second expression becomes

$$K_0 \stackrel{?}{=} K_0 (E(H_0, t) \exp(-tH_0) H_0 + \exp(-tH_0)).$$

It is straightforward to check that  $E(H_0, t) \exp(-tH_0) H_0 + \exp(-tH_0) \stackrel{!}{=} I_p$ .  $\square$

Now we can rewrite  $\dot{Y}$  directly as

$$\dot{Y} = Y(Y^T Y)^{-1} H_0 Y^T Y + W(W^T W)^{-1} K_0 \exp(-tH_0) Y^T Y. \quad (34)$$

Summarising all the transformations, we have almost proved the following result.

**LEMMA 4.7** Let  $A(t) = [Y(t) \ Z(t)]$  be defined as in Prop. 4.3 with the partitioning (4). In addition, assume that  $Y(0) \perp Z(0)$ , then  $Y(t)$  is the solution of the initial value problem

$$\dot{Y}(t) = (Y(t) \exp(-tH_0) (Y_0^T Y_0)^{-1} H_0 + W(t) (Z_0^T Z_0)^{-1} K_0) Y_0^T Y_0 \exp(tH_0), \quad Y(0) = Y_0, \quad (35)$$

$$\dot{W}(t) = -Y(t) \exp(-tH_0) K_0^T, \quad W(0) = Z_0, \quad (36)$$

with  $W(t) \in \mathbb{R}^{n \times (n-p)}$ . Furthermore,  $Z(t)$  is given by

$$Z(t) = W(t) + Y(t) \exp(-tH_0) E(H_0, t) K_0^T. \quad (37)$$

*Proof.* Equations (36) and (37) follow from the construction above. Observe that from (31) it follows that  $d(W^T W)/dt = 0$  and so  $W(t)^T W(t) = Z_0^T Z_0$ . Plugging this and (29) into (34), we obtain (35).  $\square$

#### 4.4 Closed-form solution of the geodesics

Now we are ready to solve the geodesics in closed form. First, we make the assumption that  $Z_0$  is an orthonormal basis for the orthogonal component of  $Y_0$ , so  $Z_0 := Y_{\perp 0}$  and  $Z_0^T Z_0 = I_{n-p}$ . This can always be done by transporting the foot of the geodesic along the equivalence class  $[A]$ . Next, we introduce the new variable

$$\tilde{Y}(t) := Y(t) \exp(-tH_0) (Y_0^T Y_0)^{-1/2}.$$

With  $(Y_0^T Y_0)^{1/2}$  we denote the unique principal matrix square root of  $Y_0^T Y_0$ . Since  $Y_0^T Y_0 \succ 0$ , this square root is symmetric positive definite and it satisfies  $(Y_0^T Y_0)^{1/2} (Y_0^T Y_0)^{1/2} = Y_0^T Y_0$ . It can be easily computed by means of an eigenvalue decomposition, see Higham (2008). Taking the derivative of  $\tilde{Y}(t)$ , we obtain an IVP equivalent to the one in Lemma 4.7. Let  $\Omega_0$  denote the following skew-symmetric matrix:

$$\Omega_0 := (Y_0^T Y_0)^{-1/2} H_0 (Y_0^T Y_0)^{1/2} - (Y_0^T Y_0)^{1/2} H_0 (Y_0^T Y_0)^{-1/2}.$$

Then the new IVP is homogeneous and linear with constant coefficients:

$$\frac{d}{dt} \begin{bmatrix} \tilde{Y} & W \end{bmatrix} = \begin{bmatrix} \tilde{Y} & W \end{bmatrix} \begin{bmatrix} \Omega_0 & -(Y_0^T Y_0)^{1/2} K_0^T \\ K_0 (Y_0^T Y_0)^{1/2} & 0 \end{bmatrix}, \quad \begin{aligned} \tilde{Y}(0) &= Y_0 (Y_0^T Y_0)^{-1/2}, \\ W(0) &= Y_{\perp 0}. \end{aligned} \quad (38)$$

Its solution is

$$\begin{bmatrix} \tilde{Y}(t) & W(t) \end{bmatrix} = \begin{bmatrix} Y_0 (Y_0^T Y_0)^{-1/2} & Y_{\perp 0} \end{bmatrix} \exp \left( t \begin{bmatrix} \Omega_0 & -(Y_0^T Y_0)^{1/2} K_0^T \\ K_0 (Y_0^T Y_0)^{1/2} & 0 \end{bmatrix} \right).$$

Although this closed-form solution is straightforward to compute, it is rather expensive since it involves the matrix exponential of an  $n \times n$  matrix. Furthermore, we would like to avoid having to use matrices  $Y_{\perp 0}$  and  $K_0$  explicitly.

Thanks to the choice  $Z_0 := Y_{\perp 0}$ , the horizontal lift of the initial velocity  $\eta_{\pi(A_0)}$  of a geodesic can be written like in (16):

$$\bar{\eta}_{A_0} = [V \ 0], \quad V = Y_0(Y_0^T Y_0)^{-1} H_0(Y_0^T Y_0) + Y_p \in \mathbb{R}^{n \times p}, \quad Y_p \perp Y_0. \quad (39)$$

We can identify matrix  $V$  as  $\dot{Y}(0)$  in Lemma 4.7. This results in the identity  $Y_p = Y_{\perp 0} K_0 Y_0^T Y_0$ . Substituting  $\widetilde{W}(t) := W(t) K_0 (Y_0^T Y_0)^{1/2}$  in (38) we get an IVP without  $Y_{\perp 0}$  and  $K_0$ :

$$\frac{d}{dt} \begin{bmatrix} \widetilde{Y} & \widetilde{W} \end{bmatrix} = \begin{bmatrix} \widetilde{Y} & \widetilde{W} \end{bmatrix} \begin{bmatrix} \Omega_0 & -\Sigma_0 \\ I_p & 0 \end{bmatrix}, \quad \begin{aligned} \widetilde{Y}(0) &= Y_0(Y_0^T Y_0)^{-1/2}, \\ \widetilde{W}(0) &= Y_p(Y_0^T Y_0)^{-1/2}. \end{aligned} \quad (40)$$

where  $\Sigma_0 := (Y_0^T Y_0)^{-1/2} Y_p^T Y_p (Y_0^T Y_0)^{-1/2}$ . Now we can summarize all the transformations in the following proposition. For ease of exposition, we formulate it stand-alone.

**PROPOSITION 4.8** Let  $Y_0 \in \mathbb{R}_*^{n \times p}$ ,  $H_0 \in \mathcal{S}_p^{\text{sym}}$  and  $Y_p \in \mathbb{R}^{n \times p}$  with  $Y_p^T Y_0 = 0$  be given. Define  $\Omega_0 := (Y_0^T Y_0)^{-1/2} H_0 (Y_0^T Y_0)^{1/2} - (Y_0^T Y_0)^{1/2} H_0 (Y_0^T Y_0)^{-1/2}$ ,  $\Sigma_0 := (Y_0^T Y_0)^{-1/2} Y_p^T Y_p (Y_0^T Y_0)^{-1/2}$  and

$$\begin{bmatrix} X_{11}(t) & X_{12}(t) \\ X_{21}(t) & X_{22}(t) \end{bmatrix} := \text{expm} \left( t \begin{bmatrix} \Omega_0 & -\Sigma_0 \\ I_p & 0 \end{bmatrix} \right).$$

Then the geodesic of  $(\text{GL}_n/\text{Stab}_E, g)$  through  $\pi([Y_0 \ Y_{\perp 0}])$  along  $\eta$  with horizontal lift

$$\bar{\eta} := [Y_0(Y_0^T Y_0)^{-1} H_0(Y_0^T Y_0) + Y_p \ 0_{n \times (n-p)}]$$

is the curve  $t \mapsto \pi([Y(t) \ Y_{\perp}(t)])$  where

$$Y(t) := \left( Y_0(Y_0^T Y_0)^{-1/2} X_{11}(t) + Y_p(Y_0^T Y_0)^{-1/2} X_{21}(t) \right) (Y_0^T Y_0)^{1/2} \text{expm}(t H_0).$$

*Proof.* The solution of (40) is

$$\begin{bmatrix} \widetilde{Y}(t) & \widetilde{W}(t) \end{bmatrix} = [Y_0(Y_0^T Y_0)^{-1/2} \ Y_p(Y_0^T Y_0)^{-1/2}] \text{expm} \left( t \begin{bmatrix} \Omega_0 & -\Sigma_0 \\ I_p & 0 \end{bmatrix} \right).$$

Undoing the transformation of  $\widetilde{Y}$  gives  $Y(t) = \widetilde{Y}(t)(Y_0^T Y_0)^{1/2} \text{expm}(t H_0)$ . By the construction above, matrix  $Y(t)$  represents the first  $p$  columns of a representative of a geodesic on  $(\text{GL}_n/\text{Stab}_E, \bar{g})$  and this is sufficient to define the whole geodesic.  $\square$

## 4.5 Metric space

By the Hopf–Rinow theorem,  $(\text{GL}_n/\text{Stab}_E, g)$  is a complete metric space since the geodesics can be extended indefinitely. This means that given two s.p.s.d. matrices in  $S_+(p, n)$ , we can always construct a geodesic of minimal length that connects these two matrices. The length of this geodesic defines the distance function on  $(\text{GL}_n/\text{Stab}_E, g)$  and, as we will see in the next section, on  $S_+(p, n)$  also.

A practical use of this distance function requires an efficient algorithm to construct the connecting geodesics, preferably in closed form. We have not found such a closed-form solution for two arbitrary matrices. As alternative, one can numerically solve a relatively simple boundary value problem since the ODE can be integrated analytically as in Prop. 4.8. This is however beyond the scope of the current paper and we will not explore this further.

## 5 Isometric embedding in the space of real matrices

In this section, we will give an interpretation of the homogeneous space geometry with the right-invariant metric. It has been shown in Helmke & Moore (1994) that  $S_+(p, n)$  is also an embedded submanifold of  $\mathbb{R}^{n \times n}$ . Since the structure of a submanifold is easier to understand than that of a quotient manifold, we will see how the two relate. Specifically, we will construct an isometry between the two manifolds.

Since we are dealing with three manifolds now, we will fix some notation in order to avoid confusion. As depicted in the table below, we use  $\bar{\cdot}$  for objects of  $GL_n$  and  $\tilde{\cdot}$  for  $S_+(p, n)$ , except for  $A \in GL_n$  which simply denotes a matrix. For an arbitrary function or map, its domain will determine the notation. (Observe that the metric and the connection satisfy this convention.)

Manifold	$GL_n/Stab_E$	$GL_n$	$S_+(p, n)$
Elements	$\pi(A)$	$A$	$\tilde{S}$
Vector fields	$\eta$	$\bar{\eta}$	$\tilde{\eta}$
Metric	$g$	$\bar{g}$	$\tilde{g}$
Connection	$\nabla$	$\bar{\nabla}$	$\tilde{\nabla}$

### 5.1 Related elements

In Prop. 3.3 we have constructed a diffeomorphism  $\Theta_E$  between  $GL_n/Stab_E$  and  $S_+(p, n)$ : there is a one-to-one correspondence between the equivalence class  $\pi(A) \in GL_n/Stab_E$  and the fixed rank matrix  $\Theta_E(\pi(A)) \in S_+(p, n)$ . The meaning of this map is not difficult to understand. Suppose we select  $A = [Y \ Z]$  as a representative of  $\pi(A)$ . Now  $\Theta_E(\pi(A)) = \theta_E(A)$  gives a fixed rank matrix on  $S_+(p, n)$  by selecting the first  $p$  columns of  $A$ , i.e.  $Y \in \mathbb{R}_*^{n \times p}$ , and forming the matrix  $\tilde{S} = YY^T$ , an s.p.s.d. matrix of rank  $p$ . The equivalence along the fibres is also clear: matrix  $A_* = [YQ \ Z_*]$ ,  $Q \in O_p$  belongs to the same fibre  $\pi^{-1}(\pi(A))$  as  $A$  and  $\theta_E(A_*)$  gives indeed the same s.p.s.d. matrix  $\tilde{S} = YY^T$ .

This equivalence along the fibre was a useful property to exploit when deriving the expressions of the geodesics in §4. In what follows we will continue using it.

### 5.2 Related tangent vectors

One can relate a vector field on  $GL_n/Stab_E$  to a vector field on  $S_+(p, n)$  by means of the differential of  $\Theta_E$ . Suppose we take  $A = [Y \ Z]$  as representative of  $\pi^{-1}(\pi(A))$  with corresponding s.p.s.d. matrix  $\tilde{S} = YY^T \in S_+(p, n)$ . Then the differential is

$$F_{\pi(A)} := D\Theta_E(\pi(A)) : T_{\pi(A)} GL_n/Stab_E \rightarrow T_{\tilde{S}} S_+(p, n),$$

where we introduced the notation  $F_{\pi(A)}$ . A tangent vector  $\eta_{\pi(A)}$  is so-called *F-related* to a tangent vector  $\tilde{\eta}_{\tilde{S}}$  if it satisfies

$$F_{\pi(A)}(\eta_{\pi(A)}) = \tilde{\eta}_{\tilde{S}}.$$

This relation is unique because  $\Theta_E$  is a diffeomorphism (Boothby, 1986, Th. IV.2.7.). In other words,  $F_{\pi(A)}$  is an isomorphism between vector spaces. By slight abuse of notation, this relation carries over directly to vector fields: we say that the vector fields  $\eta$  and  $\tilde{\eta}$  are *F-related* if  $F(\eta) = \tilde{\eta}$ .

Since  $GL_n/Stab_E$  contains abstract elements, we worked with the horizontal lift of a tangent vector. We can relate these horizontal lifts in a similar way to the tangent vectors of  $S_+(p, n)$ . If we take the derivative of the identity  $\theta_E = \Theta_E \circ \pi$  and use property (12) for horizontal lifts, we get

$$D\theta_E(A)[\bar{\eta}_A] = D\Theta_E(\pi(A))[D\pi(A)[\bar{\eta}_A]] = D\Theta_E(\pi(A))[\eta_{\pi(A)}], \quad \text{for all } \bar{\eta}_A \in \mathcal{H}_A. \quad (41)$$

We see that the differential of  $\Theta_E$  can be computed by taking a classical derivative of the matrix valued function  $\theta_E$  where the directions lie in the horizontal space. Let us examine the underlying isomorphisms. From Prop. 3.6 we know that

$$D\pi(A)|_{\mathcal{H}_A} : \mathcal{H}_A \rightarrow T_{\pi(A)}\mathrm{GL}_n/\mathrm{Stab}_E$$

is a bijection. So, by restricting the domain of  $D\theta_E$  to the horizontal space, map

$$\bar{F}_A := D\theta_E(A)|_{\mathcal{H}_A} : \mathcal{H}_A \rightarrow T_{\tilde{S}}\mathrm{S}_+(p, n). \quad (42)$$

is again an isomorphism. The corresponding diagram,

$$\begin{array}{ccc} \mathcal{H}_A & & \\ \downarrow D\pi(A) & \searrow \bar{F}_A & \\ T_{\pi(A)}\mathrm{GL}_n/\mathrm{Stab}_E & \xrightarrow{F_{\pi(A)}} & T_{\tilde{S}}\mathrm{S}_+(p, n) \end{array} \quad (43)$$

expresses the identity  $\bar{F}_A = F_{\pi(A)} \circ D\pi(A)|_{\mathcal{H}_A}$ .

Before we proceed, we recall the following characterisation of the tangent space of  $\mathrm{S}_+(p, n)$  from Vandereycken & Vandewalle (2010, Prop. 5.2).

**PROPOSITION 5.1** The tangent space at  $\tilde{S} = YY^T \in \mathrm{S}_+(p, n)$  is satisfies

$$T_{\tilde{S}}\mathrm{S}_+(p, n) = \left\{ YHY^T + Y_\perp KY^T + YK^TY_\perp^T : H \in \mathcal{S}_p^{\text{sym}}, K \in \mathbb{R}^{(n-p) \times p} \right\}.$$

Now, the following proposition shows how these relations can be computed explicitly in case of  $A = [Y \ Y_\perp]$ .

**PROPOSITION 5.2** Let  $\bar{\eta}$  be the horizontal lift of a vector field  $\eta$  on  $\mathrm{GL}_n/\mathrm{Stab}_E$ . Then  $\bar{\eta}$  is  $\bar{F}$ -related to a unique vector field  $\tilde{\eta}$  on  $\mathrm{S}_+(p, n)$ . For  $A = [Y \ Y_\perp]$ , this relation can be computed as

$$\bar{F}_A : [Y(Y^T Y)^{-1}H(Y^T Y) + Y_\perp K \ 0_{n \times (n-p)}] \mapsto Y\mathcal{L}(H)Y^T + Y_\perp KY^T + YK^TY_\perp^T, \quad (44)$$

$$\bar{F}_A^{-1} : YHY^T + Y_\perp KY^T + YK^TY_\perp^T \mapsto [Y(Y^T Y)^{-1}\mathcal{L}^{-1}(H)(Y^T Y) + Y_\perp K \ 0_{n \times (n-p)}], \quad (45)$$

with  $\mathcal{L}$  the bijection

$$\mathcal{L} : \mathcal{S}_p^{\text{sym}} \rightarrow \mathcal{S}_p^{\text{sym}}, X \mapsto (Y^T Y)^{-1}X(Y^T Y) + (Y^T Y)X(Y^T Y)^{-1}.$$

*Proof.* From diagram (43) we have  $\bar{F}_A = F_{\pi(A)} \circ D\pi(A)|_{\mathcal{H}_A}$ . Take  $A = [Y \ Z]$  and  $\tilde{S} = YY^T$ . Since  $\dim(\mathcal{H}_A) = \dim(T_{\tilde{S}}\mathrm{S}_+(p, n))$ , we have that  $\bar{F}_A$  is a bijection for every  $A$ . Now restrict  $A$  to be of the form  $A = [Y \ Y_\perp]$ . The horizontal lift of a tangent vector  $\eta_{\pi(A)}$  will be of the form  $\bar{\eta}_A = [Y(Y^T Y)^{-1}H(Y^T Y) + Y_\perp K \ 0]$ . Working out the derivative of  $\theta_E(A)$ , we get

$$\begin{aligned} \bar{F}_A(\bar{\eta}_A) &= D\theta_E(A)[\bar{\eta}_A] = \bar{\eta}_A EA^T + AE\bar{\eta}_A^T \\ &= Y(Y^T Y)^{-1}H(Y^T Y)Y^T + Y_\perp KY^T + Y(Y^T Y)H(Y^T Y)^{-1}Y^T + YK^TY_\perp^T \\ &= Y\mathcal{L}(H)Y^T + Y_\perp KY^T + YK^TY_\perp^T = \tilde{\eta}_{\tilde{S}}. \end{aligned}$$

Based on Prop. 5.1 it is clear that  $\bar{F}_A(\bar{\eta}_A)$  corresponds to only one tangent vector in  $T_{\tilde{S}}\mathrm{S}_+(p, n)$ . Likewise, the inverse  $\bar{F}_A^{-1}$  is also unique: given the tangent vector  $\tilde{\eta}_{\tilde{S}} = YHY^T + Y_\perp KY^T + YK^TY_\perp^T$  in  $T_{\tilde{S}}\mathrm{S}_+(p, n)$ , matrix  $\bar{\eta}_A = [Y(Y^T Y)^{-1}\mathcal{L}^{-1}(H)(Y^T Y) + Y_\perp K \ 0]$  represents the horizontal lift at  $A = [Y \ Y_\perp]$ . Since  $Y^T Y \succ 0$ , we know from Lemma 3.7 that  $\mathcal{L}$  is a bijection. Hence equation  $\mathcal{L}(X) = H$  has a unique and symmetric solution.  $\square$

It is illustrative to verify that the maps (44)–(45) are indeed invariant along the fibre  $\pi^{-1}(\pi(A))$ .

### 5.3 Related metrics

Next, we relate the metrics. Let  $\eta, \nu$  be two vector fields on  $\mathrm{GL}_n/\mathrm{Stab}_E$  with related vector fields  $\tilde{\eta}, \tilde{\nu}$  on  $S_+(p, n)$ . The aim is to find an inner product  $\tilde{g}$  on  $S_+(p, n)$  such that, by definition,

$$\tilde{g}(\tilde{\eta}, \tilde{\nu}) := g(\eta, \nu).$$

We call the inner products  $g$  and  $\tilde{g}$  then again  $F$ -related. Recall from Prop. 3.6 that the inner product  $g(\eta, \nu)$  is actually defined in terms of the horizontal lifts  $\bar{\eta}, \bar{\nu}$  since

$$g(\eta, \nu) := \bar{g}(\bar{\eta}, \bar{\nu}).$$

This makes the inner product  $\bar{g}$  in addition  $\bar{F}$ -related to  $\tilde{g}$ .

Since  $S_+(p, n)$  is embedded in the Euclidean space  $\mathbb{R}^{n \times n} \simeq \mathbb{R}^{n^2}$ , we will construct  $\tilde{g}$  as the weighted Euclidean metric

$$\tilde{g}_{\tilde{S}}(\eta_1, \eta_2) := \mathrm{vec}(\eta_1)^T \mathcal{W} \mathrm{vec}(\eta_2), \quad \text{for all } \eta_1, \eta_2 \in T_{\tilde{S}} S_+(p, n),$$

for some matrix  $\mathcal{W} \in \mathbb{R}^{n^2 \times n^2}$ . We have used the vectorisation operator  $\mathrm{vec} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n^2}$ , which makes a vector from a matrix by column-wise stacking. In order that  $\tilde{g}$  is an inner product,  $\mathcal{W}$  should be symmetric and positive definite for all vectors in  $T_{\tilde{S}} S_+(p, n)$ . However,  $\mathcal{W}$  may be indefinite on  $\mathbb{R}^{n^2 \times n^2}$ .

The following proposition gives such a matrix  $\mathcal{W}$ . It makes use of the Kronecker product  $\otimes$  and the following property of the vec operator (Horn & Johnson, 1991, Ch. 4)

$$\mathrm{vec}(AXB) = (B^T \otimes A) \mathrm{vec}(X), \quad A \in \mathbb{R}^{m \times n}, X \in \mathbb{R}^{n \times p}, B \in \mathbb{R}^{p \times q}. \quad (46)$$

**PROPOSITION 5.3** Let  $\tilde{S} = YY^T \in S_+(p, n)$  and define  $M := Y^T Y$ . Then the relation

$$\tilde{g}_{\tilde{S}}(\tilde{\eta}_1, \tilde{\eta}_2) := \frac{1}{2} \mathrm{vec}(\tilde{\eta}_1)^T \mathcal{W} \mathrm{vec}(\tilde{\eta}_2), \quad \forall \tilde{\eta}_1, \tilde{\eta}_2 \in T_{\tilde{S}} S_+(p, n), \quad (47)$$

with

$$\mathcal{W} = (Y \otimes Y)(M^3 \otimes M + M \otimes M^3)^{-1}(Y^T \otimes Y^T) + YM^{-3}Y^T \otimes Y_\perp Y_\perp^T + Y_\perp Y_\perp^T \otimes YM^{-3}Y^T.$$

defines a Riemannian metric  $\tilde{g}$  on  $S_+(p, n)$ . Furthermore,  $\tilde{g}$  is  $F$ -related to the metric  $g$  on  $\mathrm{GL}_n/\mathrm{Stab}_E$  of Prop. 3.6.

*Proof.* We first check that (47) is well defined and that it fulfills the three axioms of an inner product. Linearity is trivial. Next, observe that we can write  $\mathcal{W} = \mathcal{Y} \mathcal{C} \mathcal{Y}^T$  with

$$\mathcal{Y} = [Y \otimes Y \quad Y \otimes Y_\perp \quad Y_\perp \otimes Y \quad Y_\perp \otimes Y_\perp]$$

and

$$\mathcal{C} = \begin{bmatrix} (M^3 \otimes M + M \otimes M^3)^{-1} & 0 & 0 & 0 \\ 0 & M^{-3} \otimes I_{n-p} & 0 & 0 \\ 0 & 0 & I_{n-p} \otimes M^{-3} & 0 \\ 0 & 0 & 0 & 0_{(n-p)^2 \times (n-p)^2} \end{bmatrix}.$$

Since  $M \succ 0$ , one can verify that  $\mathcal{C} \succeq 0$ , from which, by Sylvester's law of inertia,  $\mathcal{W} \succeq 0$ . This immediately gives symmetry. Positive-definiteness follows from the fact that  $\tilde{g}_{\tilde{S}}(X, X)$  vanishes only for matrices  $X \in \mathbb{R}^{n \times n}$  of the form  $Y_\perp LY_\perp^T$  for some  $L \in \mathbb{R}^{(n-p) \times (n-p)}$ . Since these matrices do not belong to the tangent space,  $\tilde{g}_{\tilde{S}}(\tilde{\eta}, \tilde{\eta}) = 0$  implies  $\tilde{\eta} = 0$ . Furthermore, the metric is independent of the choice of  $Y$  and  $Y_\perp$ . Suppose we take  $Z = YQ$ , for some  $Q \in O_p$  and  $Z_\perp = Y_\perp P$  for some  $P \in O_{n-p}$ . Then it is straightforward to show that matrix

$$(Z \otimes Z)(W^3 \otimes W + W \otimes W^3)^{-1}(Z^T \otimes Z^T) + ZW^{-3}Z^T \otimes Z_\perp Z_\perp^T + Z_\perp Z_\perp^T \otimes ZW^{-3}Z^T$$

with  $W = Z^T Z = Q^T M Q$  is equal to  $\mathcal{W}$ . Hence, the inner product stays the same. Second, we will prove by construction that (47) coincides with the right-invariant metric (13). Let us take the usual related base points  $A = [Y \ Y_\perp]$  and  $\tilde{S} = YY^T$  and related tangent vectors  $\tilde{\eta}_1 = \overline{F}(\bar{\eta}_1)$  and  $\tilde{\eta}_2 = \overline{F}(\bar{\eta}_2)$ . All these tangent vectors can be parameterized by matrices  $H_1, H_2 \in \mathcal{S}_p^{\text{sym}}$  and  $K_1, K_2 \in \mathbb{R}^{(n-p) \times p}$  satisfying the relations in Prop. 5.2. In case of  $\tilde{\eta}_1$  and  $\bar{\eta}_1$ . This gives

$$\tilde{\eta}_1 = YH_1Y^T + Y_\perp K_1Y^T + YK_1^TY_\perp, \quad \bar{\eta}_1 = [YM^{-1}\mathcal{L}^{-1}(H_1)M + Y_\perp K_1 \ 0]$$

and similarly for  $\tilde{\eta}_2$  and  $\bar{\eta}_2$ . First, we work out  $\tilde{g}_{\tilde{S}}(\tilde{\eta}_1, \tilde{\eta}_2)$ . Using property (46), we get that

$$\begin{aligned} \text{vec}(\tilde{\eta}_1) &= (Y \otimes Y) \text{vec}(H_1) + (Y \otimes Y_\perp) \text{vec}(K_1) + (Y_\perp \otimes Y) \text{vec}(K_1^T) \\ &= \mathcal{Y} \begin{bmatrix} \text{vec}(H_1) \\ \text{vec}(K_1) \\ \text{vec}(K_1^T) \\ 0_{(n-p)^2 \times 1} \end{bmatrix}. \end{aligned}$$

Now we have that

$$\begin{aligned} 2\tilde{g}_{\tilde{S}}(\tilde{\eta}_1, \tilde{\eta}_2) &= \begin{bmatrix} \text{vec}(H_2) \\ \text{vec}(K_2) \\ \text{vec}(K_2^T) \\ 0_{(n-p)^2 \times 1} \end{bmatrix}^T \mathcal{Y}^T \mathcal{Y} \mathcal{C} \mathcal{Y}^T \mathcal{Y} \begin{bmatrix} \text{vec}(H_1) \\ \text{vec}(K_1) \\ \text{vec}(K_1^T) \\ 0_{(n-p)^2 \times 1} \end{bmatrix} \\ &= \text{vec}(H_2)^T (M^{-1} \otimes M + M \otimes M^{-1})^{-1} \text{vec}(H_1) \\ &\quad + \text{vec}(K_2)^T (M^{-1} \otimes I) \text{vec}(K_1) + \text{vec}(K_2^T)^T (I \otimes M^{-1}) \text{vec}(K_1^T) \\ &= \text{vec}(H_2)^T \text{vec}(\mathcal{L}^{-1}H_1) + \text{vec}(K_2)^T \text{vec}(K_1 M^{-1}) + \text{vec}(K_2^T)^T \text{vec}(M^{-1} K_1^T). \end{aligned}$$

Changing to matrices and using the definition of  $\mathcal{L}$  from Prop. 5.2 gives

$$\begin{aligned} 2\tilde{g}_{\tilde{S}}(\tilde{\eta}_1, \tilde{\eta}_2) &= \text{tr}(\mathcal{L}(\mathcal{L}^{-1}(H_2)) \mathcal{L}^{-1}(H_1)) + \text{tr}(M^{-1} K_2^T K_1) + \text{tr}(K_2 M^{-1} K_1^T) \\ &= 2\text{tr}(M \mathcal{L}^{-1}(H_2) M^{-1} \mathcal{L}^{-1}(H_1)) + 2\text{tr}(M^{-1} K_2^T K_1). \end{aligned}$$

Now applying formula (17) for  $\bar{g}_A(\bar{\eta}_1, \bar{\eta}_2)$ , we get immediately that

$$\bar{g}_A(\bar{\eta}_1, \bar{\eta}_2) = \text{tr}(M \mathcal{L}^{-1}(H_2) M^{-1} \mathcal{L}^{-1}(H_1) + M^{-1} K_2^T K_1) = \tilde{g}_{\tilde{S}}(\tilde{\eta}_1, \tilde{\eta}_2).$$

Since the metrics are independent of the choice of  $Y$  and  $A$ , this concludes the proof.  $\square$

The following is immediate.

#### COROLLARY 5.4 Map

$$\Theta_E : (\text{GL}_n / \text{Stab}_E, g) \rightarrow (\text{S}_+(p, n), \tilde{g})$$

is an isometry with metric  $g$  of Prop. 3.6 and metric  $\tilde{g}$  of Prop. 5.3.

Observe that any change to  $\mathcal{W}$  that is restricted to the normal space of  $\tilde{S}$  will have no influence on (47). Specifically, we can choose any  $\mathcal{W} = \mathcal{Y} \mathcal{C} \mathcal{Y}^T$  with

$$\mathcal{Y} = [Y \otimes Y \quad Y \otimes Y_\perp \quad Y_\perp \otimes Y \quad Y_\perp \otimes Y_\perp]$$

and

$$\mathcal{C} = \begin{bmatrix} (M^3 \otimes M + M \otimes M^3)^{-1} & 0 & 0 & X_1 \\ 0 & M^{-3} \otimes I_{n-p} & 0 & X_2 \\ 0 & 0 & I_{n-p} \otimes M^{-3} & X_3 \\ X_1^T & X_2^T & X_3^T & X_4 \end{bmatrix}, \quad \begin{aligned} X_1 &\in \mathbb{R}^{p^2 \times p^2} \\ X_2 &\in \mathbb{R}^{p(n-p) \times p(n-p)} \\ X_3 &\in \mathbb{R}^{p(n-p) \times p(n-p)} \\ X_4 &\in \mathcal{S}_{(n-p)^2}^{\text{sym}} \end{aligned}.$$

## 5.4 Related connection

By virtue of the naturality of the Levi-Civita connection (Lee, 1997, Prop. 5.6), isometry  $\Theta_E$  of Prop. 3.2 induces a  $D\Theta_E$ -related connection  $\tilde{\nabla}$  on  $(S_+(p, n), \tilde{g})$ . In other words,  $\tilde{\nabla} := D\Theta_E \circ \nabla$  with  $\nabla$  the connection on  $(GL_n / Stab_E, g)$ .

We would like to express this again in terms of horizontal lifts. Recall from Prop. 5.2 that the bijection  $\bar{F}_A^{-1}$  allows us to lift each vector field on  $S_+(p, n)$  to a unique vector field on  $GL_n$  with  $A = [Y \ Y_\perp]$  the  $\Theta_E$ -related element of  $\tilde{S} = YY^T$ . Furthermore, this lifted vector field can be parameterized by a classic matrix function:

$$\bar{\eta}(A) := \bar{F}_A^{-1}(\tilde{\eta}(\tilde{S})) = [Z_\eta(Y) \ 0_{n \times (n-p)}] \quad \text{with } Z_\eta : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}^{n \times p}. \quad (48)$$

Now, Prop. 3.11 allows us to express the connection  $\tilde{\nabla}$  of  $(S_+(p, n), \tilde{g})$  in terms of the horizontal lifts as

$$\tilde{\nabla}_{\tilde{\nu}} \tilde{\eta} = \bar{F}(P^h(\bar{\nabla}_{\bar{\nu}} \bar{\eta})) = \bar{F}(\bar{\nabla}_{\bar{\nu}} \bar{\eta}),$$

where we used the property that  $P^h$  belongs to the null space of  $\bar{F}$ . Summarizing, this results in the following proposition.

**PROPOSITION 5.5** Let  $\tilde{\eta}, \tilde{\nu}$  be two vector fields on  $S_+(p, n)$ . Let  $Z_\eta, Z_\nu$  be the parametrization (48) of the related vector fields  $\bar{\eta}, \bar{\nu}$  on  $GL_n / Stab_E$ . Then the Levi-Civita connection in  $\tilde{S} = YY^T$  w.r.t. metric  $\tilde{g}$  of Prop. 5.3 is given by

$$\tilde{\nabla}_{\tilde{\nu}} \tilde{\eta}(\tilde{S}) = (DZ_\eta(Y)[Z_\nu] + W) Y^T + Y (DZ_\eta(Y)[Z_\nu] + W)^T \quad (49)$$

where

$$W := \frac{1}{2} \{ Y(Y^T Y)^{-1} (Z_\eta^T Z_\nu + Z_\nu^T Z_\eta) - Z_\nu(Y^T Y)^{-1} (Z_\eta^T Y + Y^T Z_\eta) - Z_\eta(Y^T Y)^{-1} (Z_\nu^T Y + Y^T Z_\nu) \}.$$

## 5.5 Related geodesics

The geodesics on  $(S_+(p, n), \tilde{g})$  are simply the  $\Theta_E$ -related geodesics of  $(GL_n / Stab_E, g)$ .

**PROPOSITION 5.6** Let  $\tilde{S}_0 = Y_0 Y_0^T \in S_+(p, n)$  with  $Y_0 \in \mathbb{R}_*^{n \times p}$  be given. Define  $N_0 := (Y_0^T Y_0)^{1/2}$  and  $Z(t) := Y_0 N_0^{-1} X_{11}(t) + Y_p N_0^{-1} X_{21}(t)$  with

$$\begin{bmatrix} X_{11}(t) & X_{12}(t) \\ X_{21}(t) & X_{22}(t) \end{bmatrix} := \expm \left( t \begin{bmatrix} N_0^{-1} H_0 N_0 - N_0 H_0 N_0^{-1} & -N_0^{-1} Y_p^T Y_p N_0^{-1} \\ I_p & 0_{p \times p} \end{bmatrix} \right).$$

Then the geodesic in  $\tilde{S}_0$  along

$$Y_0(N_0^{-1} H_0 N_0 + N_0 H_0 N_0^{-1}) Y_0^T + Y_p Y_0^T + Y_0 Y_p^T,$$

with given  $H_0 \in \mathcal{S}_p^{\text{sym}}$  and  $Y_0 \perp Y_p \in \mathbb{R}^{n \times p}$  is the curve

$$t \mapsto \tilde{S}(t) := Z(t) N_0 \expm(2tH_0) N_0 Z(t)^T.$$

*Proof.* All geodesics on  $(S_+(p, n), \tilde{g})$  are of the form  $\tilde{S}(t) = Y(t) Y(t)^T$  with  $Y(t)$  from Prop. 4.8. The initial tangent vector is identified by taking the derivative in  $t = 0$ .  $\square$

## 6 Special geodesics and quasi-geodesics

Observe that every  $Y \in \mathbb{R}_*^{n \times p}$  can be written as  $Y = UC$  with orthonormal  $U \in \mathbb{R}_*^{n \times p}$  and  $C \in GL_p$ . The corresponding s.p.s.d. matrix can be decomposed as  $\tilde{S} = UWU^T$  with  $W = CC^T \succ 0$ . In this section we will see whether we can derive a geodesic as a decomposed curve also: one curve for the orthonormal part  $U(t)$  and one for the small matrix  $W(t) \succ 0$ . The general answer will turn out to be negative although the resulting curve will be a first-order approximation of a geodesic, called a *quasi-geodesic*.

## 6.1 The case $K_0 = 0$ .

Take  $K_0 = 0$  in (38), or equivalently  $Y_p = 0$  in Prop. 4.8, then the geodesic, which we denote by  $Y_K(t)$ , is given by

$$Y_K(t) = Y_0(Y_0^T Y_0)^{-1/2} \expm(t\Omega_0)(Y_0^T Y_0)^{1/2} \expm(tH_0). \quad (50)$$

Observe that this curve does not change the column span of  $Y_0$ . This geodesic is in fact directly related to the geodesics on  $(S_+(p, p), \tilde{g}) \simeq (\mathrm{GL}_p/\mathrm{O}_p, g)$ . Since we can always write  $Y_0 = U_0 C_0$  with orthonormal  $U_0 \in \mathbb{R}_{*}^{n \times p}$  and  $C_0 \in \mathrm{GL}_p$ , we have  $Y_K(t) = U_0 C(t)$  with

$$C(t) = C_0(C_0^T C_0)^{-1/2} \expm(t\Omega_0)(C_0^T C_0)^{1/2} \expm(tH_0).$$

and

$$\Omega_0 = (C_0^T C_0)^{-1/2} H_0 (C_0^T C_0)^{1/2} - (C_0^T C_0)^{1/2} H_0 (C_0^T C_0)^{-1/2}.$$

Suppose we take  $C_0$  as a representative in  $\mathrm{GL}_p/\mathrm{O}_p$  for matrix  $C_0 C_0^T \in S_+(p, p)$ . Looking at Prop. 4.8, we see that  $\pi(C(t))$  is indeed a geodesic on  $(\mathrm{GL}_p/\mathrm{O}_p, g)$ .

## 6.2 The case $H_0 = 0$ .

Let  $L_0 := K_0(Y_0^T Y_0)^{1/2}$ . If  $H_0 = 0$  and  $p < n$ , we can write the matrix exponential in expression (38) as follows

$$\expm\left(t \begin{bmatrix} 0 & -L_0^T \\ L_0 & 0 \end{bmatrix}\right) = \begin{bmatrix} V & 0 & 0 \\ 0 & U_1 & U_2 \end{bmatrix} \begin{bmatrix} \cos(t\Sigma) & -\sin(t\Sigma) & 0 \\ \sin(t\Sigma) & \cos(t\Sigma) & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} V^T & 0 \\ 0 & U_1^T \\ 0 & U_2^T \end{bmatrix}$$

with

$$L_0 = [U_1 \ U_2] \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T$$

a partitioned singular value decomposition. The geodesic, now denoted  $Y_H(t)$ , satisfies

$$Y_H(t) = (Y_0(Y_0^T Y_0)^{-1/2} V \cos(t\Sigma) + Y_0^\perp U_1 \sin(t\Sigma)) V^T (Y_0^T Y_0)^{1/2}. \quad (51)$$

Observe that  $\text{span}(Y_H(t))$  is no longer constant along this geodesic.

In this case,  $Y_H(t)$  is related to a geodesic on the Grassmann manifold of linear subspaces equipped with the natural metric. To see this, observe that  $\text{span}(Y_H(t))$  coincides with the expression in Th. 3.6 of Absil *et al.* (2004) for a geodesic in the Grassmann manifold. Contrary to the Grassmann manifold where  $Y_H(t)$  belongs to the equivalence class  $Y_{\mathrm{GL}_p}$ , the equivalence of  $Y_H(t)$  in  $\mathrm{GL}_n/\mathrm{Stab}_E$  is restricted to the orthogonal group only, i.e.,  $Y_{\mathrm{O}_p}$ ; see also §3.3.

## 6.3 A quasi-geodesic

We have seen that we can classify the geodesics into two disjoint types, depending on whether they change  $\text{span}(Y(t))$  or not. These two types correspond to either  $H_0$  or  $K_0$  zero. In case  $H_0$  and  $K_0$  are both non-zero, we can construct a curve by composition of these two geodesics. Take again  $Y_0 = U_0 C_0$ . In order that the composition of the geodesics match, we rewrite the geodesic by isolating the constant orthonormal matrix and obtain

$$Y_K(t) = \underbrace{U_0 C_0 (C_0^T C_0)^{-1/2}}_{=: \tilde{U}_0, \ \tilde{U}_0^T \tilde{U}_0 = I_p} \gamma_K(t) \quad \text{with } \gamma_K(t) := \expm(t\Omega_0)(C_0^T C_0)^{1/2} \expm(tH_0).$$

In the same way, we obtain

$$Y_H(t) = \gamma_H(t)(C_0^T C_0)^{1/2} \quad \text{with } \gamma_H(t) := (U_0 C_0 (C_0^T C_0)^{-1/2} V \cos(t\Sigma) + Y_0^\perp U_1 \sin(t\Sigma)) V^T$$

Now we have  $Y_K(t) = \gamma_H(0)\gamma_K(t)$  and  $Y_H(t) = \gamma_H(t)\gamma_K(0)$ . The new curve for  $H_0 \neq 0$  and  $K_0 \neq 0$  is then

$$\gamma_{HK}(t) := \gamma_H(t)\gamma_K(t),$$

which is complete, i.e., it stays in  $\mathrm{GL}_n/\mathrm{Stab}_E$  for all  $t$ .

This curve is obviously not a geodesic, but it is a first-order approximation of it in the following sense: take  $Y(t)Y(t)^T$  the geodesic on  $(\mathrm{S}_+(p, n), \tilde{g})$  with  $Y(0) = \gamma_{HK}(0)$ , then

$$\mathrm{dist}_{\tilde{g}}(\gamma_{HK}(t)\gamma_{HK}(t)^T, Y(t)Y(t)^T) = O(t^2), \quad t \rightarrow 0.$$

Here  $\mathrm{dist}_{\tilde{g}}$  is the distance on the metric space  $(\mathrm{S}_+(p, n), \tilde{g})$  defined by the length of the minimal geodesic. In §5.3 metric  $\tilde{g}$  was shown to be a weighted Euclidean metric, so we can bound this distance by the usual Frobenius norm of the embedding space. This gives the following equivalent property of quasi-geodesics

$$\|\gamma_{HK}(t)\gamma_{HK}(t)^T - Y(t)Y(t)^T\|_F = O(t^2), \quad t \rightarrow 0. \quad (52)$$

Curves of this type are called quasi-geodesics of order one (Nishimori & Akaho, 2005) or first-order retractions (Absil *et al.*, 2008, Def. 4.1.1). Amongst other uses, they appear in optimization on manifolds as a cheap but equally good substitute of the exponential map and geodesics; see Absil *et al.* (2008) for a general overview.

To verify that  $\gamma_{HK}(t)\gamma_{HK}(t)^T$  is indeed of first-order, apply the Baker–Campbell–Hausdorff formula (Hairer *et al.*, 2006, III.4.2) to split an exponential:

$$\mathrm{expm}(tA)\mathrm{expm}(tB) = \mathrm{expm}(t(A+B)) + O(t^2), \quad t \rightarrow 0.$$

Using this expansion for the exponential of (38), we obtain with  $L_0 := K_0(Y_0^T Y_0)^{1/2}$

$$\mathrm{expm}\left(t \begin{bmatrix} \Omega_0 & -L_0^T \\ L_0 & 0 \end{bmatrix}\right) = \mathrm{expm}\left(t \begin{bmatrix} 0 & -L_0^T \\ L_0 & 0 \end{bmatrix}\right) \mathrm{expm}\left(t \begin{bmatrix} \Omega_0 & 0 \\ 0 & 0 \end{bmatrix}\right) + O(t^2).$$

Now working out each exponential like in the cases above, we arrive at  $Y(t) = \gamma_{HK}(t) + O(t^2), t \rightarrow 0$ .

Curve  $\gamma_{HK}(t)$  of (52) shows some resemblance to the curves in Bonnabel & Sepulchre (2009). In §7.4, we will come back to these curves and compare them more thoroughly.

## 7 Comparison with other metrics

In §3.4 we have already established that for  $p < n$  there is no longer a natural choice for the metric as in the case  $p = n$ . In this paper, we have chosen the right-invariant metric since it turned the canonical projection  $\pi : \mathrm{GL}_n \rightarrow \mathrm{GL}_n/\mathrm{Stab}_E$  into a Riemannian submersion, a useful property to exploit for deriving geodesics. In the literature there exist other metrics however. Each choice has advantages and disadvantages. Below we will briefly compare them with our choice and focus most attention on the properties of the geodesics.

Note that, in order not to make the notation of this section too dense, we depart from the notational conventions of §5 and simply use plain matrices and vector fields that belong to  $\mathrm{S}_+(p, n)$ .

### 7.1 Embedded submanifold with the Euclidean metric.

One can regard  $\mathrm{S}_+(p, n)$  as a submanifold embedded in  $\mathbb{R}^{n \times n}$  and equip it with the usual Euclidean metric, as was done in, e.g., Helmke & Moore (1994), Helmke & Shayman (1995), Orsi *et al.* (2004), Orsi *et al.* (2006), Vandereycken *et al.* (2009), Vandereycken & Vandewalle (2010). The advantage of this familiar metric is that it allows us to interpret many geometric objects in a straightforward way. Take e.g. two vector fields  $\nu, \eta$  on  $\mathrm{S}_+(p, n)$ , then the Levi-Civita connection in  $S \in \mathrm{S}_+(p, n)$  satisfies

$$\nabla_\nu \eta(S) = P_S(D\eta(S)[\nu]),$$

with  $P_S$  the usual orthogonal projection onto the tangent space  $T_SS_+(p,n)$  and  $D\eta$  a classic Euclidean directional derivative of a matrix-valued function  $\eta$ .

Hence, it may seem appropriate to use this metric, but it is not the most appealing from a theoretical perspective since  $S_+(p,n)$  is not a complete metric space. In Vandereycken *et al.* (2009), the authors derive the equations of motion of the geodesics and clearly show that they can not always be extended indefinitely. Still, the simplicity of many expressions makes this an interesting geometry.

We have seen in §5 that our homogeneous space geometry also coincides with an embedded submanifold but now with a much more involved and nonconstant metric. It is however still a weighted Euclidean metric. In principle this allows for the same interpretation of the Levi-Civita connection involving a projection but now w.r.t. to the metric of Prop. 5.3.

## 7.2 Quotient manifold $\mathbb{R}_*^{n \times p}/O_p$ with the Euclidean metric

One can factor every  $S \in S_+(p,n)$  as  $S = YY^T$  with orthonormal  $Y \in \mathbb{R}_*^{n \times p}$ . This factorization is unique up to the action of the orthogonal group, i.e., transformation  $Y \mapsto YQ$  gives the same matrix  $S$  for all  $Q \in O_p$ . This allows one to describe  $S_+(p,n)$  as the quotient manifold

$$S_+(p,n) \simeq \mathbb{R}_*^{n \times p}/O_p.$$

In Journée *et al.* (2010), Meyer *et al.* (2010), Bonnabel *et al.* (2010), Sepulchre *et al.* (2010), the authors equip this quotient manifold with the Euclidean metric. The horizontal space at  $Y \in \mathbb{R}_*^{n \times p}$  is given in (15) of Journée *et al.* (2010):

$$\mathcal{H}_Y^{\text{Eucl}} = \{Z \in \mathbb{R}_*^{n \times p} : Z^T Y = Y^T Z\},$$

which, by counting dimensions, is equivalent to

$$\mathcal{H}_Y^{\text{Eucl}} = [Y(Y^T Y)^{-1} \mathcal{S}_p^{\text{sym}} + Y_\perp \mathbb{R}^{(n-p) \times p}].$$

It is not difficult to see that this description is equivalent to equipping  $\text{GL}_n/\text{Stab}_E$  with a metric  $g_A^{\text{Eucl}}$  which is derived from the Euclidean metric on  $\text{GL}_n$ ,

$$\bar{g}^{\text{Eucl}}(\bar{\eta}, \bar{\nu}) := \text{tr}(\bar{\eta}^T \bar{\nu}).$$

Indeed, reiterating the steps of the derivation in §3, but now with  $\bar{g}^{\text{Eucl}}$ , we get

$$\mathcal{H}_A^{\text{Eucl}} = A^{-T} \begin{bmatrix} \mathcal{S}_p^{\text{sym}} & 0 \\ \mathbb{R}^{(n-p) \times p} & 0 \end{bmatrix}$$

as horizontal space at  $A \in \text{GL}_n$ . In  $A = [Y \ Y_\perp]$ , the horizontal lifts are of the form

$$\mathcal{H}_A^{\text{Eucl}} = [Y(Y^T Y)^{-1} \mathcal{S}_p^{\text{sym}} + Y_\perp \mathbb{R}^{(n-p) \times p} \ 0_{n \times (n-p)}].$$

Notice that the first  $p$  columns of  $\mathcal{H}_A^{\text{Eucl}}$  are exactly  $\mathcal{H}_Y^{\text{Eucl}}$  of above. Metric  $g_A^{\text{Eucl}}$  on  $\text{GL}_n/\text{Stab}_E$  is now the restriction of  $\bar{g}^{\text{Eucl}}$  to  $\mathcal{H}_A^{\text{Eucl}}$ , which is the same as in Journée *et al.* (2010).

Since  $\pi^{\text{Eucl}} : (\text{GL}_n, \bar{g}_{\text{Eucl}}) \rightarrow (\text{GL}_n/\text{Stab}_E, g_A^{\text{Eucl}})$  is again a Riemannian submersion, we have that the geodesics are the projection of horizontal geodesics on  $(\text{GL}_n, \bar{g}^{\text{Eucl}})$ . These will be straight curves

$$t \mapsto Y_0 + t \dot{Y}_0, \quad \text{for all } Y_0 \in \mathbb{R}_*^{n \times p}, \dot{Y}_0 \in \mathcal{H}_{Y_0}^{\text{Eucl}},$$

which are obviously not complete; the underlying reason being that  $(\text{GL}_n, \bar{g}_{\text{Eucl}})$  is not complete. This is the primary reason why we disregarded the Euclidean metric in §3.5.

### 7.3 Quotient manifold $\mathbb{R}_*^{n \times p}/O_p$ with a special metric

In Absil *et al.* (2009), the authors equip  $\mathbb{R}_*^{n \times p}/O_p$  with the following specially chosen metric on  $\mathbb{R}_*^{n \times p}$ :

$$g_Y^{\text{Left}}(Z_1, Z_2) = \text{tr}(Z_1^T P_Y^\perp(Z_2) + Z_1^T Y(Y^T Y)^{-2} Y^T Z_2)$$

where  $P_Y^\perp := I - P_Y$  denotes the usual orthogonal projection. In this case, the horizontal space at  $Y \in \mathbb{R}_*^{n \times p}$  satisfies

$$\mathcal{H}_Y^{\text{Left}} = [Y \mathcal{S}_p^{\text{sym}} + Y_\perp \mathbb{R}^{(n-p) \times p}].$$

This simple expression allows to pick a specific affine (not the Levi-Civita) connection which results in a particularly lean expression for a Newton equation on  $\mathbb{R}_*^{n \times p}/O_p$ . We refer to Absil *et al.* (2009) for details. Here, we only want to point out that this metric coincides with equipping  $\text{GL}_n/\text{Stab}_E$  with the left-invariant metric

$$\bar{g}_A^{\text{Left}}(\eta_A, \eta_A) = \bar{g}_I^{\text{Left}}(A^{-1}\eta_A, A^{-1}\eta_A) = \text{tr}(\eta_A^T A^{-T} A^{-1} \eta_A). \quad (53)$$

Analogously as in the previous section, quotient manifold  $(\text{GL}_n/\text{Stab}_E, g_{\text{Left}})$  has

$$\mathcal{H}_A^{\text{Left}} = A \begin{bmatrix} \mathcal{S}_p^{\text{sym}} & 0 \\ \mathbb{R}^{(n-p) \times p} & 0 \end{bmatrix}$$

as horizontal space at  $A \in \text{GL}_n$ . In  $A = [Y \ Y_\perp]$ , this gives

$$\mathcal{H}_A^{\text{Left}} = [Y \mathcal{S}_p^{\text{sym}} + Y_\perp \mathbb{R}^{(n-p) \times p} \ 0_{n \times (n-p)}] = [\mathcal{H}_Y^{\text{Left}} \ 0_{n \times (n-p)}].$$

For the previous choice of  $A$ , let  $\eta = [V_\eta \ 0] \in \mathcal{H}_A^{\text{Left}}$  with  $V_\eta = Y S_\eta + Y_\perp K_\eta \in \mathcal{H}_Y^{\text{Left}}$  and similarly for  $\eta$  and  $V_\eta$ . Then metric  $g_{\text{Left}}$  satisfies

$$g_A^{\text{Left}}(\eta_A, \eta_A) = \text{tr} \left( \begin{bmatrix} S_\eta & K_\eta^T \\ 0 & 0 \end{bmatrix} A^T A^{-T} A^{-1} A \begin{bmatrix} S_\eta & 0 \\ K_\eta & 0 \end{bmatrix} \right) = \text{tr}(S_\eta S_\eta + K_\eta^T K_\eta),$$

which is obviously the same as

$$\begin{aligned} g_Y^{\text{Left}}(V_\eta, V_\eta) &= \text{tr}((S_\eta Y^T + K_\eta^T Y_\perp^T) Y (Y^T Y)^{-2} Y^T (Y S_\eta + Y_\perp K_\eta)) \\ &\quad + \text{tr}((S_\eta Y^T + K_\eta^T Y_\perp^T) Y_\perp Y_\perp^T (Y S_\eta + Y_\perp K_\eta)). \end{aligned}$$

The downside of this description is that this horizontal space does not define a connection on the principal bundle  $\text{GL}_n(\text{GL}_n/\text{Stab}_E, \text{Stab}_E)$  since  $\mathcal{H}_{AH}^{\text{Left}} \neq \mathcal{H}_A^{\text{Left}} H$  for all  $H \in \text{Stab}_E$ . Consequently, it is not immediately obvious whether

$$\pi^{\text{Left}} : (\text{GL}_n, \bar{g}^{\text{Left}}) \rightarrow (\text{GL}_n/\text{Stab}_E, g^{\text{Left}})$$

is a Riemannian submersion or not. We do have however  $\mathcal{H}_{YQ}^{\text{Left}} = \mathcal{H}_Y^{\text{Left}} Q$  for all  $Q \in O_p$ . So projection

$$\pi^{\text{Left}} : (\mathbb{R}_*^{n \times p}, \bar{g}^{\text{Left}}) \rightarrow (\text{GL}_n/\text{Stab}_E, g^{\text{Left}})$$

is Riemannian. However, the geodesics of  $(\mathbb{R}_*^{n \times p}, g^{\text{Left}})$  are most likely rather difficult to find, and it is not known whether they are complete or not.

### 7.4 Quotient manifold $(\text{St}(p, n) \times \text{S}_+(p, p))/O_p$ with a polar metric

We can factor every  $S \in \text{S}_+(p, n)$  as  $S = UPU^T$  with orthonormal  $U \in \mathbb{R}_*^{n \times p}$  and  $P \in \text{S}_+(p, p)$ . This factorization is unique up to the action of the orthogonal group, i.e., transformation

$$U \mapsto UQ, \quad P \mapsto Q^T PQ, \quad \text{for all } Q \in O_p$$

gives the same matrix  $S$ . This allows one to describe  $S_+(p, n)$  as the quotient manifold

$$S_+(p, n) \simeq (St(p, n) \times S_+(p, p)) / O_p,$$

with  $St(p, n)$  the Stiefel manifold of  $n \times p$  orthonormal matrices. In Bonnabel & Sepulchre (2009), the authors equip this manifold with a so-called *polar metric* which is a linear combination of the natural metrics on  $St(p, n)$  and  $S_+(p, p)$ . Since  $St(p, n)$  and  $S_+(p, p)$  both have a rich geometry, the metric on  $(St(p, n) \times S_+(p, p)) / O_p$  inherits most of the useful invariance properties (but not all) of the reductive space  $S_+(n, n)$ .

A difficulty regarding this approach is that the quotient map

$$\pi^{\text{Polar}} : (St(p, n) \times S_+(p, p)) \rightarrow (St(p, n) \times S_+(p, p)) / O_p$$

is not a Riemannian submersion. Although the horizontal geodesics on  $(St(p, n) \times S_+(p, p))$  are available in closed form as (Bonnabel & Sepulchre, 2009, Th. 5.2),

$$t \mapsto U(t)R^2(t)U(t)^2, \quad \text{with } U(t) \in St(p, n) \text{ and } R^2(t) \in S_+(p, p), \quad (54)$$

their projection under  $\pi^{\text{Polar}}$  will not be geodesics on  $(St(p, n) \times S_+(p, p)) / O_p$ ; they will only be quasi-geodesics. In addition, it is not obvious if this description will eventually lead to complete or efficient geodesics. These curves do however have nice properties: they are complete, available in closed-form and it is straightforward to construct a curve connecting to s.p.s.d. matrices, see, e.g., Meyer *et al.* (2009), Meyer *et al.* (2010), Bonnabel *et al.* (2010), Bonnabel & Sepulchre (2010), Sepulchre *et al.* (2010).

These quasi-geodesics of Bonnabel & Sepulchre (2009) show some resemblance to our quasi-geodesics: compare the curve (54) of Bonnabel & Sepulchre (2009, Th. 5.2) to  $\gamma_{HK}(t)\gamma_{HK}(t)^T$  of §6.3. While the curves for the orthonormal part, i.e.,  $U(t)$  and  $\gamma_H(t)$ , are both based on geodesics on the Grassmann manifold, the curves for the small matrix  $S_+(p, p)$ , i.e.,  $R^2(t)$  and  $\gamma_K(t)$  are based on different geodesics. In Bonnabel & Sepulchre (2009), this curve  $R^2(t)$  is a geodesic on  $S_+(p, p)$  for the natural (affine-invariant) metric

$$g_P^{\text{Natural}}(\eta, \nu) := \text{tr}(\eta P^{-1} \nu P^{-1}) = \text{vec}(\eta)^T (P \otimes P)^{-1} \text{vec}(\nu), \quad \text{for all } \eta, \nu \in T_P S_+(p, p),$$

while in our case, the curve  $\gamma_K(t)\gamma_K^T(t)$  is based on the related right-invariant metric for  $S_+(p, p)$  from Prop. 5.3. When  $p = n$  and  $P = YY^T$ , this metric becomes

$$g_P^{\text{Right}}(\eta, \nu) := \text{vec}(\eta)^T (Y \otimes Y) \{(Y^T Y)^3 \otimes (Y^T Y) + (Y^T Y) \otimes (Y^T Y)^3\}^{-1} (Y^T \otimes Y^T) \text{vec}(\nu).$$

Since we can always take  $Y = P^{1/2}$ , the symmetric square root of  $P$ , the metric can also be written as

$$g_P^{\text{Right}}(\eta, \nu) := \text{vec}(\eta)^T (P^2 \otimes I + I \otimes P^2)^{-1} \text{vec}(\nu).$$

Admittedly, metric  $g_P^{\text{Right}}$  is more complicated than  $g_P^{\text{Natural}}$  and finding a connecting geodesic between two s.p.s.d. matrices is more involved for  $g_P^{\text{Right}}$ . On the other hand, since curve  $\gamma_{HK}(t)$  is not a geodesic anyway (for both metric spaces), this is probably not a major disadvantage.

## 8 Conclusions

We introduced a homogenous space geometry for  $S_+(p, n)$ , the symmetric positive semidefinite matrices of fixed rank. By choosing the right-invariant metric on  $GL_n$  we made the canonical projection onto  $GL_n / \text{Stab}_E \simeq S_+(p, n)$  a Riemannian submersion. This had the appealing property that the complete horizontal geodesics of  $GL_n$  could be used as pre-image of the geodesics on  $GL_n / \text{Stab}_E$ . The derivation of an efficient closed-form expression of these geodesics opens the door to the practical application of this complete space to rank-constrained problems involving positive semidefinite matrices. Since the quotient space  $GL_n / \text{Stab}_E$  consists of abstract equivalence classes as elements, we embedded it isometrically in the space of real matrices. This should allow for a more concrete understanding of the vector fields, the metric and the geodesics in terms of classic matrices. As final contribution we compared our geometry with other existing geometries in the literature. Specifically, we compared the geodesics and the quasi-geodesics for the different metrics.

## Acknowledgements

The authors thank S. Bonnabel and R. Sepulchre for helpful discussions about the geometry of fixed-rank matrices.

## Funding

B.V. is a Research Assistant of the Research Foundation Flanders (Belgium). B.V. and S.V. were supported in part by the Research Council K.U.Leuven, CoE EF/05/006 Optimization in Engineering (OPTEC). B.V., P.-A.A. and S.V. present research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

## References

- ABSIL, P.-A., MAHONY, R. & SEPULCHRE, R. (2004) Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.*, **80**, 199–220.
- ABSIL, P.-A., BAKER, C. G. & GALLIVAN, K. A. (2007) Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, **7**, 303–330.
- ABSIL, P.-A., MAHONY, R. & SEPULCHRE, R. (2008) *Optimization Algorithms on Matrix Manifolds*. Princeton University Press.
- ABSIL, P.-A., ISHTEVA, M., DE LATHAUWER, L. & VAN HUFFEL, S. (2009) A geometric Newton method for Oja’s vector field. *Neural Comput.*, **21**, 1415–1433.
- ADLER, R. L., DEDIEU, J.-P., MARGULIES, J. Y., MARTENS, M. & SHUB, M. (2002) Newton’s method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, **22**.
- ALVAREZ, F., BOLTE, J. & MUNIER, J. (2008) A unifying local convergence result for Newton’s method in Riemannian manifolds. *Found. Comput. Math.*, **8**, 197–226.
- BONNABEL, S., MEYER, G. & SEPULCHRE, R. (2010) Adaptive filtering for estimation of a low-rank positive semidefinite matrix. *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems*. MTNS 2010.
- BONNABEL, S. & SEPULCHRE, R. (2009) Geometric distance and mean for positive semi-definite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.*, **31**, 1055–1070.
- BONNABEL, S. & SEPULCHRE, R. (2010) Rank-preserving geometric means of positive semi-definite matrices. *Proc. of the 19th International Symposium on Mathematical Theory of Networks and Systems*. MTNS 2010.
- BOOTHBY, W. M. (1986) *An Introduction to Differentiable Manifolds and Riemannian Geometry*, second edn. Academic Press.
- BOYD, S. & VANDENBERGHE, L. (2004) *Convex Optimization*. Cambridge University Press.
- CHEEGER, J. & EBIN, D. G. (1975) *Comparison Theorems in Riemannian Geometry*. North-Holland Publishing Co.
- DEDIE, J.-P., PRIOURET, P. & MALAJOVICH, G. (2003) Newton’s method on Riemannian manifolds: Covariant alpha theory. *IMA J. Numer. Anal.*, **23**, 395–419.

- FERREIRA, O. P. & SVAITER, B. F. (2002) Kantorovich's theorem on Newton's method in Riemannian manifolds. *J. Complexity*, **18**.
- GALLOT, S., HULIN, D. & LAFONTAINE, J. (2004) *Riemannian Geometry*, third edn. Springer-Verlag.
- HAIRER, E., LUBICH, C. & WANNER, G. (2006) *Geometric Numerical Integration*, second edn. Springer-Verlag.
- HELMKE, U. & MOORE, J. B. (1994) *Optimization and Dynamical Systems*. Springer-Verlag London Ltd.
- HELMKE, U. & SHAYMAN, M. A. (1995) Critical points of matrix least squares distance functions. *Linear Algebra Appl.*, **215**, 1–19.
- HIGHAM, N. J. (2008) *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics.
- HORN, R. & JOHNSON, C. R. (1991) *Topics in Matrix Analysis*. Cambridge University Press.
- HUBER, P. J. (1981) *Robust Statistics*. Wiley.
- JOURNÉE, M., BACH, F., ABSIL, P.-A. & SEPULCHRE, R. (2010) Low-rank optimization on the cone of positive semidefinite matrices. *SIAM J. Optim.*, **20**, 2327–2351.
- KOBAYASHI, S. & NOMIZU, K. (1963) *Foundations of Differential Geometry*. Interscience Publishers, a division of John Wiley & Sons, New York-London.
- LANCKRIET, G. R. G., CRISTIANINI, N., BARTLETT, P., EL GHAOUI, L. & JORDAN, M. I. (2004) Learning the kernel matrix with semidefinite programming. *J. Mach. Learn. Res.*, **5**, 27–72.
- LEE, J. M. (1997) *Riemannian Manifolds: An Introduction to Curvature*, vol. 176. Springer-Verlag, pp. xv+224.
- LEE, S., CHOI, M., KIM, H. & PARK, F. C. (2007) Geometric direct search algorithms for image registration. *IEEE Trans. Image Process.*, **16**, 2215–2224.
- LI, C. & WANG, J. (2006) Newton's method on Riemannian manifolds: Smale's point estimate theory under the  $\gamma$ -condition. *IMA J. Numer. Anal.*, **26**, 228–251.
- MAHONY, R. E. (1994) Optimization algorithms on homogeneous spaces: with applications in linear systems theory. *Ph.D. thesis*, Department of Systems Engineering, Canberra, Australia.
- MARSDEN, J. E. & TUDOR, T. S. (1999) *Introduction to Mechanics and Symmetry*, corrected second edn. Springer.
- MEYER, G., JOURNÉE, M., BONNABEL, S. & SEPULCHRE, R. (2009) From subspace learning to distance learning: a geometrical optimization approach. *Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing*. IEEE, pp. 385–388.
- MEYER, G., BONNABEL, S. & SEPULCHRE, R. (2010) Regression on fixed-rank positive semidefinite matrices: a Riemannian approach. *Submitted. Available at arXiv:1006.1288v1*.
- MILLER, M. I., TROUVE, A. & YOUNES, L. (2003) The metric spaces, Euler equations, and normal geodesic image motions of computational anatomy. *Proceedings of the International Conference on Image Processing*. ICIP, pp. II – 635–638.
- NESTEROV, Y. & TODD, M. (2008) On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.*, **2**, 333–361.

- NISHIMORI, Y. & AKAHO, S. (2005) Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. *Neurocomputing*, **67**, 106–135.
- O’NEILL, B. (1983) *Semi-Riemannian Geometry*. Academic Press Inc.
- ORSI, R., HELMKE, U. & MOORE, J. B. (2004) A Newton-like method for solving rank constrained linear matrix inequalities. *Proceedings of the 43rd IEEE Conference on Decision and Control*. IEEE, pp. 3138–3144.
- ORSI, R., HELMKE, U. & MOORE, J. B. (2006) A Newton-like method for solving rank constrained linear matrix inequalities. *Automatica J. IFAC*, **42**, 1875–1882.
- PENNEC, X., FILLARD, P. & AYACHE, N. (2006) A Riemannian framework for tensor computing. *Int. J. Comput. Vision*, **66**, 41–66.
- PENZL, T. (1998) Numerical solution of generalized Lyapunov equations. *Adv. Comput. Math.*, **8**, 33–48.
- POSTNIKOV, M. M. (2001) *Geometry VI: Riemannian geometry*. Springer.
- SEPULCHRE, R., ABSIL, P.-A. & BONNABEL, S. (2010) Géométrie des matrices semi-définies positives de rang fixé: un peu de théorie, et beaucoup d’applications. *Proceedings of the Sixième Conférence Internationale Francophone d’Automatique*. CIFA.
- SMITH, S. T. (2005) Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.*, **53**, 1610–1630.
- STEWART, G. W. (2001) *Matrix Algorithms, Volume II: Eigensystems*. Society for Industrial and Applied Mathematics.
- STOER, J. & BULIRSCH, R. (1992) *Introduction to Numerical Analysis*, second edn. Springer-Verlag.
- VANDEREYCKEN, B., ABSIL, P.-A. & VANDEWALLE, S. (2009) Embedded geometry of the set of symmetric positive semidefinite matrices of fixed rank. *Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing*. IEEE, pp. 389–392.
- VANDEREYCKEN, B. & VANDEWALLE, S. (2010) A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. *SIAM J. Matrix Anal. Appl. (to appear)*.
- YANG, Y. (2006) Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *J. Optim. Theory Appl.*, **132**, 245–265.