Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices*

Estelle Massart† P.-A. Absil†

October 12, 2019

Abstract: This paper explores the well-known identification of the manifold of rank p positive-semidefinite matrices of size n with the quotient of the set of full-rank n-by-p matrices by the orthogonal group in dimension p. The induced metric corresponds to the Wasserstein metric between centered degenerate Gaussian distributions, and is a generalization of the Bures–Wasserstein metric on the manifold of positive-definite matrices. We compute the Riemannian logarithm, and show that the local injectivity radius at any matrix C is the square root of the p^{th} largest eigenvalue of C. As a result, the global injectivity radius on this manifold is zero. Finally, this paper also contains a detailed description of this geometry, recovering previously known expressions by applying the standard machinery of Riemannian submersions.

Keywords: positive-semidefinite matrices, low-rank, Riemannian quotient manifold, geodesics, Riemannian logarithm, data fitting.

1 Introduction

Positive-semidefinite (PSD) matrices appear in many domains, ranging from optimization (and more particularly semidefinite programming [JBAS10, BV04]) to machine learning and statistics, where they arise respectively as kernels and covariance matrices [LCB$^+$04, Hub81]. They also appear as diffusion tensors in brain imaging [BCS13], as covariance descriptors in image set classification [FHP16] and as Gram matrices in tomography [Rau18].

We consider here situations where the rank of the matrices is assumed to be fixed. This is for example the case when the data points are low-rank representatives of large PSD matrices. The data belong then to the set $\mathcal{S}_+(p,n)$ of fixed-rank PSD matrices of size n and rank p. For example, optimization algorithms on $\mathcal{S}_+(p,n)$ were proposed for distance learning [BMS10, MBS11], distance matrix completion [MMS11], and role model extraction [MHB$^+$16]. Other works face the problem of fitting a curve to a set of points on $\mathcal{S}_+(p,n)$. Geodesic interpolation of Gram matrices is used in [LB14] to generate intermediary protein conformations. In [GMM$^+$17], the authors fit a curve to a set of covariance matrices representing a wind field, while [KDB$^+$18] proposes to use interpolation of Gram matrices for facial expression recognition in videos. Interpolation on $\mathcal{S}_+(p,n)$ was also used in [BHS18] in the framework of dynamical community detection, and in [MGS$^+$19] for parametric model order reduction.

The set $\mathcal{S}_+(p,n)$ does not have a vector space structure: the sum of two PSD matrices of rank p has usually a rank larger than p. However, this set can be turned into a Riemannian manifold; see, e.g., [VAV09]. Many algorithms working on a Riemannian manifold perform part of the computations in a tangent space (a local first order approximation of the manifold), and map the result back onto the manifold. This is for example the case of the steepest descent algorithm, whose iterations are made of two steps: the next iterate is first computed in the tangent space of the current iterate, and then mapped on the manifold. Two tools are therefore useful to implement such algorithms: the

*This work was supported by (i) the Fonds de la Recherche Scientifique – FNRS and the Fonds Wetenschappelijk Onderzoek – Vlaanderen under EOS Project no 30468160 and (ii) "Communauté française de Belgique - Actions de Recherche Concertées" (contract ARC 14/19-060).

†ICTEAM Institute, UCLouvain, Avenue Georges Lemaître 4 bte L4.05.01, 1348 Louvain-la-Neuve, Belgium

exponential and logarithm maps. The exponential map sends a point from a given tangent space to the manifold. The logarithm is its reciprocal operator: it maps a point from the manifold to a given tangent space.

The definition of the exponential and logarithm maps depends on the geometry of the manifold. Different geometries were proposed for the manifold $\mathcal{S}_+(p,n)$, none of them having the desirable property of turning $\mathcal{S}_+(p,n)$ into a geodesically complete manifold with closed-form expressions for both the exponential and the logarithm maps. Completeness implies that the exponential map is defined on the whole tangent bundle (the collection of all tangent spaces to the manifold). For the steepest descent algorithm, this means that any new iterate computed in the tangent space can be mapped back to the manifold. The authors of [VAV09] consider the set $\mathcal{S}_+(p,n)$ as an embedded submanifold of $\mathbb{R}^{n\times n}$. A closed-form expression is provided for the exponential map of some tangent vectors, while numerical solvers have to be used for the others. In [BS09], the set is seen as a quotient manifold $\mathcal{S}_+(p,n) \simeq (\mathrm{St}(n,p) \times \mathcal{P}_p)/\mathcal{O}_p$, with $\mathrm{St}(n,p)$ the (orthogonal) Stiefel manifold, $\mathcal{P}_p$ the manifold of positive-definite matrices and $\mathcal{O}_p$ the orthogonal group. With this geometry, the manifold is complete, and the authors obtain expressions for horizontal geodesics in the structure space $\mathrm{St}(n,p)\times\mathcal{P}_p$. These horizontal geodesics define a curve in $\mathcal{S}_+(p,n)$, but this curve is not necessarily a geodesic. The work [VAV13] manages to turn the set $\mathcal{S}_+(p,n)$ into a complete Riemannian manifold, by endowing it with a homogeneous space geometry. Expressions are provided for the exponential map, but not for the logarithm.

In this paper, we follow the route of [JBAS10] by identifying $\mathcal{S}_+(p,n)$ with the quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, where $\mathbb{R}_*^{n\times p}$ is the set of full-rank $n\times p$ matrices and $\mathcal{O}_p$ is the orthogonal group of order p. The total space $\mathbb{R}_*^{n\times p}$ is equipped with the Euclidean metric.

There are two main reasons to consider this geometry. The first one is the fact that the computation cost of the resulting exponential and logarithm maps is low in comparison with the other above-mentioned geometries (the expression for the exponential map is already given in [VAV13, §7.2], and we compute in this paper the Riemannian logarithm). As a result, this geometry is particularly suitable for numerical computations.

The second motivation to consider this geometry is the fact that the associated distance coincides with the Wasserstein distance between degenerate centered Gaussian distributions. Indeed, any degenerate centered Gaussian distribution is parameterized by a positive-semidefinite covariance matrix. The Wasserstein metric between degenerate centered Gaussian distributions induces then a distance between positive-semidefinite matrices, that coincides with the distance on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ computed in this paper. The latter is also a direct generalization of the Bures–Wasserstein distance between positive-definite matrices, presented in, e.g., [Tak11, BJL16].

The main drawback of this geometry is that it does not turn the manifold into a complete metric space. This drawback is mitigated by two observations. First, this situation is not isolated, see, e.g., [AO14] that proposes several retractions (first-order approximations of the exponential map) on the low-rank manifold, that are not defined everywhere. Secondly, several recent works take into account situations where the exponential map (or more generally the retraction) is not defined everywhere, see, e.g., [BAC19, KM18].

Optimization on the manifold $\mathcal{S}_+(p,n)$, endowed with the quotient geometry considered here, has already been used in several papers, see [BMS10, JBAS10, MBS11, MMS11]. The authors of [JBAS10] develop optimization algorithms on (constrained subsets of) $\mathcal{S}_+(p,n)$, and apply them to the computation of the maximal cut of a graph, and to sparse principal component analysis. The works [BMS10] and [MBS11] use first-order optimization algorithms on $\mathcal{S}_+(p,n)$, for several geometries including the quotient representation considered here, for learning a regression model parameterized by a PSD matrix. The work [MMS11] proposes a gradient descent and a trust region algorithm on $\mathcal{S}_+(p,n)$ for low-rank distance matrix completion.

Less related to the present work, we also mention that the quotient $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ was used in [AIDV08] to build a Newton method for computing the zeroes of the Oja's vector field, while its complex variant $\mathbb{C}_*^{n\times p}/\mathcal{O}_p$ was used in [HGZ17] to solve the phase retrieval problem. However, in these last two examples, the total space is endowed with another metric. We finally indicate that related manifolds, defined as quotients (by the orthogonal group) of subsets of $\mathbb{R}_*^{n\times p}$, were investigated in [DH18].

All these works use the (already known) expressions for, e.g., the Riemannian exponential, projectors on the horizontal and vertical spaces, and the gradient of a cost function. The main contribution of the present paper is the computation of the Riemannian logarithm and injectivity radius on this manifold. This last concept is required to guarantee convergence of several optimization [ATV13, Bon13] and consensus algorithms [TAV13]. It also allows to ensure continuity of the results of some data fitting algorithms [AGSW16, GMA18]. Moreover, we also obtain expressions for the resulting distance on the manifold, Lie derivatives, and summarize some recent results on the curvature of the manifold.

During the writing of the present paper, we made our implementation of the Riemannian logarithm available in the Manopt toolbox [BMAS14]. These implementations have been used in [GMM$^+$17] for wind field modeling, in [MGS$^+$19] for parametric model order reduction, in [SAV$^+$19] for MEG/EEG signal processing and in [SDB$^+$19] for action recognition in video frames.

The structure of the paper is as follows. Section 2 briefly presents the required geometric concepts on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$. In Section 3, we study the domain of the exponential map. Section 4 is devoted to the logarithm map, while we consider respectively, in Section 5, 6 and 7, the Riemannian distance, the injectivity radius, and the Lie derivative. Finally, Section 8 provides a numerical application.

In the appendix (Section A), for reference purposes, we give an overview of known or easy-to-obtain facts about $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$.

2 Quotient geometry of $\mathcal{S}_+(p,n)$: preliminaries

This section provides the required background regarding the identification of the manifold $\mathcal{S}_+(p,n)$ with the quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, where

$$\mathbb{R}_*^{n\times p} := \{Y \in \mathbb{R}^{n\times p} : \det\left(Y^\top Y\right) \neq 0\}$$

is the set of full-rank $n \times p$ matrices, and

$$\mathcal{O}_p := \{Q \in \mathbb{R}^{p\times p} : Q^\top Q = I\}$$

is the orthogonal group in dimension p. More detail (with the corresponding proofs) is given in Section A.

2.1 The quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$

The identification of the manifold $\mathcal{S}_+(p,n)$ with the quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is motivated by the following result (see Section A for the proof).

Proposition 2.1. *Let $Y_1, Y_2 \in \mathbb{R}_*^{n\times p}$. Then $Y_1 Y_1^\top = Y_2 Y_2^\top$ if and only if $Y_2 = Y_1 Q$ for some $Q \in \mathcal{O}_p$, and*

$$\mathcal{S}_+(p,n) = \{YY^\top : Y \in \mathbb{R}_*^{n\times p}\}.$$

Proposition 2.1 gives us a convenient way of representing a matrix $S \in \mathcal{S}_+(p,n)$: pick $Y \in \mathbb{R}_*^{n\times p}$ such that $S = YY^\top$. Observe that the matrix S contains n^2 elements, whereas Y only contains np elements, which is an appreciable improvement in the frequently encountered case where p is much smaller than n. However, this representation comes with a redundancy which is gracefully tackled by a quotient set approach. This approach consists of considering as a single *point* the set $Y\mathcal{O}_p := \{YQ : Q \in \mathcal{O}_p\}$ of all Y's that yield the same S. The set of all those *points* (also referred to as *equivalence classes*) is the quotient set $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$.

Let us now introduce the mapping $\phi : \mathbb{R}_*^{n\times p} \to \mathcal{S}_+(p,n) : Y \mapsto YY^\top$, the mapping $\Phi : \mathbb{R}_*^{n\times p}/\mathcal{O}_p \to \mathcal{S}_+(p,n) : Y\mathcal{O}_p \mapsto YY^\top$, and the quotient map $\pi : \mathbb{R}_*^{n\times p} \to \mathbb{R}_*^{n\times p}/\mathcal{O}_p$. As illustrated by Figure 1, the mapping Φ is a bijection, and thus provides an identification of $\mathcal{S}_+(p,n)$ with the quotient set $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$.

It is well known that $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is a *quotient manifold* of dimension $pn - \frac{p(p-1)}{2}$. The mapping Φ is a diffeomorphism between the quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ and the submanifold $\mathcal{S}_+(p,n)$ of $\mathbb{R}^{n\times n}$.

$$
\begin{tikzcd}
\mathbb{R}_*^{n\times p} \arrow[d, "\pi"'] \arrow[dr, "\phi"] & \\
\mathbb{R}_*^{n\times p}/\mathcal{O}_p \arrow[r, "\mathrm{bij}", "\Phi"'] & \mathcal{S}_+(p,n)
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
Y \arrow[d, "\pi"'] \arrow[dr, "\phi"] & \\
Y\mathcal{O}_p \arrow[r, "\mathrm{bij}", "\Phi"'] & YY^\top
\end{tikzcd}
$$

Figure 1: Commutative diagram for ϕ, π and Φ.

2.2 Horizontal and vertical spaces

Let the manifold $\mathbb{R}_*^{n\times p}$ be endowed with the Euclidean metric $\langle Z_1, Z_2\rangle_Y := \operatorname{tr}(Z_1^\top Z_2)$ for all $Z_1, Z_2 \in \mathcal{T}_Y\mathbb{R}_*^{n\times p}$, where the notation $\operatorname{tr}(A)$ stands for the trace of the matrix A. The tangent space $\mathcal{T}_Y\mathbb{R}_*^{n\times p}$, which can be identified with $\mathbb{R}^{n\times p}$, is the direct sum of the *vertical* and *horizontal* spaces at Y, defined as:

$$\mathcal{V}_Y := \{Y\Omega : \Omega = -\Omega^\top\} \qquad \text{and} \qquad \mathcal{H}_Y := \{Z \in \mathbb{R}^{n\times p} : Y^\top Z - Z^\top Y = 0\}. \tag{1}$$

The vertical space $\mathcal{V}_Y$ is the tangent space at Y to the equivalence class $\pi(Y)$, while the horizontal space $\mathcal{H}_Y$ is its orthogonal complement. The latter can also be written as

$$\mathcal{H}_Y = \{Y(Y^\top Y)^{-1}H + Y_\perp K : H = H^\top \in \mathbb{R}^{p\times p}, K \in \mathbb{R}^{(n-p)\times p}\},$$

where $Y_\perp \in \mathbb{R}^{n\times(n-p)}$ is orthonormal and satisfies $Y^\top Y_\perp = 0$.

Any tangent vector $\xi_{\pi(Y)} \in \mathcal{T}_{\pi(Y)}\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is represented by a unique horizontal vector $\bar{\xi}_Y \in \mathcal{H}_Y$, satisfying $\mathrm{D}\pi(Y)[\bar{\xi}_Y] = \xi_{\pi(Y)}$. Throughout this paper, we use the notation $\bar{\xi}$ to denote the unique horizontal lift of the vector field ξ on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, in the sense of [KN63, Chap. II, Prop. 1.2].

The projections of any $Z \in \mathbb{R}^{n\times p}$ on the horizontal and vertical spaces, denoted by $\mathrm{P^v}_Y(Z)$ and $\mathrm{P^h}_Y(Z)$, are given by:

$$\mathrm{P^v}_Y(Z) = Y\,\mathbf{T}_{Y^\top Y}^{-1}(Y^\top Z - Z^\top Y), \tag{2}$$

$$\mathrm{P^h}_Y(Z) = Z - \mathrm{P^v}_Y(Z). \tag{3}$$

where $\mathbf{T}_E^{-1}(Z)$ is the solution X to the Sylvester equation $EX + XE = Z$, which is unique if E is symmetric positive-definite.

Observe that $\mathbf{T}_E^{-1}(Z)$ can be easily computed by diagonalization of the positive-definite matrix $E := Y^\top Y$. Indeed, let $E := U\Lambda U^\top$ be an eigenvalue decomposition. The Sylvester equation $EX + XE = Z$ becomes

$$U\Lambda U^\top X + XU\Lambda U^\top = Z.$$

Left- and rightmultiplying this equation by respectively $U^\top$ and U yields:

$$\Lambda U^\top XU + U^\top XU\Lambda = U^\top ZU,$$

which can be easily solved for $\tilde{X} := U^\top XU$. This methodology comes from [BR97, §10].

2.3 Riemannian metric

The Euclidean metric on $\mathbb{R}_*^{n\times p}$ induces a Riemannian metric on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$. For all $\xi_{\pi(Y)}, \eta_{\pi(Y)} \in \mathcal{T}_{\pi(Y)}\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, with horizontal lifts $\bar{\xi}_Y, \bar{\eta}_Y \in \mathcal{H}_Y$, let us define

$$g_{\pi(Y)}(\xi_{\pi(Y)}, \zeta_{\pi(Y)}) := \operatorname{tr}\left(\bar{\xi}_Y{}^\top \bar{\zeta}_Y\right). \tag{4}$$

This metric turns the quotient map π into a Riemannian submersion.

2.4 Retraction and exponential map

The authors of [JBAS10] also provide a retraction on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$:

$$R_{\pi(Y)}\xi_{\pi(Y)} := \pi(Y+\bar{\xi}_Y). \tag{5}$$

In view of [VAV13, §7.2], (5) is actually the exponential map on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$.

Theorem 2.2. *Let us define the set*

$$\mathcal{D}_Y := \{\bar{\xi}_Y \in \mathcal{H}_Y | \mathrm{rank}(Y+t\bar{\xi}_Y) = p \ \forall t \in [0,1]\}.$$

For all $\xi_{\pi(Y)} \in \mathrm{D}\pi(Y)[\mathcal{D}_Y]$, the exponential map on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is given by:

$$\mathrm{Exp}_{\pi(Y)}\xi_{\pi(Y)} = \pi(Y+\bar{\xi}_Y), \tag{6}$$

i.e., geodesics in $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ are images, through the quotient map π, of straight lines $t \mapsto Y+t\bar{\xi}_Y$ in $\mathbb{R}_*^{n\times p}$, restricted to the time interval around $t=0$ where $Y+t\bar{\xi}_Y$ remains full rank.

3 Domain of the exponential map

As mentioned in the previous section, geodesics in the quotient are projections of horizontal geodesics in $\mathbb{R}_*^{n\times p}$, i.e., straight lines that remain in $\mathbb{R}_*^{n\times p}$. This last condition introduces some restrictions on the initial velocity of the curve, as for an arbitrary $\bar{\xi}_Y \in \mathcal{H}_Y$, the curve $t \mapsto Y+t\bar{\xi}_Y$ does not necessarily remain full-rank. This section is aimed at describing the set of allowed horizontal vectors, i.e., the set $\mathcal{D}_Y$ in Theorem 2.2. We first need to introduce some notation.

Definition 3.1. We define the mapping

$$\overline{\mathrm{EXP}}_Y : \mathcal{H}_Y \to \mathbb{R}^{n\times p} : \bar{\xi}_Y \mapsto Y+\bar{\xi}_Y.$$

The set that is mapped through $\overline{\mathrm{EXP}}_Y$ on the set of full-rank matrices is:

$$\mathcal{F}_Y := \overline{\mathrm{EXP}}_Y^{-1}(\mathbb{R}_*^{n\times p}) = \{\bar{\xi}_Y \in \mathcal{H}_Y : \mathrm{rank}(\overline{\mathrm{EXP}}_Y\bar{\xi}_Y) = p\}.$$

The domain $\mathcal{D}_Y$ of the exponential map on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is the largest star-shaped domain emanating from Y that is contained in $\mathcal{F}_Y$:

$$\mathcal{D}_Y := \{\bar{\xi}_Y \in \mathcal{F}_Y | t\bar{\xi}_Y \in \mathcal{F}_Y \ \forall t \in [0,1]\} = \{\bar{\xi}_Y \in \mathcal{H}_Y | \mathrm{rank}(Y+t\bar{\xi}_Y) = p \ \forall t \in [0,1]\}.$$

We define $\overline{\mathrm{Exp}}_Y$ as the restriction of $\overline{\mathrm{EXP}}_Y$ to the set $\mathcal{D}_Y$:

$$\overline{\mathrm{Exp}}_Y := \overline{\mathrm{EXP}}_Y|_{\mathcal{D}_Y},$$

and we have $\mathrm{Exp}_{\pi(Y)}(\xi) = \pi \circ \overline{\mathrm{Exp}}_Y(\bar{\xi}_Y)$ for all $\xi \in \mathrm{D}\pi(Y)[\mathcal{D}_Y]$. We also define

$$\mathrm{EXP}_{\pi(Y)}(\xi) := \pi \circ \overline{\mathrm{EXP}}_Y(\bar{\xi}_Y),$$

with $\xi \in D\pi(Y)[\mathcal{F}_Y]$. Finally, we let $\mathcal{F}_Y^c := \mathcal{H}_Y \backslash \mathcal{F}_Y$ and $\mathcal{D}_Y^c := \mathcal{H}_Y \backslash \mathcal{D}_Y$.

The goal of the following results is to study the set $\mathcal{D}_Y$, i.e., the domain of $\overline{\mathrm{Exp}}_Y$. We first describe the larger set $\mathcal{F}_Y$.

Proposition 3.2. *Let $\bar{\xi}_Y \in \mathcal{H}_Y$, decomposed as $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H + Y_\perp K$. The condition $\bar{\xi}_Y \in \mathcal{F}_Y^c$, i.e., $\mathrm{rank}(\overline{\mathrm{EXP}}_Y\bar{\xi}_Y) < p$, is equivalent to the condition:*

$$\ker(I+(Y^\top Y)^{-1}H) \bigcap \ker(K) \neq \{0\}, \tag{7}$$

where the notation $\ker(A)$ stands for the null space of the matrix A, i.e., the set of vectors v satisfying $Av=0$.

Proof. The inequality $\mathrm{rank}(\overline{\mathrm{EXP}}_Y\bar{\xi}_Y) < p$ is equivalent to the existence of a non-zero vector $v \in \mathbb{R}^p$ such that $(\overline{\mathrm{EXP}}_Y\bar{\xi}_Y)v = 0$. This can be written as $(Y+\bar{\xi}_Y)v = Y(I+(Y^\top Y)^{-1}H)v+Y_\perp Kv = 0$. Since the ranges of $Y(I+(Y^\top Y)^{-1}H)$ and $Y_\perp K$ are disjoint, and since Y and $Y_\perp$ have full column-rank, the last equality is true if and only if $v \in \ker(I+(Y^\top Y)^{-1}H)\bigcap\ker(K)$. □

We say that t is a *singular time* of $\bar{\xi}_Y \in \mathcal{H}_Y$ if $t\bar{\xi}_Y \in \mathcal{F}_Y^c$. We also introduce the notation $A \succ 0$ and $A \succeq 0$ to say that A is positive-definite or positive-semidefinite, respectively. The next results readily follow.

Corollary 3.3. *Let $\bar{\xi}_Y \in \mathcal{H}_Y$, decomposed as $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H + Y_\perp K$.*

1. *If t is a singular time of $\bar{\xi}_Y$, then $t \neq 0$ and $-1/t$ is an eigenvalue of $(Y^\top Y)^{-1}H$. (The eigenvalues of $(Y^\top Y)^{-1}H$ are real since it is similar to $(Y^\top Y)^{-1/2}H(Y^\top Y)^{-1/2}$.)*
2. *The number of singular times of $\bar{\xi}_Y$ is at most p.*
3. *$\bar{\xi}_Y$ has p singular times if and only if $(Y^\top Y)^{-1}H$ has p distinct nonzero eigenvalues and $K = 0$.*
4. *$\bar{\xi}_Y \in \mathcal{D}_Y$ if and only if for all $\lambda \in \Lambda((Y^\top Y)^{-1}H)\bigcap(-\infty,-1]$, there holds:*
$$\ker\left(I - \frac{1}{\lambda}(Y^\top Y)^{-1}H\right)\bigcap\ker(K) = \{0\}, \tag{8}$$
with $\Lambda(A)$ the spectrum of A.
5. *If $H \succeq 0$, then $t\bar{\xi}_Y := t(Y(Y^\top Y)^{-1}H + Y_\perp K) \in \mathcal{D}_Y \subset \mathcal{F}_Y$ for all $t \geq 0$.*
6. *There exist vectors $\bar{\xi}_Y$ without singular times. This is, e.g., the case when $H = 0$ and also when $\ker(K) = \{0\}$. Hence there exist complete geodesics in $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$.*
7. *$\mathcal{D}_Y \subset \mathcal{F}_Y$*
8. *$\mathcal{D}_Y, \mathcal{F}_Y, \mathcal{D}_Y^c, \mathcal{F}_Y^c$ are nonempty and unbounded.*
9. *Since $\mathcal{D}_Y^c$ is nonempty, $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is not complete.*

We show that if $||\bar{\xi}_Y||_{\mathrm{F}}$ is small enough (where $||A||_{\mathrm{F}}$ is the Frobenius norm of the matrix A), then $\bar{\xi}_Y \in \mathcal{D}_Y \subset \mathcal{F}_Y$. We compute the maximum value of $||\bar{\xi}_Y||_{\mathrm{F}}$ that guarantees that $\bar{\xi}_Y \in \mathcal{D}_Y \subset \mathcal{F}_Y$.

Proposition 3.4. *Let $Y = U\Sigma V^\top$ be a singular value decomposition, with singular values $\sigma_1 \geq ... \geq \sigma_p$. Then*
$$\min_{\bar{\xi}_Y\in\mathcal{F}_Y^c} ||\bar{\xi}_Y||_{\mathrm{F}} = \min_{\bar{\xi}_Y\in\mathcal{D}_Y^c} ||\bar{\xi}_Y||_{\mathrm{F}} = \sigma_p, \tag{9}$$
and the vector $\bar{\xi}_Y^* := -\sigma_p^2 Y(Y^\top Y)^{-1}v_p v_p^\top$ is a minimizer, with v_p a right singular vector associated to σ_p.

Proof. The first equality in (9) is obtained using the fact that $\mathcal{F}_Y^c \subset \mathcal{D}_Y^c$, and that for any $\bar{\xi}_Y \in \mathcal{D}_Y^c\backslash\mathcal{F}_Y^c$, there exists $t\bar{\xi}_Y \in \mathcal{F}_Y^c$, with $t \in [0,1]$. Since $\bar{\xi}_Y \in \mathcal{F}_Y^c$ implies $\tilde{Y} := Y + \bar{\xi}_Y \notin \mathbb{R}_*^{n\times p}$, we have
$$\min_{\bar{\xi}_Y\in\mathcal{F}_Y^c} ||\bar{\xi}_Y||_{\mathrm{F}} \geq \min_{\tilde{Y}\notin\mathbb{R}_*^{n\times p}} ||\tilde{Y} - Y||_{\mathrm{F}} = \sigma_p,$$
where the second equality comes from the Schmidt-Mirsky theorem [Ste98, Chap 1., Thm. 4.32]. Next, we have that $\bar{\xi}_Y^* \in \mathcal{H}_Y$, and
$$Y + \bar{\xi}_Y^* = Y(Y^\top Y)^{-1}\left(Y^\top Y - \sigma_p^2 v_p v_p^\top\right) = Y(Y^\top Y)^{-1}\left(\sum_{i=1}^{p-1}\sigma_i^2 v_i v_i^\top\right),$$
hence $Y + \bar{\xi}_Y^* \notin \mathbb{R}_*^{n\times p}$. Thus $\bar{\xi}_Y^* \in \mathcal{F}_Y^c$. Observe now that $(Y^\top Y)^{-1} = V\Sigma^{-2}V^\top$, hence the norm of $\bar{\xi}_Y^*$ can be written as:
$$||\bar{\xi}_Y^*||_{\mathrm{F}} = \sigma_p^2||Y(Y^\top Y)^{-1}v_p v_p^\top||_{\mathrm{F}} = \sigma_p^2||U\Sigma^{-1}V^\top v_p v_p^\top||_{\mathrm{F}} = \sigma_p^2||u_p\sigma_p^{-1}v_p^\top||_{\mathrm{F}} = \sigma_p\sqrt{\mathrm{tr}\left(v_p u_p^\top u_p v_p^\top\right)} = \sigma_p.$$

We have thus obtained

$$\sigma_p = ||\bar{\xi}_Y^*||_{\mathrm{F}} \geq \min_{\bar{\xi}_Y \in \mathcal{F}_Y^c} ||\bar{\xi}_Y||_{\mathrm{F}} = \min_{\bar{\xi}_Y \in \mathcal{D}_Y^c} ||\bar{\xi}_Y||_{\mathrm{F}} \geq \sigma_p,$$

and the conclusion follows. □

Corollary 3.5. *If $||\bar{\xi}_Y||_{\mathrm{F}} < \sigma_p(Y)$ (with $\sigma_p(Y)$ the smallest singular value of Y), then $\bar{\xi}_Y \in \mathcal{D}_Y \subset \mathcal{F}_Y$.*

Corollary 3.6. *The largest open ball in $\mathcal{D}_Y$ centered at 0_Y, the null element of the horizontal space $\mathcal{H}_Y$, has radius $\sigma_p(Y)$.*

The next result shown that $\mathcal{D}_Y$ is generic in $\mathcal{H}_Y$ in the sense of measure theory.

Proposition 3.7. *Let $\bar{\xi}_Y$ be drawn from a nonsingular probability distribution on $\mathcal{H}_Y$ (i.e., a probability distribution that is not concentrated on a set of Lebesgue measure zero). If $p < n$, then $\bar{\xi}_Y \in \mathcal{D}_Y$ (and thus $\mathcal{F}_Y$) with probability one.*

Proof. Due to Corollary 3.3, the vector $\bar{\xi}_Y$ belongs to $\mathcal{D}_Y$ (and thus $\mathcal{F}_Y$) with probability one if and only if

$$\begin{aligned} &P\left(\ker\left(I - \frac{1}{\lambda}(Y^\top Y)^{-1}H\right) \bigcap \ker(K) = \{0\}\ \forall \lambda \in \Lambda((Y^\top Y)^{-1}H) \bigcap (-\infty, -1]\right) = 1 \\ \Leftrightarrow &\sum_{\lambda \in \Lambda((Y^\top Y)^{-1}H) \bigcap (-\infty,-1]} P\left(\exists v \neq 0 | v \in \ker\left(I - \frac{1}{\lambda}(Y^\top Y)^{-1}H\right) \text{ and } Kv = 0\right) = 0. \end{aligned}$$

The matrix $(Y^\top Y)^{-1}H$ has distinct eigenvalues with probability one. As a result, with probability one, the eigenspaces associated to each eigenvalue are one-dimensional. In other words, there exists a unique vector $\tilde{v}_\lambda \in \ker(I - \frac{1}{\lambda}(Y^\top Y)^{-1}H)$ such that $||\tilde{v}_\lambda|| = 1$ and $\ker(I - \frac{1}{\lambda}(Y^\top Y)^{-1}H) = \{t\tilde{v}_\lambda | t \in \mathbb{R}\}$. The above-mentioned condition becomes:

$$\begin{aligned} &\sum_{\lambda \in \Lambda((Y^\top Y)^{-1}H) \bigcap (-\infty,-1]} P\left(\exists v \neq 0 | v \in \ker\left(I - \frac{1}{\lambda}(Y^\top Y)^{-1}H\right) \text{ and } Kv = 0\right) = 0 \\ \Leftrightarrow &\sum_{\lambda \in \Lambda((Y^\top Y)^{-1}H) \text{and} (-\infty,-1]} P\left(\langle K_{1,:}, \tilde{v}_\lambda \rangle = 0 \text{ and } \dots \text{ and } \langle K_{n-p,:}, \tilde{v}_\lambda \rangle = 0\right) = 0. \end{aligned}$$

The last equality is valid as the probability of an arbitrary vector to be orthogonal to a set of $n-p$ arbitrary vectors is zero. □

The next result shows that $\mathcal{D}_Y$ is also generic in $\mathcal{H}_Y$ in the sense of the topology.

Proposition 3.8. *The set $\mathcal{D}_Y$ is open relative to $\mathcal{H}_Y$. Moreover, the closure of $\mathcal{D}_Y$ is $\mathcal{H}_Y$ when $p < n$.*

Proof. We prove the first claim, namely that $\mathcal{D}_Y$ is open relatively to $\mathcal{H}_Y$. Due to Corollary 3.3, $\bar{\xi}_Y := Y(Y^\top Y)^{-1}H + Y_\perp K \in \mathcal{D}_Y$ if and only if, for all $\lambda \in (-\infty, -1]$, $\ker(M_{\lambda,H,K}) = \{0\}$, with $M_{\lambda,H,K}$ defined as

$$M_{\lambda,H,K} := \begin{bmatrix} I - \frac{1}{\lambda}(Y^\top Y)^{-1}H \\ K \end{bmatrix}. \tag{10}$$

Let us define the function

$$f : (\lambda, H, K) \mapsto \det\left(M_{\lambda,H,K}^\top M_{\lambda,H,K}\right).$$

Then, $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H + Y_\perp K \in \mathcal{D}_Y$ if and only if $f(\lambda, H, K) \neq 0$ for all $\lambda \in (-\infty, -1]$. To complete the proof of the first claim, it remains to show that for all δH and δK sufficiently small, it holds that $f(\lambda, H+\delta H, K+\delta K) \neq 0$ for all $\lambda \in (-\infty, -1]$, i.e., $\bar{\xi}_Y + \delta\bar{\xi}_Y := Y(Y^\top Y)^{-1}(H+\delta H) + Y_\perp(K+\delta K) \in \mathcal{D}_Y$. There holds

$$f(\lambda, H, K) = \det\left(I - \frac{1}{\lambda}H(Y^\top Y)^{-1} - \frac{1}{\lambda}(Y^\top Y)^{-1}H + \frac{1}{\lambda^2}H(Y^\top Y)^{-2}H + K^\top K\right).$$

Let $\rho_{H,K} := \inf_{\lambda\in(-\infty,-1]} f(\lambda, H, K)$. Since $(-\infty,-1] \ni \lambda \mapsto f(\lambda, H, K)$ is continuous, has no zero, and $\lim_{\lambda\to-\infty} f(\lambda, H, K) = \det\left(I + K^\top K\right) \geq 1$, it follows that $\rho_{H,K} > 0$. Since $\mathrm{D}\det(A)[\dot{A}] = \mathrm{tr}\left(\mathrm{adj}(A)\dot{A}\right)$, there exists $L_{H,K}$ such that for all δH and δK sufficiently small and for all $\lambda \in (-\infty,-1]$:

$$|f(\lambda, H + \delta H, K + \delta K) - f(\lambda, H, K)| \leq L_{H,K} (\|\delta H\|+\|\delta K\|) .$$

Then for all $\delta H, \delta K$ such that $\|\delta H\| < \frac{\rho_{H,K}}{2L_{H,K}}$ and $\|\delta K\| < \frac{\rho_{H,K}}{2L_{H,K}}$, one has $f(\lambda, H + \delta H, K + \delta K) > 0$ for all $\lambda \in (-\infty,-1]$, and the proof of the first claim is complete.

We now prove the second claim, i.e., $\mathrm{cl}(\mathcal{D}_Y) = \mathcal{H}_Y$ if $p < n$. Let $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H + Y_\perp K \in \mathcal{D}_Y^c$, and let $\epsilon > 0$. There exists δH symmetric with $\|\delta H\| < \epsilon$ such that all the eigenvalues of $(Y^\top Y)^{-1}(H + \delta H)$ are distinct. Let V be a matrix whose columns are the eigenvectors of $(Y^\top Y)^{-1}(H+\delta H)$. Observe that V is invertible since $(Y^\top Y)^{-1}(H + \delta H)$ is similar to the symmetric matrix $(Y^\top Y)^{-1/2}(H + \delta H)(Y^\top Y)^{-1/2}$. Let $\delta K := EV^{-1}$, where E is the matrix of all ones of size $(n-p)\times p$; this is a valid choice since it is assumed that $p < n$. Observe that $\delta K V e_i \neq 0$ for all i, and thus $\hat{\rho} := \min\{\rho > 0 : \exists i : (K + \rho\delta K)Ve_i = 0\} > 0$ since the min is over a finite (possibly empty) subset of $[0,\infty)$. We have that, for all $\rho \in (0, \hat{\rho})$,

$$\bar{\xi}_Y + \delta\bar{\xi}_Y := \bar{\xi}_Y + Y(Y^\top Y)^{-1}\delta H + \rho Y_\perp \delta K = Y(Y^\top Y)^{-1}(H + \delta H) + Y_\perp(K + \rho\delta K)$$

is in $\mathcal{D}_Y$. Indeed, for all $\lambda \in (-\infty,-1]$, either λ is not an eigenvalue of $(Y^\top Y)^{-1}(H + \delta H)$ and thus $\ker\left(I - \frac{1}{\lambda}(Y^\top Y)^{-1}(H + \delta H)\right) = \{0\} = \ker(M_{\lambda,H+\delta H,K+\rho\delta K})$; or λ is an eigenvalue of $(Y^\top Y)^{-1}(H + \delta H)$ with eigenvector v_λ (which is one of the columns of V), thus $\ker\left(I - \frac{1}{\lambda}(Y^\top Y)^{-1}H\right) = \mathbb{R}v_\lambda$ but $(K + \rho\delta K)v_\lambda \neq 0$, and thus $\ker(M_{\lambda,H+\delta H,K+\rho\delta K}) = \{0\}$ again. In conclusion, there is $\delta\bar{\xi}_Y$ arbitrarily small such that $\bar{\xi}_Y + \delta\bar{\xi}_Y \in \mathcal{D}_Y$.

□

This last result addresses the existence of closed geodesics on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$.

Proposition 3.9. *There is no closed geodesic, i.e.,* $\mathrm{Exp}_{\pi(Y)}\xi = \pi(Y)$ *if and only if* $\xi = 0_{\pi(Y)}$.

Proof. Assume that there exists $\bar{\xi}_Y \in \mathcal{H}_Y$, i.e., $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H + Y_\perp K$, such that $Y + \bar{\xi}_Y = YQ$ for some matrix $Q \in \mathcal{O}_p$. This can be written as:

$$Y + Y(Y^\top Y)^{-1}H + Y_\perp K = YQ,$$

which implies that $K = 0$. So, there remains:

$$Y + Y(Y^\top Y)^{-1}H = YQ.$$

Left-multiplying both sides of the equation by $Y^\top$ yields:

$$Y^\top Y + H = Y^\top YQ.$$

If $Y^\top Y + H \succeq 0$, then the only solution is $Q = I$ and $H = 0$ (due to the uniqueness of the polar decomposition, see [LT85, §5.7]). Otherwise, if $Y^\top Y + H \not\succeq 0$, $(Y^\top Y)^{-1}H$ must have an eigenvalue $\lambda \leq -1$, with associate eigenvector $v \neq 0$. As $Kv = 0$ (since $K = 0$), this implies by Corollary 3.3 that $\bar{\xi}_Y \notin \mathcal{D}_Y$. □

4 Logarithm map

In this section, we compute the inverse exponential map, i.e., the logarithm map. We show that, for $Y_1, Y_2 \in \mathbb{R}_*^{n\times p}$ such that $Y_1^\top Y_2$ is nonsingular, there is only one minimizing geodesic in $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ from $\pi(Y_1)$ to $\pi(Y_2)$, hence the logarithm map is well defined. Its horizontal lift at Y_1 is obtained from the polar decomposition of $Y_1^\top Y_2$.

The next lemma will be useful to characterize all the solutions to the equation $\mathrm{EXP}_{\pi(Y_1)}\xi = \pi(Y_2)$.

Lemma 4.1. *Given a matrix $A \in \mathbb{R}^{p\times p}$ of rank r, a pair of matrices H, Q, with $H \in \mathbb{R}^{p\times p}$ symmetric and $Q \in \mathcal{O}_p$, satisfies $A = HQ$ if and only if H and Q can be written as:*

$$H = [UU_\perp] \begin{pmatrix} \pm D_{11} & & & \\ & \ddots & & \\ & & \pm D_{rr} & \\ & & & 0_{p-r} \end{pmatrix} [UU_\perp]^\top =: [UU_\perp] \begin{pmatrix} D\mathcal{I} & \\ & 0_{p-r} \end{pmatrix} [UU_\perp]^\top, \tag{11}$$

$$Q = [UU_\perp] \begin{pmatrix} \mathcal{I} & \\ & \tilde{Q} \end{pmatrix} [VV_\perp]^\top, \qquad \tilde{Q} \in \mathcal{O}_{p-r}, \tag{12}$$

with $A = [UU_\perp]\mathrm{Diag}\,(D, 0_{p-r})\,[VV_\perp]^\top$ a singular value decomposition. In the right-hand side of (11), *$\mathcal{I}$ is a diagonal matrix whose diagonal element $\mathcal{I}_{ii}$ is equal to 1 if the positive sign is chosen for D_{ii} and -1 otherwise. As a result, there are at least (in view of the lack of essential uniqueness of the singular value decomposition in case of repeated singular values) 2^p possible factorizations $A = HQ$ if $r \in \{p-1, p\}$ and an infinite number of factorizations if $r \leq p-2$.*

Proof. If H and Q satisfy respectively (11) and (12), then $HQ = [UU_\perp]\mathrm{Diag}\,(D, 0_{p-r})\,[VV_\perp]^\top = A$. Conversely, let $H = H^\top$ and $Q \in \mathcal{O}_p$ be such that $A = HQ$. This implies that $H^2 = AA^\top = UDU^\top$. By the spectral theorem, there exists P and Λ such that $H = [PP_\perp]\mathrm{Diag}\,(\Lambda, 0_{p-r})\,[PP_\perp]^\top$. Hence, $H^2 = P\Lambda^2P^\top$. The equality $H^2 = AA^\top$ implies that the eigenvalues of H^2 are the same as the eigenvalues of $AA^\top$. Moreover, the eigenspaces must coincide as well. So, the columns of P form a basis for the eigenspaces of $AA^\top$ and, as a result, the product (11) generates all possible solutions. Observe now that (11) implies that $A = H[UU_\perp]\mathrm{Diag}\,(\mathcal{I}, M)\,[VV_\perp]^\top$, for any $M \in \mathbb{R}^{(p-r)\times(p-r)}$. Requiring $Q := [UU_\perp]\mathrm{Diag}\,(\mathcal{I}, M)\,[VV_\perp]^\top$ to be orthogonal leads to (12). □

Observe that if $\mathcal{I}$ is the identity matrix, then H is positive semidefinite and H and Q are respectively the factors H_{pol} and Q_{pol} of a polar decomposition of A:

$$A =: H_{\mathrm{pol}} Q_{\mathrm{pol}}.$$

If $r = p$ (i.e., A is invertible), H_{pol} is then positive definite and this decomposition is unique [LT85, §5.7]. We refer then to H_{pol} and Q_{pol} as respectively the symmetric and orthogonal polar factors of A.

To compute the inverse of the exponential map (i.e., the logarithm), we have to solve for ξ the equation $\mathrm{EXP}_{\pi(Y_1)}\xi = \pi(Y_2)$. We first discuss the possible solutions in the case $Y_1^\top Y_2$ nonsingular, and we exhibit the one with the smallest norm. Before that, we introduce the following definition, that will be useful for the rest of the paper.

Definition 4.2. The set $\mathcal{M}_Y$ is defined as:

$$\mathcal{M}_Y := \{\bar{\xi}_Y \in \mathcal{H}_Y : Y^\top \overline{\mathrm{EXP}}_Y(t\bar{\xi}_Y) \in \mathbb{R}_*^{p\times p}\ \forall t \in [0,1]\}.$$

Proposition 4.3. *The following inclusions hold:*

$$\mathcal{M}_Y \subseteq \mathcal{D}_Y \subseteq \mathcal{F}_Y.$$

Proof. Suppose that $\bar{\xi}_Y \in \mathcal{M}_Y$, but $\bar{\xi}_Y \notin \mathcal{D}_Y$. Then, there exists $t \in [0,1]$ such that $\overline{\mathrm{EXP}}_Y(t\bar{\xi}_Y) \notin \mathbb{R}_*^{n\times p}$. This implies that $Y^\top\overline{\mathrm{EXP}}_Y(t\bar{\xi}_Y) \notin \mathbb{R}_*^{p\times p}$, which contradicts the fact that $\bar{\xi}_Y \in \mathcal{M}_Y$. The second inclusion comes from Corollary 3.3. □

Proposition 4.4. *Let $Y_1, Y_2 \in \mathbb{R}_*^{n\times p}$ such that $Y_1^\top Y_2$ is nonsingular. Then, the equation* $\mathrm{EXP}_{\pi(Y_1)}\xi = \pi(Y_2)$ *(i.e., $\pi(Y_1 + \bar{\xi}_{Y_1}) = \pi(Y_2)$) has at least 2^p solutions, given by*

$$\bar{\xi}_{Y_1} = Y_2 Q - Y_1, \qquad \textit{with} \qquad Q = V\mathcal{I}U^\top,$$

with U and V the matrices containing the left and right singular vectors of the product $Y_1^\top Y_2$, and $\mathcal{I}$ any diagonal matrix whose elements belong to the set $\{-1,1\}$. Equivalently,

$$\bar{\xi}_{Y_1} = Y_1(Y_1^\top Y_1)^{-1}(\tilde{H} - Y_1^\top Y_1) + Y_{1\perp}Y_{1\perp}^\top Y_2\tilde{Q}^\top \tag{13}$$

where $\tilde{H} = \tilde{H}^\top$ (not necessarily positive definite) and $\tilde{Q} \in \mathcal{O}_p$ satisfy $Y_1^\top Y_2 = \tilde{H}\tilde{Q}$. The shortest of those vectors is unique and given by

$$\bar{\xi}^*_{Y_1} = Y_2Q^* - Y_1 = Y_1(Y_1^\top Y_1)^{-1}(H_{\mathrm{pol}} - Y_1^\top Y_1) + Y_{1\perp}Y_{1\perp}^\top Y_2 Q_{\mathrm{pol}}^\top, \tag{14}$$

where $Y_1^\top Y_2 = H_{\mathrm{pol}}Q_{\mathrm{pol}}$ is the polar decomposition and $Q^* := Q_{\mathrm{pol}}^\top = VU^\top$. Moreover, $\bar{\xi}^*_{Y_1} \in \mathcal{M}_{Y_1} \subset \mathcal{D}_{Y_1}$, and it is the unique solution in $\mathcal{M}_{Y_1}$.

Proof. The condition $\pi\left(Y_1 + \bar{\xi}_{Y_1}\right) = \pi(Y_2)$ is equivalent to $Y_1 + \bar{\xi}_{Y_1} = Y_2Q$ for some matrix $Q \in \mathcal{O}_p$. The condition $\bar{\xi}_{Y_1} \in \mathcal{H}_{Y_1}$ is then equivalent to:

$$H = H^\top \quad \text{with} \quad H := Y_1^\top \bar{\xi}_{Y_1} = Y_1^\top(Y_2Q - Y_1).$$

Therefore, $Y_1^\top Y_2 = \left(Y_1^\top Y_1 + H\right) Q^\top =: \tilde{H}\tilde{Q}$. The solutions to this equation were obtained in Lemma 4.1. The equivalent expression (13) comes from the fact that $Y_1^\top \bar{\xi}_{Y_1} = H = \tilde{H} - Y_1^\top Y_1$, and $Y_{1\perp}^\top \bar{\xi}_{Y_1} = Y_{1\perp}^\top(Y_2\tilde{Q}^\top - Y_1) = Y_{1\perp}^\top Y_2\tilde{Q}^\top$ (and the fact that $Y_{1\perp}$ is orthonormal).

The minimization of the norm $||Y_2Q - Y_1||_{\mathrm{F}}$ is a particular instance of the orthogonal Procrustes problem, see [GV96, §12.4.1]. The solution is $Q^* = Q_{\mathrm{Pol}}^\top$, with Q_{Pol} the orthogonal polar factor of $Y_1^\top Y_2$. This corresponds to the choice $\mathcal{I} = I$ in Lemma 4.1.

We now show that $\bar{\xi}^*_{Y_1} \in \mathcal{M}_{Y_1}$, i.e., that $Y(t) := Y_1 + t\bar{\xi}^*_{Y_1}$ is such that $Y_1^\top Y(t)$ remains in $\mathbb{R}^{p\times p}_*$ (i.e., is full-rank) for $t \in [0,1]$. Observe that $Y_1^\top Y(t) = Y_1^\top Y_1 + t(H_{\mathrm{pol}} - Y_1^\top Y_1) = (1-t)Y_1^\top Y_1 + tH_{\mathrm{Pol}} \succ 0$ since $Y_1^\top Y_1 \succ 0$ and $H_{\mathrm{Pol}} \succ 0$. Hence $Y_1^\top Y(t) \in \mathbb{R}^{p\times p}_*$.

There remains to prove that $\bar{\xi}^*_{Y_1}$ is the unique solution in $\mathcal{M}_{Y_1}$. Any horizontal vector $\bar{\zeta}_{Y_1} \neq \bar{\xi}_{Y_1}$ satisfying $\pi(Y_1 + \bar{\zeta}_{Y_1}) = \pi(Y_2)$ can be written as $\bar{\zeta}_{Y_1} := Y_2\tilde{Q} - Y_1$, for some orthogonal matrix $\tilde{Q} \neq I$. Due to the uniqueness of the polar decomposition of nonsingular matrices, the product

$$Y_1^\top Y_2\tilde{Q} = Y_1^\top(Y_1 + \bar{\zeta}_{Y_1}),$$

which is symmetric according to (1), is not positive semidefinite. Since $Y_1^\top Y_1 \succ 0$, there must exist $t^* \in [0,1]$ such that $\det\left(Y_1^\top(Y_1 + t^*\bar{\zeta}_{Y_1})\right) = 0$. This implies that $\bar{\zeta}_{Y_1} \notin \mathcal{M}_{Y_1}$. □

Before being able to define the logarithm entirely, we discuss the possible solutions to the equation $\mathrm{EXP}_{\pi(Y_1)}\xi = \pi(Y_2)$ in the case $Y_1^\top Y_2$ singular. We show that the shortest vector satisfying this equation is not unique. As a result, there are several shortest paths going from $\pi(Y_1)$ to $\pi(Y_2)$, and the logarithm is not uniquely defined.

Proposition 4.5. *Let $Y_1, Y_2 \in \mathbb{R}^{n\times p}_*$ such that $Y_1^\top Y_2 = [UU_\perp]\mathrm{Diag}\left(D, 0_{p-r}\right)[VV_\perp]^\top$ is a singular value decomposition (i.e., $Y_1^\top Y_2$ has rank $r < p$). Then, if $r = p-1$, the equation $\mathrm{EXP}_{\pi(Y_1)}\xi = \pi(Y_2)$ (i.e., $\pi(Y_1 + \bar{\xi}_{Y_1}) = \pi(Y_2)$) has at least 2^p solutions, given by*

$$\bar{\xi}_{Y_1} = Y_2Q - Y_1, \qquad \text{with} \qquad Q = \tilde{V}\mathcal{I}\tilde{U}^\top,$$

where $\tilde{U} := [UU_\perp]$, $\tilde{V} := [VV_\perp]$, and $\mathcal{I}$ is any diagonal matrix whose elements belong to the set $\{-1,1\}$. Otherwise, there is an infinity of solutions that can be expressed as

$$\bar{\xi}_{Y_1} = Y_2Q - Y_1, \qquad \text{with} \qquad Q = \tilde{V}\mathrm{Diag}\left(\mathcal{I}, \tilde{Q}\right)\tilde{U}^\top,$$

with $\tilde{Q} \in \mathcal{O}_{p-r}$. In any case, the shortest vectors are of the form:

$$\bar{\xi}^*_{Y_1} = Y_2Q^* - Y_1,$$

where Q^* is the transposed of the orthogonal factor of any polar decomposition $Y_1^\top Y_2 = H_{\mathrm{pol}}Q_{\mathrm{pol}}$. For example, one of the minimizers is obtained for $Q^* := \tilde{V}\tilde{U}^\top$.

Proof. The proof is similar to the one of the previous proposition. Let $\bar{\xi}_{Y_1} \in \mathcal{H}_{Y_1}$. The equality $\mathrm{EXP}_{\pi(Y_1)}\xi = \pi(Y_2)$ is equivalent to:

$$Y_1^\top Y_2 = \left(Y_1^\top Y_1 + H\right) Q^\top,$$

for some matrices $H = H^\top$ and $Q \in \mathcal{O}_p$, and the possible solutions are provided in Lemma 4.1. The fact that, among those solutions, $\bar{\xi}^*_{Y_1}$ is a minimizer of $||\bar{\xi}_{Y_1}||_{\mathrm{F}}$ is again a consequence of [GV96, §12.4.1]. □

We now move to the definition of the logarithm map.

Definition 4.6. For $Y_1, Y_2 \in \mathbb{R}^{n\times p}_*$ such that $Y_1^\top Y_2$ is nonsingular, we introduce the mapping

$$\overline{\mathrm{Log}}_{Y_1} Y_2 := \bar{\xi}^*_{Y_1}$$

where $\bar{\xi}^*_{Y_1}$ is defined as in (14). Let also $\mathrm{Log}_{\pi(Y_1)}\pi(Y_2)$ denote the Riemannian logarithm on $\mathbb{R}^{n\times p}_*/\mathcal{O}_p$, namely the shortest vector $\xi_{\pi(Y_1)}$ such that $\mathrm{Exp}_{\pi(Y_1)}\xi_{\pi(Y_1)} = \pi(Y_2)$, i.e., $\pi(Y_1 + \bar{\xi}_{Y_1}) = \pi(Y_2)$ and $\mathrm{rank}(Y_1 + t\bar{\xi}_{Y_1}) = p$ for all $t \in [0, 1]$.

A consequence of Propositions 4.4 and 4.5 is that the shortest path between $\pi(Y_1)$ and $\pi(Y_2)$ is unique if and only if $Y_1^\top Y_2$ is nonsingular. As a result, the Riemannian logarithm is only uniquely defined for points satisfying this condition. The horizontal lift of the Riemannian logarithm can then be computed as follows.

Theorem 4.7. *Let $Y_1, Y_2 \in \mathbb{R}^{n\times p}_*$ such that $Y_1^\top Y_2$ is nonsingular. Then $\mathrm{Log}_{\pi(Y_1)}\pi(Y_2)$ is uniquely defined and its horizontal lift at Y_1 is given by*

$$\overline{\mathrm{Log}_{\pi(Y_1)}\pi(Y_2)}_{Y_1} = \overline{\mathrm{Log}}_{Y_1} Y_2 = Y_2 Q^* - Y_1, \qquad Q^* := VU^\top,$$

with $Y_1^\top Y_2 =: U\Sigma V^\top$ a singular value decomposition.

Proof. This is direct from Proposition 4.4. □

An immediate consequence is the following corollary, describing minimizing curves on $\mathbb{R}^{n\times p}_*/\mathcal{O}_p$.

Corollary 4.8. *Let $\bar{\xi}_Y \in \mathcal{H}_Y$ and $t^* := \min\left\{t > 0 : t\bar{\xi}_Y \notin \mathcal{M}_Y\right\}$. Then the curve $t \mapsto Y(t) := \pi(\overline{\mathrm{EXP}}_Y t\bar{\xi}_Y) = \pi(Y + t\bar{\xi}_Y)$ is minimizing on $t \in [0, t^*)$. Assuming that $t^*\bar{\xi}_Y \in \mathcal{F}_Y$, there exist several minimizing curves between $\pi(Y)$ and $\pi(\overline{\mathrm{EXP}}_Y t^*\bar{\xi}_Y)$, the curve $t \mapsto Y(t)$ being one of them. Finally, $t \mapsto Y(t)$ is not minimizing beyond t^*.*

Corollary 4.9. *Let $t^* := \min\left\{t > 0 : t\bar{\xi}_Y \notin \mathcal{M}_Y\right\}$ and $t_* := \max\left\{t < 0 : t\bar{\xi}_Y \notin \mathcal{M}_Y\right\}$. Then $\overline{\mathrm{Log}}_Y\left(\overline{\mathrm{EXP}}_Y t\bar{\xi}_Y\right) = t\bar{\xi}_Y$ if and only if $t \in (t_*, t^*)$.*

The choice to define, in these two corollaries, the variable t^* as a minimum or maximum of the set considered (and not the infimum or supremum) relies on the following observation. Based on the definition of $\mathcal{M}_Y$, it is obvious that $t^* := \min\{t > 0 : t\bar{\xi}_Y \notin \mathcal{M}_Y\} = \min\{t > 0 : Y^\top(Y + t\bar{\xi}_Y) \notin \mathbb{R}^{p\times p}_*\}$. The set $\{t > 0 : Y^\top(Y + t\bar{\xi}_Y) \in \mathbb{R}^{p\times p}_*\}$ is open as it is the preimage of the open set $\mathbb{R}^{p\times p}_*$ by the continuous function $t \mapsto Y^\top(Y + t\bar{\xi}_Y)$. Therefore, its complement is closed and the minimum exists with the convention that $\min(\emptyset) = +\infty$. The same argument applies to the case $t_* := \max\{t < 0 : t\bar{\xi}_Y \notin \mathcal{M}_Y\} = \max\{t < 0 : Y^\top(Y + t\bar{\xi}_Y) \notin \mathbb{R}^{p\times p}_*\}$.

5 Riemannian distance

The following result provides the distance between $\pi(Y_1)$ and $\pi(Y_2)$ with respect to the metric (4). Observe that, when $p = n$, the result reduces to the Bures–Wasserstein distance explored in [BJL16].

Proposition 5.1. *Let $Y_1, Y_2 \in \mathbb{R}_*^{n\times p}$ with the singular value decomposition $Y_1^\top Y_2 = U\Sigma V^\top$. The distance between $\pi(Y_1)$ and $\pi(Y_2)$ is:*

$$\mathrm{d}(\pi(Y_1), \pi(Y_2)) = ||Y_2 Q^* - Y_1||_{\mathrm{F}}, \qquad Q^* = VU^\top.$$

Equivalently, on $S_+(p,n)$ endowed with the Riemannian metric g,

$$\mathrm{d}(S_1, S_2) = \left[\mathrm{tr}\,(S_1) + \mathrm{tr}\,(S_2) - 2\mathrm{tr}\left(\left(S_1^{1/2} S_2 S_1^{1/2}\right)^{1/2}\right)\right]^{1/2}. \tag{15}$$

Proof. This is a direct consequence of Propositions 4.4 and 4.5. If $Y_1^\top Y_2$ is invertible, the distance is the norm of the logarithm map. Otherwise, there exists several shortest paths between the two equivalence classes, and Q^* corresponds to one of those paths (it corresponds to the choice $\mathrm{Diag}\left(\mathcal{I}, \tilde{Q}\right) = I$ in Proposition 4.5). □

According to [Gel90], the distance (15) coincides with the Wasserstein distance between the degenerate centered Gaussian distributions parameterized by covariance matrices S_1 and S_2.

6 Injectivity radius

For ξ small enough, the exponential map is a diffeomorphism: it is smooth and has a smooth inverse [O'N83, Chap. 3, Prop. 30]. In this section, we show that the exponential map $\xi_{\pi(Y)} \mapsto \mathrm{Exp}_{\pi(Y)}\xi_{\pi(Y)}$ is a diffeomorphism on $\mathrm{D}\pi(Y)[\mathcal{M}_Y]$. Hence, the injectivity radius of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ at a point $\pi(Y)$ is the radius of the largest ball contained in $\mathcal{M}_Y$ and centered at 0_Y. We first derive two results about $\mathcal{M}_Y$.

Proposition 6.1. *If $p = n$, then $\mathcal{M}_Y = \mathcal{D}_Y$. If $p < n$, then $\mathcal{M}_Y$ is a proper subset of $\mathcal{D}_Y$.*

Proof. If $p = n$, every $\bar{\xi}_Y \in \mathcal{H}_Y$ can be decomposed as $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H$. Then,

$$\begin{aligned}
\bar{\xi}_Y \in \mathcal{D}_Y &\Leftrightarrow \mathrm{rank}\left(\overline{\mathrm{EXP}}_Y t\bar{\xi}_Y\right) = p \quad \forall t \in [0,1] \\
&\Leftrightarrow \mathrm{rank}\left(Y(Y^\top Y)^{-1}(Y^\top Y + tH)\right) = p \quad \forall t \in [0,1] \\
&\Leftrightarrow Y^\top \overline{\mathrm{EXP}} t\bar{\xi}_Y = Y^\top Y + tH \in \mathbb{R}_*^{n\times p} \quad \forall t \in [0,1] \\
&\Leftrightarrow \bar{\xi}_Y \in \mathcal{M}_Y.
\end{aligned}$$

If $p < n$, let $Y = U\Sigma V^\top$ be a singular value decomposition, and let $\bar{\xi}_Y := -\sigma_p^2 Y(Y^\top Y)^{-1} v_p v_p^\top + Y_\perp K$ with $Kv_p \neq 0$. Then, $\bar{\xi}_Y \in \mathcal{D}_Y$ but $\bar{\xi}_Y \notin \mathcal{M}_Y$. □

Proposition 6.2. *The largest ball in $\mathcal{M}_Y$ centered at 0_Y has radius $\sigma_p(Y)$.*

Proof. By Corollary 3.6 and Proposition 4.3, the radius of the largest ball contained in $\mathcal{M}_Y$ is upper bounded by $\sigma_p(Y)$. We show that it is actually equal to $\sigma_p(Y)$. Assume that $\bar{\xi}_Y := Y(Y^\top Y)^{-1}H + Y_\perp K \notin \mathcal{M}_Y$, and define $Y(t) := \overline{\mathrm{EXP}}_Y t\bar{\xi}_Y$. Since $\bar{\xi}_Y \notin \mathcal{M}_Y$, the product $Y^\top Y(t)$ is singular for some $t^* \in [0,1]$. We show that $||t^*\bar{\xi}_Y||_{\mathrm{F}} \geq \sigma_p(Y)$, which implies that $||\bar{\xi}_Y||_{\mathrm{F}} \geq \sigma_p(Y)$. Since $Y^\top Y(t^*)$ is singular, there holds:

$$0 = \det\left(Y^\top(Y + t^* Y(Y^\top Y)^{-1}H)\right) = \det\left(Y^\top Y + t^* H\right). \tag{16}$$

This equation implies that $t^* Y(Y^\top Y)^{-1}H \notin \mathcal{F}_Y$, since $Y + t^* Y(Y^\top Y)^{-1}H = Y(Y^\top Y)^{-1}(Y^\top Y + t^* H)$ is rank deficient. According to Corollary 3.5,

$$||t^*\bar{\xi}_Y||_{\mathrm{F}}^2 = ||t^* Y(Y^\top Y)^{-1}H||_{\mathrm{F}}^2 + ||t^* Y_\perp K||_{\mathrm{F}}^2 \geq \sigma_p^2(Y) + ||t^* Y_\perp K||_{\mathrm{F}}^2 \geq \sigma_p^2(Y).$$

□

We show that the exponential map it is a diffeomorphism on $\mathcal{M}_Y$. Proposition 6.2 implies then that the injectivity radius of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ at $\pi(Y)$ is equal to $\sigma_p(Y)$, the radius of the largest ball contained in $\mathcal{M}_Y$.

Theorem 6.3. *The exponential map is a diffeomorphism on $\mathcal{M}_Y$. As a result, the injectivity radius of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ at $\pi(Y)$ is $\sigma_p(Y)$.*

Proof. Due to Propositions 4.3 and 4.4, $\mathrm{Exp}_{\pi(Y)}$ is a bijection on $\mathcal{M}_Y$. We conclude the proof by observing that the exponential and logarithm maps are both smooth, by smoothness of the polar decomposition [DE99, §2.3(c)]. □

The next result characterizes the global injectivity radius of the manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, defined as the infimum over $\pi(Y) \in \mathbb{R}_*^{n\times p}/\mathcal{O}_p$ of the injectivity radius at $\pi(Y)$.

Corollary 6.4. *The (global) injectivity radius of the manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is equal to 0.*

7 Lie derivative

If there would be a submanifold $\mathcal{S}$ of $\mathbb{R}_*^{n\times p}$ such that $\mathcal{S} \cap Y\mathcal{O}_p$ is a singleton for all $Y \in \mathbb{R}_*^{n\times p}$ and $\mathcal{T}_Y\mathcal{S} = \mathcal{H}_Y$ for all $Y \in \mathcal{S}$, then $(\pi|_{\mathcal{S}})^{-1}$ would be an isometric embedding of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ into the Euclidean space $\mathbb{R}^{n\times p}$, and we might be able to resort to submanifold theory instead of quotient manifold theory to study the Riemannian manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$. We show in this section that such a submanifold $\mathcal{S}$ does not exist when $p > 1$: the condition $\mathcal{T}_Y\mathcal{S} = \mathcal{H}_Y$ cannot be satisfied, even locally.

The Riemannian submersion theory (see, e.g., [O'N83]) enables to write the Lie bracket in the quotient $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ as the horizontal projection of the Lie bracket in $\mathbb{R}_*^{n\times p}$.

Proposition 7.1. *Let ξ, η be two vector fields on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, and let $\bar{\xi}_Y, \bar{\eta}_Y$ be their horizontal lifts at Y. The lift at Y of the Lie bracket $[\xi, \eta]$ is*

$$\overline{[\xi,\eta]}_Y = \mathrm{P}^{\mathrm{h}}[\bar{\xi},\bar{\eta}]_Y = [\bar{\xi},\bar{\eta}]_Y - Y\mathbf{T}^{-1}_{Y^\top Y}\left(2(\bar{\eta}_Y^\top\bar{\xi}_Y - \bar{\xi}_Y^\top\bar{\eta}_Y)\right),$$

where $[\bar{\xi},\bar{\eta}]_Y$ is the Lie bracket in $\mathbb{R}_*^{n\times p}$, evaluated at Y.

Proof. Lemma [O'N83, Chap. 7, Lem. 45] states that $\overline{[\xi,\eta]}_Y = \mathrm{P}^{\mathrm{h}}[\bar{\xi},\bar{\eta}]_Y$. We write $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H_\xi + (I - Y(Y^\top Y)^{-1}Y^\top)A_\xi$, and $\bar{\eta}_Y = Y(Y^\top Y)^{-1}H_\eta + (I - Y(Y^\top Y)^{-1}Y^\top)A_\eta$, where H_ξ, H_η are symmetric $p \times p$ matrices, and A_ξ, A_η are arbitrary $n \times p$ matrices. The Lie bracket $[\bar{\xi},\bar{\eta}]$ is defined as $[\bar{\xi},\bar{\eta}]_Y = D\bar{\eta}_Y[\bar{\xi}_Y] - D\bar{\xi}_Y[\bar{\eta}_Y]$, where $D\bar{\eta}_Y[\bar{\xi}_Y]$ is the directional derivative in the Euclidean space of the vector $\bar{\eta}_Y$ in the direction $\bar{\xi}_Y$, which is given by:

$$\begin{aligned} D\bar{\eta}_Y[\bar{\xi}_Y] &= \bar{\xi}_Y(Y^\top Y)^{-1}[H_\eta - Y^\top A_\eta] - Y(Y^\top Y)^{-1}(\bar{\xi}_Y^\top Y + Y^\top\bar{\xi}_Y)(Y^\top Y)^{-1}[H_\eta - Y^\top A_\eta] \\ &\quad + Y(Y^\top Y)^{-1}[DH_\eta[H_\xi] - \bar{\xi}_Y^\top A_\eta] + (I - Y(Y^\top Y)^{-1}Y^\top)DA_\eta[A_\xi], \end{aligned}$$

and similarly for $D\bar{\xi}_Y[\bar{\eta}_Y]$. According to (3) and (2), the vertical projection of the Lie bracket is $\mathrm{P}^{\mathrm{v}}[\bar{\xi},\bar{\eta}]_Y = Y\mathbf{T}^{-1}_{Y^\top Y}(2\mathrm{skew}(Y^\top[\bar{\xi},\bar{\eta}]_Y))$, while the horizontal projection is $\mathrm{P}^{\mathrm{h}}[\bar{\xi},\bar{\eta}]_Y = [\bar{\xi},\bar{\eta}]_Y - \mathrm{P}^{\mathrm{v}}[\bar{\xi},\bar{\eta}]_Y$. Let us compute $Y^\top D\bar{\eta}_Y[\bar{\xi}_Y]$. We obtain:

$$Y^\top D\bar{\eta}_Y[\bar{\xi}_Y] = H_\xi(Y^\top Y)^{-1}[H_\eta - Y^\top A_\eta] - 2H_\xi(Y^\top Y)^{-1}[H_\eta - Y^\top A_\eta] + DH_\eta[H_\xi] - \bar{\xi}_Y^\top A_\eta.$$

Replacing $\bar{\xi}_Y$ by its definition in the last term of the previous equation yields:

$$Y^\top D\bar{\eta}_Y[\bar{\xi}_Y] = -H_\xi(Y^\top Y)^{-1}H_\eta + DH_\eta[H_\xi] - A_\xi^\top(I - Y(Y^\top Y)^{-1}Y^\top)A_\eta.$$

Again, the expression for $Y^\top D\bar{\xi}_Y[\bar{\eta}_Y]$ is obtained by just switching $\bar{\xi}_Y$ and $\bar{\eta}_Y$ in the previous equation. Putting everything together, we obtain, for the product $Y^\top[\bar{\xi},\bar{\eta}]_Y$:

$$Y^\top[\bar{\xi},\bar{\eta}]_Y = \bar{\eta}_Y^\top\bar{\xi}_Y - \bar{\xi}_Y^\top\bar{\eta}_Y + DH_\eta[H_\xi] - DH_\xi[H_\eta].$$

So, the projection of $[\bar{\xi},\bar{\eta}]_Y$ on the vertical space is $\mathrm{P}^{\mathrm{v}}[\bar{\xi},\bar{\eta}]_Y = Y\mathbf{T}^{-1}_{Y^\top Y}\left(2(\bar{\eta}_Y^\top\bar{\xi}_Y - \bar{\xi}_Y^\top\bar{\eta}_Y)\right)$, and $\mathrm{P}^{\mathrm{h}}[\bar{\xi},\bar{\eta}]_Y = [\bar{\xi},\bar{\eta}]_Y - \mathrm{P}^{\mathrm{v}}[\bar{\xi},\bar{\eta}]_Y$. □

Proposition 7.2. *If $p > 1$, the horizontal distribution (i.e., the set of all horizontal vectors to $\mathbb{R}_*^{n\times p}$) is not involutive.*

Proof. This is a direct consequence of the previous result, in which the vertical projection of the Lie bracket $[\bar{\xi}, \bar{\eta}]_Y$ is seen to be generally non-zero provided that $p > 1$. □

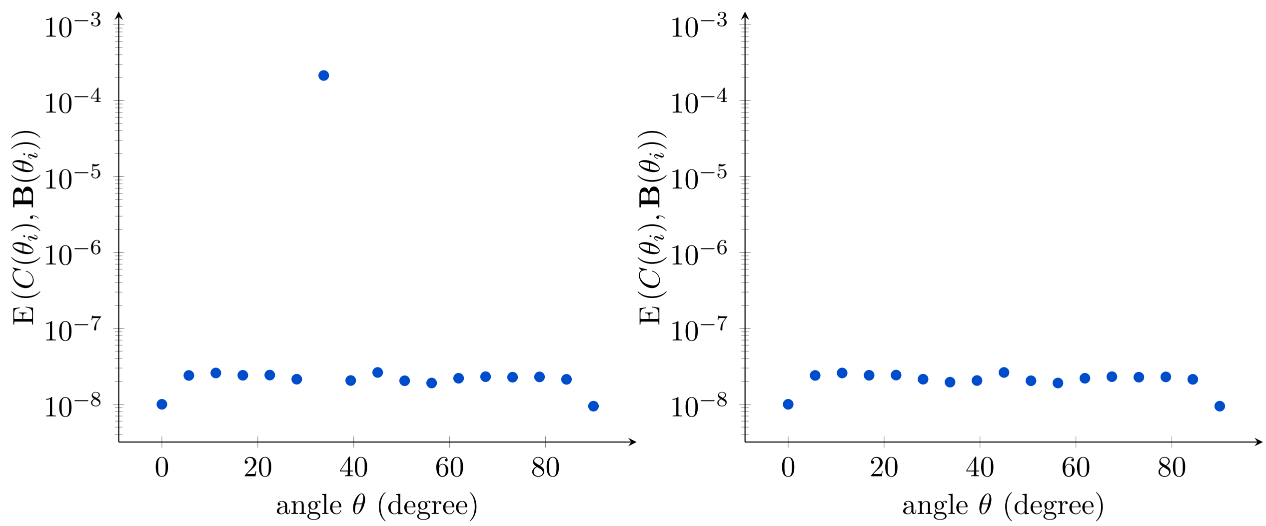
By the Frobenius theorem [AMR88, §4.4.3], there exists no integral manifold for the horizontal distribution if $p > 1$.

8 Numerical illustrations

Section 6 provides an expression for the injectivity radius of the manifold. It is equal to the smallest singular value of the matrix Y: it is close to zero when the matrix Y is close to the boundary of $\mathbb{R}_*^{n\times p}$. This can have adverse effects in some practical applications. We provide in this section an example, that we have obtained from an application in wind field estimation, and already presented in [GMM⁺17, GMA18].

In [GMA18, §4.3], the authors apply piecewise Bézier interpolation to the covariance matrices characterizing the wind field. The wind field is represented by a Gaussian stochastic process, characterized by a mean field and a covariance matrix. Those two parameters depend on some external conditions, such as the prevailing wind in the region of interest. They were computed for several prevailing wind magnitudes, by running computationally expensive computational fluid dynamics (CFD) models. To avoid running too many CFD models, the authors fit a curve to the known covariance matrices, to recover matrices associated with other (intermediary) wind magnitudes.

The authors considered a set of covariance matrices $C(\theta_1), C(\theta_3), C(\theta_5), \ldots, C(\theta_{33})$. The matrices are of size 3024×3024, and the rank was estimated to be 20. They observed that the algorithm fails to interpolate one of the data points, as illustrated on Figure 2a. The lack in interpolation is a consequence of the fact that the points are too far away with respect to the injectivity radius, resulting in a discontinuity in the logarithm during one of the steps of the algorithm. Adapting the rank of the data (i.e., keeping a reduced number of columns) solves this problem, as illustrated on Figure 2b. The interested reader is referred to [GMA18] for more information.



(a) $p = 20$ **(b)** $p = 18$

Figure 2: Interpolation error between the curve $\mathcal{B}(\theta)$ and the data points $C(\theta_1), C(\theta_3), C(\theta_5), \ldots, C(\theta_{33})$. The figure on the left was already presented in [GMA18, Fig.4]. We observe that the error made on the data point $C_{\theta_{13}}$ is several orders of magnitude higher than the error made on the others. This is no longer the case when the parameter p, corresponding to the rank, is reduced, as indicated by the figure on the right.

9 Conclusion

It is well known that the set $\mathcal{S}_+(p,n)$ of $n \times n$ symmetric positive-semidefinite matrices of rank p is an embedded submanifold of $\mathbb{R}^{n\times n}$. We have shown that this manifold is diffeomorphic (i.e., equivalent) to the quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, the diffeomorphism being $\Phi : \mathbb{R}_*^{n\times p}/\mathcal{O}_p \to \mathcal{S}_+(p,n) : Y\mathcal{O}_p \mapsto YY^\top$.

The Riemannian submersion theory yields a natural Riemannian metric g on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$. It turns out that this Riemannian metric g is not equivalent to the Riemannian submanifold metric of $\mathcal{S}_+(p,n)$. In particular, whereas the geodesics of $\mathcal{S}_+(p,n) \subset \mathbb{R}^{n\times n}$ do not generally have a known analytical expression [VAV09], the geodesics of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ admit a particularly simple expression: mapped on $\mathcal{S}_+(p,n)$ through Φ, they read

$$t \mapsto (Y + t\bar{\xi}_Y)(Y + t\bar{\xi}_Y)^\top,$$

where $Y \in \mathbb{R}_*^{n\times p}$ and $\bar{\xi}_Y \in \mathbb{R}^{n\times p}$ with $Y^\top\bar{\xi}_Y$ symmetric.

As a consequence, steepest descent along geodesics for an optimization problem of the form $\min_{S\in\mathcal{S}_+(p,n)} f(S)$ in this quotient geometry boils down to the Euclidean steepest descent for the optimization problem $\min_Y \bar{f}(Y)$, where $\bar{f}(Y) := f(YY^\top)$ (assuming that we are never unlucky enough to go through a rank-deficient Y). This makes this geometry very simple for optimization purposes, while it is considerably harder to do steepest descent along geodesics for $\min_{S\in\mathcal{S}_+(p,n)} f(S)$ in the above-mentioned submanifold geometry.

However, observe that things become a bit less straightforward when using the Hessian, or if the objective function involves distances, or in computational problems where the logarithm is needed.

We have given in Theorem 4.7 a formula for the endpoint geodesic problem (i.e., the Riemannian logarithm), and we have characterized the minimizing geodesics in Corollary 4.8.

Moreover, we have shown that the injectivity radius of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ at $Y\mathcal{O}_p$ is $\sigma_p(Y)$, the smallest singular value of Y. A numerical example has illustrated that certain curve-fitting methods on $\mathcal{S}_+(p,n)$ may produce unsatisfactory results when data points have a small $\sigma_p(Y)$. We suspect that resorting to other geometries—such as the submanifold geometry [VAV09] or the geometry with complete geodesics proposed in [VAV13]—would not remedy the issue, and moreover those other geometries are impractical for this curve-fitting task because no formula is available for their Riemannian logarithm.

To conclude, expressions for the main geometric tools on $\mathcal{S}_+(p,n) \simeq \mathbb{R}_*^{n\times p}/\mathcal{O}_p$ are summarized in Table 1.

Equivalence classes	$Y\mathcal{O}_p := \{YQ : Q \in \mathcal{O}_p\}$
Dimension	$pn - \frac{p(p-1)}{2}$
Vertical space	$\mathcal{V}_Y = \{Y\Omega : \Omega = -\Omega^\top \in \mathbb{R}^{p\times p}\}$
Horizontal space	$\mathcal{H}_Y = \{Z \in \mathbb{R}^{n\times p} : Y^\top Z - Z^\top Y = 0\}$
Metric on $\mathcal{T}_{\pi(Y)}\mathbb{R}_*^{n\times p}/\mathcal{O}_p$	$g_{\pi(Y)}(\xi_{\pi(Y)}, \eta_{\pi(Y)}) = \mathrm{tr}\left(\bar{\xi}_Y^\top \bar{\eta}_Y\right)$
Projection on $\mathcal{V}_Y$	$\mathrm{P^v}_Y(Z) = Y\mathbf{T}^{-1}_{Y^\top Y}(Y^\top Z - Z^\top Y)$
Projection on $\mathcal{H}_Y$	$\mathrm{P^h}_Y(Z) = Z - \mathrm{P^v}_Y(Z)$
Gradient	$\overline{\mathrm{grad} f}_Y = \mathrm{grad}\overline{f}(Y)$, with $\overline{f} := f \circ \pi$
Connection	$\overline{\nabla_\eta \xi}_Y = \mathrm{P^h}_Y(\mathrm{D}\bar{\xi}(Y)[\bar{\eta}_Y])$
Exponential	$\mathrm{Exp}_{\pi(Y)}\xi_{\pi(Y)} = \pi(Y + \bar{\xi}_Y)\ \forall \xi_Y \in \mathcal{D}_Y$
Logarithm	If $\det\left(Y_1^\top Y_2\right) \neq 0$, $\overline{\mathrm{Log}_{\pi(Y_1)}\pi(Y_2)}_{Y_1} = Y_2 Q^* - Y_1$, with $Y_1^\top Y_2 = HQ^{*\top}$ the polar decomposition of $Y_1^\top Y_2$.
Riemannian distance on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$	$\delta(\pi(Y_1), \pi(Y_2)) = \|\|Y_2 Q^* - Y_1\|\|_\mathrm{F}$, Q^* defined as above
Riemannian distance on $\mathcal{S}_+(p, n)$	$\delta(S_1, S_2) = \left[\mathrm{tr}\,(S_1) + \mathrm{tr}\,(S_2) - 2\mathrm{tr}\left(\left(S_1^{1/2} S_2 S_1^{1/2}\right)^{1/2}\right)\right]^{1/2}$
Injectivity radius	$r_{\mathrm{Inj}}(\pi(Y)) = \sigma_p(Y)$
Lie derivative	$\overline{[\xi, \eta]}_Y = [\bar{\xi}, \bar{\eta}]_Y - Y\mathbf{T}^{-1}_{Y^\top Y}\left(2(\bar{\eta}_Y^\top \bar{\xi}_Y - \bar{\xi}_Y^\top \bar{\eta}_Y)\right)$
Vector transport	$\overline{T_{\eta_{\pi(Y)}}\xi_{\pi(Y)}}_{Y+\bar{\eta}_Y} = \mathrm{P^h}_{Y+\bar{\eta}_Y}\bar{\xi}_Y$
Hessian	$\overline{\mathrm{Hess} f(\pi(Y))[\xi_{\pi(Y)}]}_Y = \mathrm{P^h}_Y\left(\mathrm{Dgrad}\bar{f}(Y)[\bar{\xi}_Y]\right)$, $\bar{f} := f \circ \pi$
Riemannan curvature tensor	See (36)
Sectional curvature	$K_{\mathbb{R}_*^{n\times p}/\mathcal{O}_p}(\xi_{\pi(Y)}, \eta_{\pi(Y)}) = \frac{3\|\|Y\mathbf{T}^{-1}_{Y^\top Y}\left(\bar{\eta}_Y^\top \bar{\xi}_Y - \bar{\xi}_Y^\top \bar{\eta}_Y\right)\|\|_\mathrm{F}^2}{\langle \bar{\xi}_Y, \bar{\xi}_Y\rangle \langle \bar{\eta}_Y, \bar{\eta}_Y\rangle - \langle \bar{\xi}_Y, \bar{\eta}_Y\rangle^2}$.
Bounds on the sectional curvature	For any tangent plane (spanned by $\xi_{\pi(Y)}, \eta_{\pi(Y)}$) based at Y, there holds $K_{\mathbb{R}_*^{n\times p}/\mathcal{O}_p}(\xi_{\pi(Y)}, \eta_{\pi(Y)}) \in [0, \frac{3}{\sigma_{p-1}^2(Y) + \sigma_p^2(Y)}]$

Table 1: Summary of the expressions for the main geometric tools on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$.

A Review of the geometry of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$

The results in this section either come from the literature (mainly [JBAS10]), or can be quite easily deduced from it. We provide them in detail for reference purpose.

A.1 The quotient set $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$

We first prove the following result, referred to as Proposition 2.1 in Section 2.

Proposition A.1. *Let $Y_1, Y_2 \in \mathbb{R}_*^{n\times p}$. Then $Y_1Y_1^\top = Y_2Y_2^\top$ if and only if $Y_2 = Y_1Q$ for some $Q \in \mathcal{O}_p$, and*

$$\mathcal{S}_+(p,n) = \{YY^\top : Y \in \mathbb{R}_*^{n\times p}\}.$$

Proof. The first claim is a particularization of [BM05, Lem. 2.1]. We sketch the proof for the reader's convenience. The "if" part is straightforward. For the "only if" part, observe that $Y_1Y_1^\top = Y_2Y_2^\top$, $Y_1, Y_2 \in \mathbb{R}_*^{n\times p}$ implies $\ker(Y_1^\top) = \ker(Y_2^\top)$ and thus $\mathrm{range}(Y_1) = \mathrm{range}(Y_2)$, hence there is $Q \in \mathbb{R}^{p\times p}$ such that $Y_2 = Y_1Q$. Since $Y_1Y_1^\top = Y_1QQ^\top Y_1^\top$ and Y_1 has full column rank, it follows that $QQ^\top = I$, i.e., $Q \in \mathcal{O}_p$. For the second part of the proposition, see [VAV13, Prop. 3.1]. □

Proposition 2.1 yields an identification of $\mathcal{S}_+(p,n)$ with a quotient set. It consists of considering as a single *point* the set of all Y's that yield the same S. The set of all those *points* is the quotient set.

To make this identification precise, let ϕ denote the mapping

$$\phi : \mathbb{R}_*^{n\times p} \to \mathcal{S}_+(p,n) : Y \mapsto YY^\top.$$

Let $\sim$ denote the equivalence relation on $\mathbb{R}_*^{n\times p}$ defined by

$$Y_1 \sim Y_2 \text{ if and only if } Y_2 = Y_1Q \text{ for some } Q \in \mathcal{O}_p, \tag{17}$$

for which the equivalence class of $Y \in \mathbb{R}_*^{n\times p}$ is

$$Y\mathcal{O}_p := \{YQ : Q \in \mathcal{O}_p\}.$$

Let

$$\mathbb{R}_*^{n\times p}/\mathcal{O}_p \;:=\; \mathbb{R}_*^{n\times p}/\sim \;:=\; \{Y\mathcal{O}_p : Y \in \mathbb{R}_*^{n\times p}\} \tag{18}$$

denote the quotient of $\mathbb{R}_*^{n\times p}$ by the equivalence relation $\sim$, and let

$$\pi : \mathbb{R}_*^{n\times p} \to \mathbb{R}_*^{n\times p}/\mathcal{O}_p$$

denote the quotient map. The next result, illustrated in Figure 1, shows that there is a natural bijection between the quotient set $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ and $\mathcal{S}_+(p,n)$.

Corollary A.2. *The map $\phi : \mathbb{R}_*^{n\times p} \to \mathcal{S}_+(p,n) : Y \mapsto YY^\top$ is surjective (i.e., onto). Its fiber through $Y \in \mathbb{R}_*^{n\times p}$ is given by*

$$\phi^{-1}(YY^\top) = Y\mathcal{O}_p := \{YQ : Q \in \mathcal{O}_p\}.$$

Proof. This follows from Proposition A.1

□

The function

$$\Phi : \mathbb{R}_*^{n\times p}/\mathcal{O}_p \to \mathcal{S}_+(p,n) : Y\mathcal{O}_p \mapsto YY^\top$$

defined by $\phi = \Phi \circ \pi$ is then a bijection.

A.2 The quotient space $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$

The quotient set $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ can be endowed with a quotient set topology, which turns it into a quotient space. In this topology, the open sets in $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ are the sets whose inverse image by π are open sets in $\mathbb{R}_*^{n\times p}$ (endowed here with the metric topology), see [BC70]. It is then clear that π is a continuous function with respect to this topology.

We will show that the quotient space $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is Hausdorff and second-countable. To prove these two properties, we consider the action of the orthogonal group $\mathcal{O}_p$ on $\mathbb{R}_*^{n\times p}$. Given a point $Y \in \mathbb{R}_*^{n\times p}$, let

$$Y\mathcal{O}_p := \{YQ : Q \in \mathcal{O}_p\}$$

be the orbit of Y under the (right) action of the orthogonal group. The set of orbits can be identified, by (18), to the quotient set $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$. The action of the orthogonal group on $\mathbb{R}_*^{n\times p}$ has the following properties.

Proposition A.3. *The action of the orthogonal group on $\mathbb{R}_*^{n\times p}$ is continuous, smooth, free and proper.*

Proof. Continuity and smoothness are a direct consequence of continuity and smoothness of the matrix product. The action is free since, for any $Y \in \mathbb{R}_*^{n\times p}$, the equality $YQ_1 = YQ_2$ implies $Q_1 = Q_2$. Finally, the action is proper, as the orthogonal group is a compact Lie group (see [DK12, §1.2.A]), and any continuous action by a compact Lie group on a manifold is proper (see [Lee13, Cor. 21.6]). Note that the set $\mathbb{R}_*^{n\times p}$ is a manifold since it is an open subset of the linear space $\mathbb{R}^{n\times p}$. As such, it admits a natural structure of open submanifold of $\mathbb{R}^{n\times p}$. We refer, e.g., to [AMS08] for details. □

We then deduce the following results.

Corollary A.4. *The orbit space $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is Hausdorff.*

Proof. This is a consequence of the fact that the Lie group $\mathcal{O}_p$ acts continuously and properly on the manifold $\mathbb{R}_*^{n\times p}$ (see [Lee13, Prop. 21.4]). □

The proof that $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is second-countable is delayed to the next section.

A.3 The quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$

Since $\mathbb{R}_*^{n\times p}$ is a manifold, it makes sense to ask if the quotient $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is a manifold. (In other words, we wonder if there is a differentiable structure on the set $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ such that the differential of π at every $Y \in \mathbb{R}_*^{n\times p}$ is onto.) The next result answers the question positively (this result was already obtained in [AIDV08]).

Proposition A.5. *The equivalence relation $\sim$ (17) is regular. In other words, the quotient space $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is a quotient manifold. The dimension of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is $pn - \frac{p(p-1)}{2}$.*

Proof. The result follows from the quotient manifold theorem (see [Lee13, Thm. 21.10)]). This theorem states that, given a Lie group G acting smoothly, freely and properly on a smooth manifold $\mathcal{M}$, the orbit space $\mathcal{M}/G$ is a topological manifold of dimension equal to $\dim(\mathcal{M}) - \dim(G)$, and has a unique smooth structure with the property that the quotient map $\pi : \mathcal{M} \to \mathcal{M}/G$ is a smooth submersion. The proof follows then from Proposition A.3. □

As a consequence, the equivalence classes $Y\mathcal{O}_p$ are embedded submanifolds of $\mathbb{R}_*^{n\times p}$. The tangent space to $Y\mathcal{O}_p$ at Y is the *vertical space*:

$$\mathcal{V}_Y = Y\mathcal{S}_{\mathrm{skew}}(p) = \{Y\Omega : \Omega = -\Omega^\top \in \mathbb{R}^{p\times p}\}. \tag{19}$$

We also deduce the following results.

Corollary A.6. *The quotient space $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is second-countable.*

Proof. It is a quotient manifold of a second-countable manifold $\mathbb{R}_*^{n\times p}$, which, by [BC70, Prop.6.3.2], implies that it is itself second-countable. The manifold $\mathbb{R}_*^{n\times p}$ is second-countable as it is an open subset of the Euclidean space $\mathbb{R}^{n\times p}$, which is second-countable (see [Lee11, p. 37-38]). $\square$

Proposition A.7. *The mapping Φ is a diffeomorphism between the quotient manifold $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ and $\mathcal{S}_+(p,n)$.*

Proof. This is a consequence of the fact that both π and ϕ are submersions. Given a submersion f of a manifold $\mathcal{M}$ to a manifold $\mathcal{M}'$, Proposition 6.1.2 of [BC70] states that any function g, such that $g\circ f$ is differentiable, is also differentiable (with respect to the differentiable structure inherited from its embedding in the Euclidean space $\mathbb{R}^{n\times n}$). Using the differentiability of $\phi = \Phi\circ\pi$ and of $\pi = \Phi^{-1}\circ\phi$, we deduce respectively that Φ and Φ^{-1} are differentiable. The fact that π is a submersion is stated in the proof of Proposition A.5, so we just have to prove that $\phi : Y \mapsto YY^\top$ is a submersion. The tangent space of $\mathcal{S}_+(p,n)$ was shown in [VAV09] to be

$$\mathcal{T}_{\phi(Y)}\mathcal{S}_+(p,n) := \{(YH + Y_\perp K)Y^\top + Y(YH+Y_\perp K)^\top : H = H^\top \in \mathbb{R}^{p\times p}, K \in \mathbb{R}^{(n-p)\times p}\}, \tag{20}$$

where $Y_\perp \in \mathbb{R}^{n\times(n-p)}$ is orthonormal and satisfies $Y^\top Y_\perp = 0$. For any $\xi \in \mathcal{T}_{\phi(Y)}\mathcal{S}_+(p,n)$ there exists $\dot{Y} \in \mathcal{T}_Y\mathbb{R}_*^{n\times p} \simeq \mathbb{R}^{n\times p}$ such that $\mathrm{D}\phi(Y)[\dot{Y}] = \dot{Y}Y^\top + Y\dot{Y}^\top = \xi$ (just choose $\dot{Y} := (YH + Y_\perp K)$ above), which implies that ϕ is a submersion. $\square$

The tangent space can also be written more compactly as

$$\mathcal{T}_{\phi(Y)}\mathcal{S}_+(p,n) := \{\begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} H & K^\top \\ K & 0 \end{bmatrix} \begin{bmatrix} Y & Y_\perp \end{bmatrix}^\top : H = H^\top \in \mathbb{R}^{p\times p}, K \in \mathbb{R}^{(n-p)\times p}\} \tag{21}$$

with $Y_\perp \in \mathbb{R}^{n\times(n-p)}$ an orthonormal matrix satisfying $Y^\top Y_\perp = 0$.

Proposition A.7 shows that the manifolds $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ and $\mathcal{S}_+(p,n)$ are *equivalent*:

$$\mathbb{R}_*^{n\times p}/\mathcal{O}_p \simeq \mathcal{S}_+(p,n).$$

Through this equivalence, a point $\pi(Y) = Y\mathcal{O}_p \in \mathbb{R}_*^{n\times p}/\mathcal{O}_p$ corresponds to the point $\phi(Y) = YY^\top \in \mathcal{S}_+(p,n)$. In the rest of this paper, most results are written in terms of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, but they can readily be translated into results for $\mathcal{S}_+(p,n)$; to this end, replace π by ϕ.

A.4 Riemannian metric and horizontal space

A Riemannian metric is an inner product on the tangent spaces that varies smoothly with the foot of the tangent space. We refer, e.g., to [Boo86, AMS08] for more information.

Consider on $\mathbb{R}_*^{n\times p}$ the canonical Riemannian metric defined by

$$\langle Z_1, Z_2\rangle_Y = \operatorname{tr}\left(Z_1^\top Z_2\right), \quad Z_1, Z_2 \in T_Y\mathbb{R}_*^{n\times p}, \tag{22}$$

where $\operatorname{tr}()$ denotes the sum of the diagonal elements of its argument and $T_Y\mathbb{R}_*^{n\times p}$ is the tangent space of $\mathbb{R}_*^{n\times p}$ at Y. The latter can be identified with $\mathbb{R}^{n\times p}$ since $\mathbb{R}_*^{n\times p}$ is an open submanifold of $\mathbb{R}^{n\times p}$. Note that, letting $||Z||_\mathrm{F}$ denote the Frobenius norm of Z, we have $||Z||_\mathrm{F}^2 = \langle Z, Z\rangle$.

For every $Y \in \mathbb{R}_*^{n\times p}$, define the horizontal space $\mathcal{H}_Y$ at Y as the orthogonal complement of the vertical space $\mathcal{V}_Y$, i.e.,

$$\begin{aligned}\mathcal{H}_Y &= \{Z \in \mathbb{R}^{n\times p} : \operatorname{tr}\left((Y\Omega)^\top Z\right) = 0, \text{ for all } \Omega = -\Omega^\top \in \mathbb{R}^{p\times p}\} \\ &= \{Z \in \mathbb{R}^{n\times p} : Y^\top Z - Z^\top Y = 0\},\end{aligned} \tag{23}$$

The second equality comes because the relation $\operatorname{tr}(\Omega S) = 0$ for all $\Omega \in \mathcal{S}_{\text{skew}}(p)$ implies that S is symmetric. The dimension of $\mathcal{H}_Y$ is the dimension of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, i.e., $pn - \frac{p(p-1)}{2}$. We have the alternative description

$$\mathcal{H}_Y = \{Y(Y^\top Y)^{-1}H + Y_\perp K : H = H^\top \in \mathbb{R}^{p\times p}, K \in \mathbb{R}^{(n-p)\times p}\},$$

where $Y_\perp \in \mathbb{R}^{n\times(n-p)}$ is orthonormal and satisfies $Y^\top Y_\perp = 0$. Indeed, it is readily checked that this subspace is orthogonal to the vertical space and has the correct dimension.

For every vector field ξ on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, there exists one and only one vector field $\overline{\xi}$ on $\mathbb{R}_*^{n\times p}$ such that, for all $Y \in \mathbb{R}_*^{n\times p}$,

$$\overline{\xi}_Y \in \mathcal{H}_Y, \tag{24}$$

and

$$\mathrm{D}\pi(Y)[\overline{\xi}_Y] = \xi_{\pi(Y)}. \tag{25}$$

Condition (24) expresses that $\overline{\xi}$ is a *horizontal vector field.* The vector field $\overline{\xi}$ is called the *horizontal lift* of ξ, see [KN63, Chap. II, Prop. 1.2].

Proposition A.8. *If $\xi_{\pi(Y)} \in \mathcal{T}_{\pi(Y)}\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, then its horizontal lift satisfies $\bar{\xi}_{YQ} = \bar{\xi}_Y Q$ for all $Q \in \mathcal{O}_p$. Hence, if $\bar{\xi}_Y \in \mathcal{H}_Y$, then, for all $Q \in \mathcal{O}_p$, $\bar{\xi}_Y Q \in \mathcal{H}_{YQ}$ is the horizontal lift at YQ of $\mathrm{D}\pi(Y)[\bar{\xi}_Y]$.*

Proof. Let $\bar{\xi}_Y$ be the horizontal lift of $\xi_{\pi(Y)}$ at Y. We need to show that $\bar{\xi}_Y Q \in \mathcal{H}_{YQ}$ and $\mathrm{D}\pi(YQ)[\bar{\xi}_Y Q] = \xi_{\pi(Y)}$, and the proof is complete. For the former, since $Y^\top\bar{\xi}_Y$ is symmetric, we have that $(YQ)^\top\bar{\xi}_Y Q = Q^\top Y^\top\bar{\xi}_Y Q$ is symmetric, hence $\bar{\xi}_Y Q \in \mathcal{H}_{YQ}$. For the latter, observe that

$$\begin{aligned} D\phi(YQ)[\bar{\xi}_Y Q] &= \bar{\xi}_Y Q(YQ)^\top + YQ(\bar{\xi}_Y Q)^\top \\ &= \bar{\xi}_Y Y^\top + Y\bar{\xi}_Y^\top \\ &= \mathrm{D}\phi(Y)[\bar{\xi}_Y], \end{aligned}$$

hence $\mathrm{D}\pi(YQ)[\bar{\xi}_Y Q] = \mathrm{D}\pi(Y)[\bar{\xi}_Y] = \xi_{\pi(Y)}$. $\square$

Proposition A.9. *The relation*

$$g_{\pi(Y)}(\xi_{\pi(Y)}, \zeta_{\pi(Y)}) = \mathrm{tr}\left(\overline{\xi}_Y{}^\top\overline{\zeta}_Y\right) \tag{26}$$

defines a Riemannian metric g on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$. The metric g turns the quotient map $\pi : \mathbb{R}_*^{n\times p} \to \mathbb{R}_*^{n\times p}/\mathcal{O}_p$ into a Riemannian submersion. In other words, $(\mathbb{R}_*^{n\times p}/\mathcal{O}_p, g)$ is a Riemannian quotient manifold of $(\mathbb{R}_*^{n\times p}, \langle\cdot,\cdot\rangle)$.

Proof. We just need to show that (26) makes sense, i.e., the right-hand side depends on Y only through $\pi(Y)$. This holds, because if $\pi(Y_1) = \pi(Y_2)$, then $Y_2 = Y_1 Q$ for some $Q \in \mathcal{O}_p$, and thus $\mathrm{tr}\left(\overline{\xi}_{Y_2}{}^\top\overline{\zeta}_{Y_2}\right) = \mathrm{tr}\left(\overline{\xi}_{Y_1Q}{}^\top\overline{\zeta}_{Y_1Q}\right) = \mathrm{tr}\left(Q^\top\overline{\xi}_{Y_1}{}^\top\overline{\zeta}_{Y_1}Q\right) = \mathrm{tr}\left(\overline{\xi}_{Y_1}{}^\top\overline{\zeta}_{Y_1}\right)$. $\square$

Since the mapping $\Phi : \mathbb{R}_*^{n\times p}/\mathcal{O}_p \to \mathcal{S}_+(p,n)$ is a diffeomorphism, the following relation holds between the horizontal space at Y and the tangent space of $\mathcal{S}_+(p,n)$ (seen as an embedded submanifold of $\mathbb{R}^{n\times n}$):

$$\mathrm{D}\phi(Y)[\mathcal{H}_Y] = \mathcal{T}_{\phi(Y)}\mathcal{S}_+(p,n). \tag{27}$$

We will show in Proposition A.11 that we recover the description (21) of $\mathcal{T}_{\phi(Y)}\mathcal{S}_+(p,n)$. To this end, we introduce for every $E \in \mathbb{R}^{p\times p}$ the linear operator

$$\mathbf{T}_E : \mathbb{R}^{p\times p} \to \mathbb{R}^{p\times p} : X \mapsto EX + XE. \tag{28}$$

Lemma A.10. *If E is such that $\Lambda(E) \cap \Lambda(-E) = \emptyset$, i.e., E has no pair of opposite eigenvalues, then the operator $\mathbf{T}_E$ is invertible. In particular, if E is symmetric positive-definite, then $\mathbf{T}_E$ is invertible, and moreover $\mathbf{T}_E(\mathcal{S}_{\mathrm{skew}}(p)) = \mathcal{S}_{\mathrm{skew}}(p)$.*

Proof. The proof follows from [BR97, §10]. $\square$

As mentioned above, we now recover (27) using only tools from matrix theory.

Proposition A.11.

$$\mathrm{D}\phi(Y)[\mathcal{H}_Y] = \{\begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} \tilde{H} & K^\top \\ K & 0 \end{bmatrix} \begin{bmatrix} Y & Y_\perp \end{bmatrix}^\top : \tilde{H} = \tilde{H}^\top \in \mathbb{R}^{p\times p}, K \in \mathbb{R}^{(n-p)\times p}\}. \tag{29}$$

Proof. Let $\bar{\xi}_Y \in \mathcal{H}_Y$, decomposed as $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H + Y_\perp K$. There holds

$$\begin{aligned}
\mathrm{D}\phi(Y)[\bar{\xi}_Y] &= \bar{\xi}_Y Y^\top + Y\bar{\xi}_Y^\top, \\
&= Y\left((Y^\top Y)^{-1}H + H(Y^\top Y)^{-1}\right)Y^\top + Y_\perp K Y^\top + Y K^\top Y_\perp^\top, \\
&= \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} \tilde{H} & K^\top \\ K & 0 \end{bmatrix} \begin{bmatrix} Y & Y_\perp \end{bmatrix}^\top,
\end{aligned}$$

with $\tilde{H} := (Y^\top Y)^{-1}H + H(Y^\top Y)^{-1}$.

Conversely, let

$$\dot{S} := \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} \tilde{H} & K^\top \\ K & 0 \end{bmatrix} \begin{bmatrix} Y & Y_\perp \end{bmatrix}^\top \quad \text{with} \quad \tilde{H} = \tilde{H}^\top \in \mathbb{R}^{p\times p}, K \in \mathbb{R}^{(n-p)\times p}.$$

According to Lemma A.10, there exists one (unique) matrix H such that $\dot{S} = Y\left(H(Y^\top Y)^{-1} + (Y^\top Y)^{-1}H\right)Y^\top + YK^\top Y_\perp^\top + Y_\perp K Y^\top$. As a result, there exists one (unique) vector $\bar{\xi}_Y = Y(Y^\top Y)^{-1}H + Y_\perp K \in \mathcal{H}_Y$ such that $\mathrm{D}\phi(Y)[\bar{\xi}_Y] = \dot{S}$, which concludes the proof. □

The Riemannian metric on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ induces through the diffeomorphism Φ a metric on $\mathcal{S}_+(p,n)$, which we also denote by g in a mild abuse of notation:

$$g_{\Phi(\pi(Y))}\left(\mathrm{D}\Phi(\pi(Y))[\xi_{\pi(Y)}], \mathrm{D}\Phi(\pi(Y))[\eta_{\pi(Y)}]\right) := g_{\pi(Y)}\left(\xi_{\pi(Y)}, \eta_{\pi(Y)}\right).$$

Applying the chain rule, we observe that

$$\begin{aligned}
g_{\phi(Y)}\left(\mathrm{D}\phi(Y)[\bar{\xi}_Y], \mathrm{D}\phi(Y)[\bar{\eta}_Y]\right) &= g_{\Phi(\pi(Y))}\left(\mathrm{D}\Phi(\pi(Y))[\mathrm{D}\pi(Y)[\bar{\xi}_Y]], \mathrm{D}\Phi(\pi(Y))[\mathrm{D}\pi(Y)[\bar{\eta}_Y]]\right) \\
&= g_{\Phi(\pi(Y))}\left(\mathrm{D}\Phi(\pi(Y))[\xi_{\pi(Y)}], \mathrm{D}\Phi(\pi(Y))[\eta_{\pi(Y)}]\right)
\end{aligned}$$

so that we can write

$$g_{\phi(Y)}\left(\mathrm{D}\phi(Y)[\bar{\xi}_Y], \mathrm{D}\phi(Y)[\bar{\eta}_Y]\right) = \mathrm{tr}\left(\bar{\xi}_Y^\top \bar{\eta}_Y\right).$$

This last expression will allow us to write g as a metric on $\mathcal{S}_+(p,n)$. Replacing $\phi(Y)$ and $\mathrm{D}\phi(Y)[\cdot]$ by their definition yields

$$g_{YY^\top}\left(Y\bar{\xi}_Y^\top + \bar{\xi}_Y Y^\top, Y\bar{\eta}_Y^\top + \bar{\eta}_Y Y^\top\right) = \mathrm{tr}\left(\bar{\xi}_Y^\top \bar{\eta}_Y\right).$$

As in the proof of Proposition A.11, let us write $\bar{\xi}_Y =: Y(Y^\top Y)^{-1}H_\xi + Y_\perp K_\xi$ and

$$\dot{S} := Y\bar{\xi}_Y^\top + \bar{\xi}_Y Y^\top = \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} \tilde{H}_\xi & K_\xi^\top \\ K_\xi & 0 \end{bmatrix} \begin{bmatrix} Y & Y_\perp \end{bmatrix}^\top, \qquad \tilde{H}_\xi := H_\xi(Y^\top Y)^{-1} + (Y^\top Y)^{-1}H_\xi.$$

Analogously, we define $\bar{\eta}_Y =: Y(Y^\top Y)^{-1}H_\eta + Y_\perp K_\eta$ and

$$\check{S} := Y\bar{\eta}_Y^\top + \bar{\eta}_Y Y^\top = \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} \tilde{H}_\eta & K_\eta^\top \\ K_\eta & 0 \end{bmatrix} \begin{bmatrix} Y & Y_\perp \end{bmatrix}^\top, \qquad \tilde{H}_\eta := H_\eta(Y^\top Y)^{-1} + (Y^\top Y)^{-1}H_\eta.$$

We get:

$$\begin{aligned} g_{YY^\top}\left(\dot{S},\check{S}\right) &= \operatorname{tr}\left(\bar{\xi}_Y^\top \bar{\eta}_Y\right) \\ &= \operatorname{tr}\left(H_\xi (Y^\top Y)^{-1} H_\eta + K_\xi^\top K_\eta\right) \\ &= \operatorname{tr}\left(\left(\mathbf{T}^{-1}_{(Y^\top Y)^{-1}} \tilde{H}_\xi\right)(Y^\top Y)^{-1}\left(\mathbf{T}^{-1}_{(Y^\top Y)^{-1}} \tilde{H}_\eta\right) + K_\xi^\top K_\eta\right). \end{aligned}$$

Observe that this metric is different from the embedded submanifold metric inherited from $\mathbb{R}^{n\times n}$ described in [VAV09].

A.5 Projection

Because for every $Y \in \mathbb{R}_*^{n\times p}$ the tangent space $T_Y\mathbb{R}_*^{n\times p} \simeq \mathbb{R}^{n\times p}$ is the direct sum of the vertical space $\mathcal{V}_Y$ and the horizontal space $\mathcal{H}_Y$, every $Z \in \mathbb{R}^{n\times p}$ decomposes uniquely into the sum of a vertical term ${\mathrm{P^v}}_Y(Z)$ and a horizontal term ${\mathrm{P^h}}_Y(Z)$.

Proposition A.12.

$$ {\mathrm{P^v}}_Y(Z) = Y\,\mathbf{T}^{-1}_{Y^T Y}(Y^\top Z - Z^\top Y), \tag{30}$$

$$ {\mathrm{P^h}}_Y(Z) = Z - {\mathrm{P^v}}_Y(Z). \tag{31}$$

where $\mathbf{T}$ *is as in* (28).

Proof. The vertical projection ${\mathrm{P^v}}_Y(Z)$ is characterized by ${\mathrm{P^v}}_Y(Z) = Y\Omega$ with $\Omega \in \mathcal{S}_{\mathrm{skew}}(p)$ and $(Z - {\mathrm{P^v}}_Y(Z)) \in \mathcal{H}_Y$. According to (23), this yields $Y^\top(Z - Y\Omega) - (Z^\top - \Omega^\top Y^\top)Y = 0$, and thus $\Omega = \mathbf{T}^{-1}_{Y^\top Y}(Y^\top Z - Z^\top Y)$, which is well defined by Lemma A.10. The claims follow. ☐

A.6 Gradient

The gradient of a function defined on the quotient $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ can be obtained from the gradient in the total space $\mathbb{R}_*^{n\times p}$.

Proposition A.13. *Let* $f : \mathbb{R}_*^{n\times p}/\mathcal{O}_p \to \mathbb{R}$, *and* $\bar{f} : \mathbb{R}_*^{n\times p} \to \mathbb{R}$ *the corresponding function on* $\mathbb{R}_*^{n\times p}$ *(i.e.,* $\bar{f} = f \circ \pi$*). Then the horizontal lift at* Y *of the gradient of* f *is:*

$$\overline{\mathrm{grad} f}_Y = \mathrm{grad}\bar{f}(Y). \tag{32}$$

Proof. See [AMS08, §3.6.2]. ☐

A.7 Riemannian connection

The theory of Riemannian submersions [O'N83] (or see [AMS08, §5.3.4]) yields the following formula for the Riemannian connection ∇ (also known as the Levi-Civita connection):

$$\overline{\nabla_\eta \xi}_Y = {\mathrm{P^h}}_Y\left(\mathrm{D}\bar{\xi}(Y)[\bar{\eta}_Y]\right), \tag{33}$$

for all vector fields η, ξ on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, where ${\mathrm{P^h}}_Y$ is the horizontal projection (30) and $\mathrm{D}\bar{\xi}(Y)[\bar{\eta}_Y]$ is the derivative of $\bar{\xi}$ (viewed as a function from $\mathbb{R}^{n\times p}$ to $\mathbb{R}^{n\times p}$) at Y along $\bar{\eta}_Y$.

A.8 Hessian

Proposition A.14. *The Riemannian Hessian of a function* $f : \mathbb{R}_*^{n\times p}/\mathcal{O}_p \to \mathbb{R}$ *is given by*

$$\overline{\mathrm{Hess} f(\pi(Y))[\xi_{\pi(Y)}]}_Y = {\mathrm{P^h}}_Y\left(\mathrm{D}\mathrm{grad}\bar{f}(Y)[\bar{\xi}_Y]\right),$$

where $\bar{f} : \mathbb{R}_*^{n\times p} \to \mathbb{R}$ *is the corresponding function on* $\mathbb{R}_*^{n\times p}$.

Proof. Definition 5.5.1 of [AMS08] states that

$$\mathrm{Hess} f(x)[\xi_x] = \nabla_{\xi_x}\mathrm{grad} f.$$

The result comes then from (33). ☐

A.9 The exponential map

It is well known that geodesics on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ are obtained by projection through the quotient map π of horizontal geodesic on $\mathbb{R}_*^{n\times p}$, i.e., straight lines that remain in $\mathbb{R}_*^{n\times p}$. Now, we provide a proof of this result.

Theorem A.15. *Let $\mathcal{D}_Y$ and $\overline{\mathrm{Exp}}$ be defined as in Section 3. For all $\xi_{\pi(Y)} \in \mathrm{D}\pi(Y)[\mathcal{D}_Y]$, the exponential map on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is given by:*

$$\mathrm{Exp}_{\pi(Y)}\xi_{\pi(Y)} = \pi(\overline{\mathrm{Exp}}_Y\bar{\xi}_Y), \tag{34}$$

i.e., geodesics are images of straight lines in $\mathbb{R}_*^{n\times p}$, through the quotient map π, restricted to the time interval where $Y + t\bar{\xi}_Y$ remains full rank.

Proof. By definition, the exponential of a tangent vector $\xi_{\pi(Y)}$ is the point reached at time $t = 1$ by the geodesic emanating from $\pi(Y)$ with initial velocity $\xi_{\pi(Y)}$. It is therefore sufficient to show that the curve $t \mapsto \pi(\overline{\mathrm{Exp}}_Y t\bar{\xi}_Y)$ is a geodesic on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$, starting at $\pi(Y)$ and with initial velocity $\xi_{\pi(Y)}$. We propose here two proofs of this result.

Proof 1: Let $\bar{\xi}_Y \in \mathcal{D}_Y$. Consider the curve $\gamma : t \mapsto \overline{\mathrm{Exp}}_Y t\bar{\xi}_Y = Y + t\bar{\xi}_Y$. The curve γ is a geodesic of $\mathbb{R}_*^{n\times p}$ for all $t \in [0, 1]$, since it is a straight line that remains in $\mathbb{R}_*^{n\times p}$ by definition of $\mathcal{D}_Y$. Moreover $\dot{\gamma}(0)$ is horizontal, hence the theory of Riemannian submersions [GHL04, Prop. 2.109] yields that $\dot{\gamma}(t)$ is horizontal for all $t \in [0, 1]$, and indeed one readily checks that $\gamma(t)^\top\dot{\gamma}(t) = (Y+t\bar{\xi}_Y)^\top\bar{\xi}_Y = Y^\top\bar{\xi}_Y + t\bar{\xi}_Y^\top\bar{\xi}_Y$ is symmetric. The theory also yields that $\pi\circ\gamma = t \mapsto \pi(Y + t\bar{\xi}_Y)$ is a geodesic of $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ for $t \in [0, 1]$. Since $\pi \circ \gamma(0) = \pi(Y)$ and $\frac{\mathrm{d}}{\mathrm{d}t}(\pi \circ \gamma)(0) = \mathrm{D}\pi(Y)[\bar{\xi}_Y] = \xi_{\pi(Y)}$, the proof is complete.

Proof 2: This is a "blindly" constructive proof that does not use the property that horizontal geodesics map to quotient geodesics. To find the horizontal lift through Y of the geodesic through $\pi(Y)$ along $\mathrm{D}\pi(Y)[\bar{\xi}_Y]$, we need to find the curve through Y along $\bar{\xi}_Y$ such that $\overline{\nabla}_{\dot{Y}(t)}\dot{Y}(t) \in \mathcal{V}_{Y(t)}$ for all t. (This condition follows from the formula $\overline{\nabla_\eta\xi}_x = \mathrm{P}_x^h\overline{\nabla}_{\bar{\eta}_x}\bar{\xi}$ that relates the connection ∇ on the quotient to the connection $\overline{\nabla}$ on the total space when the quotient map is a Riemannian submersion, see Section A.7.) Since the curve must be horizontal, we have $Y^\top\dot{Y} - \dot{Y}^\top Y = 0$. Differentiating once yields $Y^\top\ddot{Y} - \ddot{Y}^\top Y = 0$. Moreover, since $\overline{\nabla}_{\dot{Y}(t)}\dot{Y}(t) = \ddot{Y}(t)$ must belong to $\mathcal{V}_{Y(t)}$ for all t, we have $\ddot{Y} = Y\Omega$ where Ω is skew-symmetric. Replacing this in the previous second-order differential equation yields $Y^\top Y\Omega + \Omega Y^\top Y = 0$. Since $Y^\top Y$ is invertible, it follows from the theory of Sylvester equations (see [Gan59, Ch. VI]) that $\Omega = 0$ is the only solution. Hence $\ddot{Y} = 0$, which implies that the horizontal lift at Y of the geodesic with initial velocity $\xi_{\pi(Y)} = D\pi(Y)[\bar{\xi}_Y]$ is $Y(t) = Y + \bar{t}\xi_Y$ for $t \in [0, 1]$. $\square$

The exponential on $(\mathcal{S}_+(p, n), g)$ is thus given by $\mathrm{Exp}_{\phi(Y)}\mathrm{D}\phi(Y)[\bar{\xi}_Y] = \phi(Y + \bar{\xi}_Y)$ (for $\bar{\xi}_Y \in \mathcal{D}_Y$), i.e.,

$$\mathrm{Exp}_{YY^\top}(Y\bar{\xi}_Y^\top + \bar{\xi}_Y Y^\top) = (Y + \bar{\xi}_Y)(Y + \bar{\xi}_Y)^\top.$$

A.10 A retraction

Following the theory in [AMS08, §4.1.2], a retraction on $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ is given by

$$R_{\pi(Y)}\xi_{\pi(Y)} = \pi(Y + \bar{\xi}_Y), \tag{35}$$

provided that this definition makes sense, i.e., the right-hand side only depends on Y through $\pi(Y)$ only. This is the case, since for all $Q \in \mathcal{O}_p$, $YQ + \bar{\xi}_{YQ} = (Y + \bar{\xi}_Y)Q$. Interestingly, with (35), we have recovered the exponential map (34). This situation is rather unusual (compare with the examples in [AMS08, §4.1.2]). This is for example not the case for the quotient geometry of $\mathcal{S}_+(p, n)$ investigated in [BS09, VAV13] or for the Grassmann manifold $\mathrm{Grass}(\mathrm{n}, \mathrm{p})$ identified to the quotient $\mathbb{R}_*^{n\times p}/\mathrm{GL}(p)$ [AMS04].

A.11 Curvature

Finally, we summarize some recent results, obtained in [MHA19], regarding the curvature of the manifold.

Theorem A.16. *Let* ξ, η, α *and* β *be vector fields on* $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$*, and let* $\bar{\xi}, \bar{\eta}, \bar{\alpha}$ *and* $\bar{\beta}$ *be their horizontal lifts. The Riemannian curvature tensor at* $\pi(Y)$ *satisfies:*

$$\begin{aligned} g\left(R_{\mathbb{R}_*^{n\times p}/\mathcal{O}_p}(\xi_{\pi(Y)}, \eta_{\pi(Y)})\alpha_{\pi(Y)}, \beta_{\pi(Y)}\right) &= \frac{1}{2}\langle \mathrm{P^v}_Y[\bar{\xi}, \bar{\eta}], \mathrm{P^v}_Y[\bar{\alpha}, \bar{\beta}]\rangle \\ &-\frac{1}{4}\left(\langle \mathrm{P^v}_Y[\bar{\eta}, \bar{\alpha}], \mathrm{P^v}_Y[\bar{\xi}, \bar{\beta}]\rangle - \langle \mathrm{P^v}_Y[\bar{\xi}, \bar{\alpha}], \mathrm{P^v}_Y[\bar{\eta}, \bar{\beta}]\rangle\right), \end{aligned} \tag{36}$$

where $[\bar{\xi}, \bar{\eta}]$ *is the Lie bracket in* $\mathbb{R}_*^{n\times p}$*, and* $\mathrm{P^v}_Y[\bar{\xi}, \bar{\eta}]$ *is given by (2).*

Corollary A.17. *Let* $\xi_{\pi(Y)}, \eta_{\pi(Y)}$ *be (independent) tangent vectors on* $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$*, with horizontal lifts* $\bar{\xi}_Y$*,* $\bar{\eta}_Y$*. The sectional curvature at* $\pi(Y)$ *in* $\mathbb{R}_*^{n\times p}/\mathcal{O}_p$ *is*

$$K_{\mathbb{R}_*^{n\times p}/\mathcal{O}_p}(\xi_{\pi(Y)}, \eta_{\pi(Y)}) = \frac{3||Y\mathbf{T}^{-1}_{Y^\top Y}\left(\bar{\eta}_Y^\top \bar{\xi}_Y - \bar{\xi}_Y^\top \bar{\eta}_Y\right)||_{\mathrm{F}}^2}{\langle \bar{\xi}_Y, \bar{\xi}_Y\rangle\langle \bar{\eta}_Y, \bar{\eta}_Y\rangle - \langle \bar{\xi}_Y, \bar{\eta}_Y\rangle^2}. \tag{37}$$

Theorem A.18. *If* $p = 1$*, the sectional curvature is always zero. If* $p \geq 2$*, the minimum over the tangent planes of the sectional curvature is zero, while the maximum is equal to* $3/(\sigma_{p-1}^2 + \sigma_p^2)$*, where* σ_i *is the* i^{th} *largest singular value of* Y*.*

References

[AGSW16] P.-A. Absil, Pierre-Yves Gousenbourger, Paul Striewski, and Benedikt Wirth. Differentiable piecewise-Bézier surfaces on Riemannian manifolds. *SIAM Journal on Imaging Sciences*, 9(4):1788–1828, 2016. doi:10.1137/16M1057978.

[AIDV08] P.-A. Absil, M. Ishteva, L. De Lathauwer, and S. Van Huffel. A geometric Newton method for Oja's vector field. *Neural Computation*, 21(5):1415–1433, 2008. doi:10.1162/neco.2008.04-08-749.

[AMR88] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1988.

[AMS04] P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80(2):199–220, 2004. doi:10.1023/B:ACAP.0000013855.14971.91.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[AO14] P.-A. Absil and I. V. Oseledets. Low-rank retractions: a survey and new results. *Computational Optimization and Applications*, 62(1):5–29, 2014. doi:10.1007/s10589-014-9714-4.

[ATV13] Bijan Afsari, Roberto Tron, and René Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013. doi:10.1137/12086282X.

[BAC19] Nicolas Boumal, P-A Absil, and Coralia Cartis. Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33, 2019. doi:10.1093/imanum/drx080.

[BC70] F. Brickell and R. S. Clark. *Differentiable Manifolds*. Van Nostrand Reinhold, London, 1970.

[BCS13] Silvère Bonnabel, Anne Collard, and Rodolphe Sepulchre. Rank-preserving geometric means of positive semi–definite matrices. *Linear Algebra and its Applications*, 438(8):3202–3216, 2013. doi:10.1016/j.laa.2012.12.009.

[BHS18] Craig Bakker, Mahantesh Halappanavar, and Arun Visweswara Sathanur. Dynamic graphs, community detection, and Riemannian geometry. *Applied Network Science*, 3(3), 2018. doi:10.1007/s41109-018-0059-2.

[BJL16] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2016. doi:10.1016/j.exmath.2018.01.002.

[BM05] Samuel Burer and Renato D. C. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103(3, Ser. A):427–444, 2005. doi:10.1007/s10107-004-0564-1.

[BMAS14] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014. URL: http://www.manopt.org.

[BMS10] S. Bonnabel, G. Meyer, and R. Sepulchre. Adaptive filtering for estimation of a low-rank positive semidefinite matrix. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems*, 2010.

[Bon13] Silvère Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions Automatic Control*, 58(9):2217–2229, 2013. doi:10.1109/TAC.2013.2254619.

[Boo86] William M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, Florida, second edition, 1986.

[BR97] Rajendra Bhatia and Peter Rosenthal. How and why to solve the operator equation AX-XB= Y. *Bulletin of the London Mathematical Society*, 29(1):1–21, 1997.

[BS09] Silvère Bonnabel and Rodolphe Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070, 2009. doi:10.1137/080731347.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, UK, 2004.

[DE99] Luca Dieci and Timo Eirola. On smooth decompositions of matrices. *SIAM Journal on Matrix Analysis and Applications*, 20(3):800–819, 1999. doi:10.1137/S0895479897330182.

[DH18] Ahmed Douik and Babak Hassibi. Low-Rank Riemannian Optimization on Positive Semidefinite Stochastic Matrices with Applications to Graph Clustering. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1299–1308, 2018.

[DK12] J. J. Duistermaat and J. A. C. Kolk. *Lie groups*. Universitext. Springer-Verlag, Berlin, 2012.

[FHP16] Masoud Faraki, Mehrtash T. Harandi, and Fatih Porikli. Image set classification by symmetric positive semi-definite matrices. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016. doi:10.1109/WACV.2016.7477621.

[Gan59] F. R. Gantmacher. *The Theory of Matrices I, II*. AMS Chelsea, Rhode Island, 1959.

[Gel90] Matthias Gelbrich. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990. doi:10.1002/mana.19901470121.

[GHL04] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry*. Universitext. Springer-Verlag, Berlin, third edition, 2004. doi:10.1007/978-3-642-18855-8.

[GMA18] Pierre-Yves Gousenbourger, Estelle Massart, and P.-A. Absil. Data fitting on manifolds with composite Bézier-like curves and blended cubic splines. *Journal of Mathematical Imaging and Vision*, 61(5):645–671, 2018. doi:10.1007/s10851-018-0865-2.

[GMM$^+$17] P.-Y. Gousenbourger, E. Massart, A. Musolas, P.-A. Absil, L. Jacques, J. M. Hendrickx, and Y. Marzouk. Piecewise-Bézier C^1 smoothing on manifolds with application to wind field estimation. In *Proceedings of the 25^{th} European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, pages 305–310, 2017.

[GV96] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore-London, third edition, 1996.

[HGZ17] Wen Huang, K. A. Gallivan, and Xiangxiong Zhang. Solving PhaseLift by low-rank Riemannian optimization methods for complex semidefinite constraints. *SIAM Journal on Scientific Computing*, 39(5):B840–B859, 2017. doi:10.1137/16M1072838.

[Hub81] Peter J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.

[JBAS10] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM Journal on Optimization*, 20(5):2327–2351, 2010. doi:10.1137/080731359.

[KDB+18] Anis Kacem, Mohamed Daoudi, Boulbaba Ben Amor, Stefano Berretti, and Juan Carlos Alvarez-Paiva. A Novel Geometric Framework on Gram Matrix Trajectories for Human Behavior Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2018. doi:10.1109/TPAMI.2018.2872564.

[KM18] Hiroyuki Kasai and Bamdev Mishra. Inexact trust-region algorithms on Riemannian manifolds. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 4249–4260. Curran Associates, Inc., 2018. URL: http://papers.nips.cc/paper/7679-inexact-trust-region-algorithms-on-riemannian-manifolds.pdf.

[KN63] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry.* Interscience Publishers, a division of John Wiley & Sons, New York-London, 1963. Volumes 1 and 2.

[LB14] X.-B. Li and F. J. Burkowski. Conformational transitions and principal geodesic analysis on the positive semidefinite matrix manifold. In *Basu M., Pan Y., Wang J. (eds) Bioinformatics Research and Applications. ISBRA 2014. Lecture Notes in Computer Science*, volume 8492, pages 334–345. Springer Cham, 2014. doi:10.1007/978-3-319-08171-7_30.

[LCB+04] Gert R. G. Lanckriet, Nello Cristianini, Peter Bartlett, Laurent El Ghaoui, and Michael I. Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5(Jan):27–72, 2004.

[Lee11] John M. Lee. *Introduction to Topological Manifolds*, volume 202 of *Graduate Texts in Mathematics.* Springer Science Business Media, New York, second edition, 2011.

[Lee13] John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics.* Springer Science Business Media, New York, second edition, 2013.

[LT85] Peter Lancaster and Miron Tismenetsky. *The theory of matrices.* Computer Sciences and Applied Mathematics. Academic Press, second edition, 1985.

[MBS11] Gilles Meyer, Silvère Bonnabel, and Rodolphe Sepulchre. Regression on fixed-rank positive semidefinite matrices: a Riemannian approach. *Journal of Machine Learning Research*, 12(Feb):593–625, 2011.

[MGS+19] Estelle Massart, Pierre-Yves Gousenbourger, Nguyen Thanh Son, Tatjana Stykel, and P.-A. Absil. Interpolation on the manifold of fixed-rank positive-semidefinite matrices for parametric model order reduction: preliminary results. In *Proceedings of the 27th European Symposium on Artifical Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, pages 281–286, 2019.

[MHA19] Estelle Massart, Julien M. Hendrickx, and P.-A. Absil. Curvature of the manifold of fixed-rank positive-semidefinite matrices endowed with the Bures-Wasserstein metric. In *Proceedings of the 4th conference on Geometric Sciences of Information (GSI)*, pages 739–748, 2019.

[MHB+16] Melissa Marchand, Wen Huang, Arnaud Browet, Paul Van Dooren, and Kyle A. Gallivan. A Riemannian optimization approach for role model extraction. In *Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems*, pages 58–64, 2016.

[MMS11] B. Mishra, G. Meyer, and R. Sepulchre. Low-rank optimization for distance matrix completion. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 4455–4460, 2011. `doi:10.1109/CDC.2011.6160810`.

[O'N83] Barrett O'Neill. *Semi-Riemannian Geometry*, volume 103 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.

[Rau18] Christian Rau. On Procrustes matching of non-negative matrices and an application to random tomography. *Linear and Multilinear Algebra*, 66(1):96–106, 2018. `doi:10.1080/03081087.2017.1284741`.

[SAV+19] David Sabbagh, Pierre Ablin, Gael Varoquaux, Alexandre Gramfort, and Denis A. Engemann. Manifold-regression to predict from MEG/EEG brain signals without source modeling. arXiv:1906.02687, 2019.

[SDB+19] Benjamin Szczapa, Mohamed Daoudi, Stefano Berretti, Alberto Del Bimbo, Pietro Pala, and Estelle Massart. Fitting, Comparison, and Alignment of Trajectories on Positive Semi-Definite Matrices with Application to Action Recognition. submitted to the 10th International Workshop on Human Behavior Understanding (in conjunction with ICCV'2019), arxiv:1908.00646, 2019.

[Ste98] G. W. Stewart. *Matrix algorithms. Vol. I.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 1998. `doi:10.1137/1.9781611971408`.

[Tak11] A. Takatsu. Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, 48(4):1005–1026, 2011.

[TAV13] Roberto Tron, Bijan Afsari, and René. Vidal. Riemannian consensus for manifolds with bounded curvature. *IEEE Transactions on Automatic Control*, 58(4):921–934, 2013. `doi:10.1109/TAC.2012.2225533`.

[VAV09] Bart Vandereycken, P.-A. Absil, and Stefan Vandewalle. Embedded geometry of the set of symmetric positive semidefinite matrices of fixed rank. In *IEEE/SP 15th Workshop on Statistical Signal Processing*, pages 389–392, 2009. `doi:10.1109/SSP.2009.5278558`.

[VAV13] Bart Vandereycken, P.-A. Absil, and Stefan Vandewalle. A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank. *IMA Journal of Numerical Analysis*, 33(2):481–514, 2013. `doi:10.1093/imanum/drs006`.