



**HAL**  
open science

## Bures-Wasserstein minimizing geodesics between covariance matrices of different ranks

Yann Thanwerdas, Xavier Pennec

► **To cite this version:**

Yann Thanwerdas, Xavier Pennec. Bures-Wasserstein minimizing geodesics between covariance matrices of different ranks. *SIAM Journal on Matrix Analysis and Applications*, 2023, 44 (3), pp.1447-1476. 10.1137/22M149168X . hal-03647321

**HAL Id: hal-03647321**

**<https://hal.science/hal-03647321>**

Submitted on 20 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# BURES-WASSERSTEIN MINIMIZING GEODESICS BETWEEN COVARIANCE MATRICES OF DIFFERENT RANKS \*

YANN THANWERDAS<sup>†</sup> AND XAVIER PENNEC<sup>‡</sup>

**Abstract.** The set of covariance matrices equipped with the Bures-Wasserstein distance is the orbit space of the smooth, proper and isometric action of the orthogonal group on the Euclidean space of square matrices. This construction induces a natural orbit stratification on covariance matrices, which is exactly the stratification by the rank. Thus, the strata are the manifolds of symmetric positive semi-definite (PSD) matrices of fixed rank endowed with the Bures-Wasserstein Riemannian metric. In this work, we study the geodesics of the Bures-Wasserstein distance. Firstly, we complete the literature on geodesics in each stratum by clarifying the set of preimages of the exponential map and by specifying the injection domain. We also give explicit formulae of the horizontal lift, the exponential map and the Riemannian logarithms that were kept implicit in previous works. Secondly, we give the expression of all the minimizing geodesic segments joining two covariance matrices of any rank. More precisely, we show that the set of all minimizing geodesics between two covariance matrices  $\Sigma$  and  $\Lambda$  is parametrized by the closed unit ball of  $\mathbb{R}^{(k-r) \times (l-r)}$  for the spectral norm, where  $k, l, r$  are the respective ranks of  $\Sigma, \Lambda, \Sigma\Lambda$ . In particular, the minimizing geodesic is unique if and only if  $r = \min(k, l)$ . Otherwise, there are infinitely many.

**Key words.** Covariance matrices, PSD matrices, Bures-Wasserstein, orbit space, geodesics, injection domain

**AMS subject classifications.** 15B48, 15A63, 53B20, 53C22, 58D17, 53-08, 53A04, 54E50, 58A35.

**1. Introduction.** Many data can be represented as covariance matrices. They are often assumed to be Symmetric Positive Definite (SPD) because it is much more convenient from the geometric point of view. Indeed, the set of SPD matrices is an open convex cone in the vector space of symmetric matrices so it has a canonical differential structure. The induced Euclidean metric is not satisfying to compute with SPD matrices because geodesics leave the space in finite time and interpolations are often non-realistic. To solve this problem, a lot of Riemannian metrics were proposed on SPD matrices, mainly  $O(n)$ -invariant metrics [39] (affine-invariant [34, 35, 32, 20, 14, 28], log-Euclidean [2, 13, 17], Bures-Wasserstein [9, 29, 36, 37, 23, 3], Bogoliubov-Kubo-Mori [33, 27], etc.), Cholesky-like metrics [21, 22] or product metrics with one metric on positive diagonal matrices and one metric on full-rank correlation matrices [38, 40].

However, this viewpoint often forgets about singular covariance matrices, that is covariance matrices with non-full rank. Altogether, they form a closed convex cone which is not anymore a manifold. First, it can be equipped with distances to provide a metric space structure. The Euclidean distance is not satisfying either here because geodesics leave the closed cone in finite time again. The main alternative is the Bures-Wasserstein distance that was introduced in many different ways. For example in [9, 29], it is the Wasserstein/Kantorovitch distance between multivariate centered Gaussian distributions, possibly degenerate. It is also the quotient distance of the Euclidean metric on square matrices by the right action of the orthogonal group. This is why it was also called the Procrustes distance [10, 16]. This viewpoint allows to split the closed cone into strata that are Riemannian manifolds whose induced geodesic distance is precisely the Bures-Wasserstein distance. In particular, the space

---

\*Preprint, April 1st, 2022.

<sup>†</sup>Université Côte d'Azur and Inria, Epione Project Team ([yann.thanwerdas@inria.fr](mailto:yann.thanwerdas@inria.fr)).

<sup>‡</sup>Université Côte d'Azur and Inria, Epione Project Team ([xavier.pennec@inria.fr](mailto:xavier.pennec@inria.fr)).

of covariance matrices equipped with this distance is a complete geodesic metric space.

The geometry of stratified spaces is a topic of interest in the community of statisticians in non-linear spaces. Examples of popular stratified spaces are the Kendall shape spaces [18], the BHV space of trees [4], the QED space of trees [11], the Graph space [7], the Wald space of forests [15], the correlation matrices or the symmetric/diagonal matrices stratified by eigenvalue multiplicity. Moreover, the space of covariance matrices with the Bures-Wasserstein distance is a metric space of non-negative curvature [37]. Spaces of this type have been much less described than metric spaces of non-positive curvature [6]. These are two motivations to study the Bures-Wasserstein geometry of covariance matrices. In this work, we focus on geodesics.

We would like to highlight three important elements that depart from usual studies on the geometry of covariance matrices. Firstly, contrarily to the affine-invariant, the log-Euclidean or the log-Cholesky metrics, the Bures-Wasserstein metric on SPD matrices (or on PSD matrices of fixed rank) is not geodesically complete. Therefore in this work, we need the notion of *definition domain* of the exponential map because it is not  $\mathbb{R}$  in general. Secondly, in the cited examples, the geodesics are globally minimizing on  $\mathbb{R}$ . It is not the case for the Bures-Wasserstein metric so we need the notions of *cut time* and *injection domain* to specify when the geodesic stops to be minimizing. Thirdly, in this work we talk about *geodesics in a metric space*, not only in a Riemannian manifold as previously. We recall the definition of these notions in Section 2.

In the principal (or regular) stratum of SPD matrices, the Bures-Wasserstein Riemannian metric was extensively studied. Since  $X \in \text{GL}(n) \mapsto XX^\top \in \text{Sym}^+(n)$  is a Riemannian submersion, many geometric operations can be computed thanks to O'Neill's equations [30]. Therefore, the curvature was derived in [36, 37, 39], the quotient geometry was described in [23, 3, 41], the exponential map was computed in [23], a Riemannian logarithm was given in [3], the injectivity radius was computed in [24] and a simplified equation of the geodesic parallel transport was proposed in [39]. In contrast to geodesically complete Riemannian metrics, it is important to specify the definition domain of the exponential map. It was characterized in [23] as the connected component of 0 in a subset of  $\mathbb{R}$ , which could be specified more explicitly. Moreover, the uniqueness of the Riemannian logarithm is not established and the injection domain seems to be unknown. Therefore in Section 4, we clarify the definition domain of the exponential map, we prove the uniqueness of the logarithm thanks to a result from [24] and we prove that the geodesics are minimizing on their domain of definition, which also provides the injection domain.

In each other stratum of PSD matrices of fixed rank  $k < n$ , the Bures-Wasserstein Riemannian metric was studied via the analogous Riemannian submersion defined by  $X \in \mathbb{R}_*^{n \times k} \mapsto XX^\top \in \text{Sym}^+(n, k)$ , where  $\mathbb{R}_*^{n \times k}$  is the open set of matrices of full rank  $k$  in  $\mathbb{R}^{n \times k}$  and  $\text{Sym}^+(n, k)$  is the set of PSD matrices of size  $n$  and rank  $k$ . The curvature was computed in [25], the exponential map, its domain of definition, the logarithm map and the injectivity radius were derived in [24]. The horizontal lift was kept implicit so these results are formulated in the total space  $\mathbb{R}_*^{n \times k}$ . We think that it is easier to understand the geometry with formulae depending directly on the tangent vector and not on its horizontal lift. Therefore in Section 5, we compute the horizontal lift and we give the expressions of the Riemannian metric, the exponential map and its definition domain in function of vectors tangent to the manifold  $\text{Sym}^+(n, k)$ . Moreover, we notice that the solutions of the logarithm equation characterized in [24] may not be in the definition domain. Thus we explain the additional condition for being a preimage of the exponential map. In addition, we

give an explicit bijective parametrization of the Riemannian logarithms, which allows us to count them. When it is unique, we give an explicit formula of the corresponding minimizing geodesic in function of the end points. This finally allows us to compute the injection domain which is kept implicit in [24].

Beyond the clarification and completion of the literature on geodesics in each stratum, our main objective is to characterize the minimizing geodesics *between* strata. Our main results are the following. (1) Any minimizing geodesic segment between two covariance matrices  $\Sigma$  and  $\Lambda$  is of constant rank on the interior of the segment. It is called the rank of the minimizing geodesic and it is greater than the ranks of  $\Sigma$  and  $\Lambda$ . (2) We give the explicit formula of all the minimizing geodesic segments in Theorem 6.2. (3) They are parametrized by the vectors of the closed unit ball of  $\mathbb{R}^{(k-r) \times (l-r)}$  for the spectral norm, where  $k, l, r$  are the respective ranks of  $\Sigma, \Lambda, \Sigma\Lambda$ . In other words, they are parametrized by matrices  $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$  with singular values in  $[0, 1]$ . (4) The minimizing geodesic segment is unique if and only if  $r = \min(k, l)$  (this includes  $\max(k, l) = n$ ). Otherwise, there are infinitely many. (5) The number of minimizing geodesics of minimal rank (i.e. of rank equal to  $\max(k, l)$ ) is 1 if  $r = \min(k, l)$ , otherwise it is 2 if  $k = l$ , otherwise it is infinite. (6) Assuming  $k \geq l$ , if  $R_0$  belongs to the Stiefel manifold  $\text{St}(k-r, l-r)$ , that is  $R_0^\top R_0 = I_{l-r}$ , then the corresponding geodesic is of minimal rank. (7) The choice of parameter  $R_0 = 0$  leads to the geodesic  $\gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2}) - \Sigma^{1/2} \Lambda)$  whose expression does not depend on the ranks of  $\Sigma$  and  $\Lambda$ . It is called the Bures-Wasserstein canonical geodesic.

In the remainder of this section, we introduce some matrix notations. In Section 2, we introduce the important concepts of geodesics and quotient space in metric spaces and manifolds. We give a particular attention to the notions that characterize if a geodesic (self-parallel curve) is minimizing: cut time, injection domain, difference between preimages of the exponential map and Riemannian logarithms. In Section 3, we recall the algebraic structure, metric topology and differential geometry of the convex cone of covariance matrices seen as the quotient of square matrices by the orthogonal group. In Sections 4 and 5, we complete the literature on the geodesics of the Bures-Wasserstein metric on SPD matrices and on singular matrices of fixed rank respectively. In Section 6, we give our main results on minimizing geodesics in the whole Bures-Wasserstein metric space of covariance matrices. We conclude in Section 7. The main proofs are deferred to the Appendix.

*Matrix notations.* Let  $n, k \in \mathbb{N}$ . In this work, we use the following manifolds of matrices:

- the vector space of  $n \times k$  matrices  $\mathbb{R}^{n \times k}$ ,
- the open subset  $\mathbb{R}_*^{n \times k} \subset \mathbb{R}^{n \times k}$  of full-rank matrices,
- in particular, the vector space of square matrices  $\text{Mat}(n) = \mathbb{R}^{n \times n}$  and the general linear group  $\text{GL}(n) = \mathbb{R}_*^{n \times n}$ ,
- the orthogonal group  $\text{O}(n)$ ,
- the Stiefel manifold  $\text{St}(n, k) = \text{O}(n)/\text{O}(n-k)$ ,
- the manifold of symmetric positive definite matrices  $\text{Sym}^+(n)$ ,
- the manifold of symmetric positive semi-definite matrices of fixed rank  $k$ ,  $\text{Sym}^+(n, k)$ ,
- the vector space of diagonal matrices  $\text{Diag}(n)$ ,
- the groups of invertible diagonal matrices  $\text{Diag}^*(n) = \text{Diag}(n) \cap \text{GL}(n)$  and positive diagonal matrices  $\text{Diag}^+(n) = \text{Diag}(n) \cap \text{Sym}^+(n)$ .

In Section 6, we need to distinguish cases where matrices may have one or two null dimensions, i.e. belonging to  $\mathbb{R}^{n \times 0}$ ,  $\mathbb{R}^{0 \times k}$  or  $\mathbb{R}^{0 \times 0}$ . Thus we recall that these spaces are isomorphic to the vector space  $\{0\}$ . Indeed, there is a unique linear map from  $\mathbb{R}^n$  to  $\mathbb{R}^0$  or from  $\mathbb{R}^0$  to  $\mathbb{R}^k$ , which is the identically null map. The canonical basis of  $\mathbb{R}^0$  is empty and the corresponding matrix in the canonical bases is called the empty matrix. It is practical to treat these spaces as non-trivial spaces to avoid writing particular cases. In particular,  $\text{St}(n, 0) = \mathbb{R}^{n \times 0}$  and  $\text{O}(0) = \text{GL}(0) = \text{Mat}(0) = \text{Diag}(0)$  are sets of cardinal 1.

We use the following notations.

- $I_n$  denotes the identity matrix of size  $n$ .
- $\mathbf{0}_n$  denotes the null matrix of size  $n \times n$ .  $\mathbf{0}_{n,k}$  denotes the null matrix of size  $n \times k$ . We may simply denote them 0 when sizes are obvious in the context.
- (Sylvester equation)  $\mathcal{S}_A(B)$  is the unique solution  $Z$  of the Sylvester equation  $AZ + ZA = B$  for  $A \in \text{Sym}^+(k)$  and  $B \in \text{Sym}(k)$ .
- (Löwner order) For all  $\Sigma \in \text{Sym}(n)$ , we say that  $\Sigma$  is positive definite (resp. positive semi-definite) and we denote  $\Sigma > 0$  (resp.  $\Sigma \geq 0$ ) when  $\Sigma$  has positive (resp. non-negative) eigenvalues. Given  $\Lambda \in \text{Sym}(n)$ , we denote  $\Sigma > \Lambda$  (resp.  $\Sigma \geq \Lambda$ ) when  $\Sigma - \Lambda > 0$  (resp.  $\Sigma - \Lambda \geq 0$ ).

We recall basic facts on symmetric matrices.

1. (Eigenvalue decomposition) For all  $\Sigma \in \text{Sym}(n)$ , there exist  $U \in \text{O}(n)$  and  $D \in \text{Diag}(n)$  such that  $\Sigma = UDU^\top$ . By removing null eigenvalues, given  $r = \text{rk}(\Sigma)$ , there also exist  $U \in \text{St}(n, r)$  and  $D \in \text{Diag}^*(r)$  such that  $\Sigma = UDU^\top$ .
2. For all  $X \in \mathbb{R}^{n \times k}$ ,  $XX^\top$  is symmetric positive semi-definite and  $\text{rk}(X) = \text{rk}(XX^\top) = \text{rk}(X^\top X)$ . In particular, if  $X \in \mathbb{R}_*^{n \times k}$ , then  $X^\top X \in \text{GL}(k)$ .
3. (Singular value decomposition) For all  $M \in \mathbb{R}^{n \times k}$ , denoting  $r = \text{rk}(M) \leq \min(n, k)$ , there exist  $U \in \text{O}(n)$ ,  $V \in \text{O}(k)$  and  $D = \begin{pmatrix} D_r & \mathbf{0}_{r, k-r} \\ \mathbf{0}_{n-r, r} & \mathbf{0}_{n-r, k-r} \end{pmatrix}$  with  $D_r \in \text{Diag}^+(r)$  such that  $M = UDV^\top$ . The diagonal entries of  $D$ , that is the  $D'_{ii}$ s for  $i \in \{1, \dots, \min(n, k)\}$  are called the singular values of  $M$ .
4. (Moore-Penrose inverse) For all  $M \in \mathbb{R}^{n \times k}$ , the unique matrix  $M^- \in \mathbb{R}^{k \times n}$  satisfying  $MM^-M = M$ ,  $M^-MM^- = M^-$ ,  $MM^- \in \text{Sym}(n)$  and  $M^-M \in \text{Sym}(k)$  is called the pseudoinverse or Moore-Penrose inverse. In this work, we only use it for symmetric matrices  $\Sigma \in \text{Sym}(k)$ . Given  $r = \text{rk}(\Sigma)$  and an eigenvalue decomposition  $\Sigma = UDU^\top = P\text{Diag}(D, 0)P^\top$  with  $U \in \text{St}(k, r)$ ,  $P = [U \ U_\perp] \in \text{O}(k)$ ,  $D \in \text{Diag}^*(r)$ , it is easy to check that  $\Sigma^- = UD^{-1}U^\top = P\text{Diag}(D^{-1}, 0)P^\top$ . Moreover, for all  $V \in \text{St}(n, k)$ ,  $(V\Sigma V^\top)^- = V\Sigma^-V^\top$ .
5. (Symmetric square root) For all  $\Sigma \in \text{Sym}(n)$ , if  $\Sigma \geq 0$ , then there exists a unique matrix  $A \in \text{Sym}(n)$ ,  $A \geq 0$  such that  $A^2 = \Sigma$ . It is called the (symmetric) square root of  $\Sigma$  and it is denoted  $\sqrt{\Sigma}$  or  $\Sigma^{1/2}$ .
6. (Non-symmetric square roots) For all  $X, Y \in \mathbb{R}^{n \times k}$ ,  $XX^\top = YY^\top$  if and only if there exists  $U \in \text{O}(k)$  such that  $XU = Y$ . Indeed, if  $Y = XU$ , then  $YY^\top = XX^\top$ . Conversely, if  $X = (x_1, \dots, x_n)^\top$  and  $Y = (y_1, \dots, y_n)^\top$  are such that  $YY^\top = XX^\top$ , then we can define the map  $\varphi : \text{span}(x_1, \dots, x_n) \rightarrow \text{span}(y_1, \dots, y_n)$  by  $\varphi(x_i) = y_i$ , extended by linearity. Indeed, it is well defined because if  $\sum_{i=1}^n \lambda_i x_i = 0$  or equivalently  $X^\top \lambda = 0$ , then  $\|Y^\top \lambda\|^2 = \lambda^\top YY^\top \lambda = \lambda^\top XX^\top \lambda = 0$  and  $Y^\top \lambda = \sum_{i=1}^n \lambda_i y_i = 0$ . Moreover,  $\varphi$  is an isometric map, thus it is injective. Since  $\text{rk}(X) = \text{rk}(XX^\top) = \text{rk}(YY^\top) =$

$\text{rk}(Y)$ , it is bijective so it is an isometry between two subvector spaces of  $\mathbb{R}^k$ . According to Witt's theorem,  $\varphi$  extends to an isometry of  $\mathbb{R}^k$ . Thus the matrix  $U = [\varphi]^\top$  in the canonical basis of  $\mathbb{R}^k$  satisfies  $XU = Y$ .

We manipulate several norms on matrices. For  $M \in \mathbb{R}^{n \times k}$  and  $x \in \mathbb{R}^k$ , we denote:

1. (Euclidean norm)  $\|M\|_2 = \text{tr}(M^\top M)^{1/2}$  and  $\|x\|_2 = (x^\top x)^{1/2}$ .
2. (Infinite norm)  $\|M\|_\infty = \max_{1 \leq i, j \leq n} |M_{ij}|$  and  $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$ .
3. (Spectral norm or Schatten's infinite norm)  $\|M\|_S = \sup_{\substack{x \in \mathbb{R}^k \\ \|x\|_2 \leq 1}} \|Mx\|_2 =$

$\|\sigma(M)\|_\infty$ , where  $\sigma(M)$  is the vector of singular values of  $M$ .

Without index,  $\|\cdot\|$  generically denotes the norm on the tangent spaces associated to the Riemannian metric at hand.

**2. Preliminary concepts.** In this section, we recall the basic definitions of geodesics and quotient spaces. We depart from works on Riemannian metrics on SPD matrices such as the affine-invariant, the log-Euclidean or the log-Cholesky metrics which are geodesically complete and where the injection domain of the exponential map is the entire manifold. We also need to introduce the notions of length and geodesic in a metric space [6, 31] and the notion of Riemannian orbit space [1].

In Section 2.1, we introduce the concepts needed to study geodesics and minimizing geodesics. We recall that geodesics are defined in a metric space, in a manifold endowed with an affine connection and in a Riemannian manifold, where the two previous notions coincide. Then we introduce our notations for preimages of the exponential map, logarithms and related notions. We define the cut time and the injection domain, extending the definition usually given in complete metric spaces, and we explain where to be cautious. In Section 2.2, we recall the notions of quotient metric space, quotient Riemannian manifold and Riemannian orbit space. We introduce the vocabulary, notations and results that we use in the next sections.

### 2.1. Geodesics.

**DEFINITION 2.1 (Curve).** *Let  $\mathcal{M}$  be a topological space. A curve on  $\mathcal{M}$  is a continuous map  $c : I \rightarrow \mathcal{M}$ , where  $I$  is an interval of  $\mathbb{R}$ . When  $I$  is a segment of  $\mathbb{R}$ , we may call  $c$  a segment.*

#### 2.1.1. Geodesics in a metric space.

**DEFINITION 2.2 (Length, length distance, length space).** [31] *Let  $(\mathcal{M}, d)$  be a metric space.*

1. (Length) *Let  $c : [a, b] \subseteq I \rightarrow \mathcal{M}$  be a curve. The length of  $c$  is defined by  $L(c) = \sup \sum_{k=0}^p d(c(t_k), c(t_{k+1})) \in [0, +\infty]$  over all subdivisions  $a = t_0 \leq t_1 \leq \dots \leq t_p \leq t_{p+1} = b$ . We say that  $c : I \rightarrow \mathcal{M}$  is rectifiable when for all  $a \leq b$  in  $I$ ,  $L(c|_{[a,b]})$  is finite.*
2. (Length distance) *The length distance between  $x \in \mathcal{M}$  and  $y \in \mathcal{M}$  is defined by  $d_L(x, y) = \inf L(c) \geq d(x, y)$  over all rectifiable curves  $c : [0, 1] \rightarrow \mathcal{M}$  from  $x = c(0)$  to  $y = c(1)$ . If  $\mathcal{M}$  is connected by rectifiable curves, then  $d_L$  is a distance.*
3. (Length space) *We say that  $(\mathcal{M}, d)$  is a length space when  $d = d_L$ .*

**LEMMA 2.3 (Length is additive and continuous).** [6, Proposition 1.20] *Let  $c : [a, b] \rightarrow \mathcal{M}$  be a rectifiable curve.*

1. *For all  $t \in [a, b]$ ,  $L(c) = L(c|_{[a,t]}) + L(c|_{[t,b]})$ .*
2. *The map  $f : t \in [a, b] \mapsto L(c|_{[a,t]})$  is non-decreasing and continuous.*

A classical example is the sphere  $\mathbb{S}^2 \subset \mathbb{R}^3$  endowed with the Euclidean distance  $d$ . The distances between the north pole  $N = (0, 0, 1)$  and the south pole  $S = (0, 0, -1)$  are  $d(N, S) = 2$  and  $d_L(N, S) = \pi$ .

DEFINITION 2.4 (Geodesics in a metric space). [31] *Let  $(\mathcal{M}, d)$  be a metric space.*

1. (Constant speed) *We say that  $c : I \rightarrow \mathcal{M}$  is a curve parameterized at constant speed when there exists  $v \geq 0$  such that for all  $t \leq t'$  in  $I$ ,  $L(c|_{[t, t']}) = v(t' - t)$ .*
2. (Unit speed) *We say that  $c : I \rightarrow \mathcal{M}$  is a curve parameterized at unit speed (or by arc length) when for all  $t \leq t'$  in  $I$ ,  $L(c|_{[t, t']}) = t' - t$ .*
3. (Globally minimizing) *We say that  $c : I \rightarrow \mathcal{M}$  is (globally) minimizing when for all  $a \leq b$  in  $I$ ,  $L(c|_{[a, b]}) = d(c(a), c(b))$ .*
4. (Locally minimizing) *We say that  $c : I \rightarrow \mathcal{M}$  is locally minimizing when for all  $t \in I$ , there exists a neighborhood  $I_0 \subseteq I$  of  $t$  such that  $c|_{I_0}$  is minimizing.*
5. (Geodesic) *A geodesic is a locally minimizing curve of constant speed. A minimizing geodesic is a globally minimizing curve of constant speed.*
6. (Geodesic space) *A geodesic (metric) space is a length space such that there exists a minimizing geodesic between any two points.*

**2.1.2. Geodesics of an affine connection.** In the following notions, the definition interval and the definition domain are of interest when the manifold is not geodesically complete and the set of preimages is of interest when the exponential map is not injective.

DEFINITION 2.5 (Geodesics of an affine connection). [31] *Let  $(\mathcal{M}, \nabla)$  be a smooth manifold equipped with an affine connection  $\nabla : \Gamma(T\mathcal{M}) \times \Gamma(T\mathcal{M}) \rightarrow \Gamma(T\mathcal{M})$ .*

1. (Geodesic) *A geodesic (or self-parallel curve) is a solution  $\gamma : I \rightarrow \mathcal{M}$  of the second-order equation  $\forall t \in I, \nabla_{\dot{\gamma}(t)} \dot{\gamma} = 0$ . The maximal solution satisfying the initial condition  $\dot{\gamma}(0) = v$  for  $v \in T_x\mathcal{M}$  is denoted  $\gamma_{(x, v)} : I_{x, v} \rightarrow \mathcal{M}$ .*
2. (Definition interval) *We call  $I_{x, v} \subseteq \mathbb{R}$  the definition interval of  $\gamma_{(x, v)}$ . It is the maximal interval of  $\mathbb{R}$  on which the geodesic  $\gamma_{(x, v)}$  is defined. In a geodesically complete manifold,  $I_{x, v} = \mathbb{R}$ .*
3. (Exponential map) *The exponential map is defined by  $\text{Exp} : v \mapsto \gamma_{(x, v)}(1)$  on the open set  $\bigcup_{x \in \mathcal{M}} \{v \in T_x\mathcal{M} \mid 1 \in I_{x, v}\} \subseteq T\mathcal{M}$ . The exponential map at  $x$ , defined by  $\text{Exp}_x(v) = \text{Exp}(v)$  for  $v \in T_x\mathcal{M}$  such that  $1 \in I_{x, v}$ , is a diffeomorphism from a neighborhood of 0 in  $T_x\mathcal{M}$  to a neighborhood of  $x$  in  $\mathcal{M}$ .*
4. (Definition domain) *The definition domain of the exponential map is  $\mathcal{D}_x = \{v \in T_x\mathcal{M} \mid 1 \in I_{x, v}\}$ .*
5. (Preimage) *A preimage of  $y \in \mathcal{M}$  by  $\text{Exp}_x$  is a vector  $v \in T_x\mathcal{M}$  such that  $1 \in I_{x, v}$  and  $\text{Exp}_x(v) = y$ . The set of preimages of  $y$  from  $x$  is denoted  $\text{Pre}_x(y) = \text{Exp}_x^{-1}(\{y\}) = \{v \in T_x\mathcal{M} \mid 1 \in I_{x, v} \text{ and } \text{Exp}_x(v) = y\}$ . Note that there might be none, one, several or infinitely many elements in  $\text{Pre}_x(y)$ .*
6. (Geodesic from  $x$  to  $y$ ) *A geodesic from  $x$  to  $y$  is a geodesic  $\gamma_{(x, v)}$  such that  $v \in \text{Pre}_x(y)$ . They are bijectively indexed by  $\text{Pre}_x(y)$  so we denote them  $\gamma_{x \rightarrow y}^v = \gamma_{(x, v)}$  for  $v \in \text{Pre}_x(y)$ . When there exists a unique preimage of  $y$  from  $x$ , we simply denote the geodesic  $\gamma_{x \rightarrow y}$ .*

**2.1.3. Geodesics of a Riemannian metric.** Any Riemannian manifold is equipped with a natural affine connection called the Levi-Civita connection. More-

over, we recall in the following definition that a connected Riemannian manifold is a metric space and even a length space.

DEFINITION 2.6 (Length, Riemannian distance). [31] *Let  $(\mathcal{M}, g)$  be a Riemannian manifold. For  $v \in T\mathcal{M}$ , we denote its norm  $\|v\| = \sqrt{g(v, v)}$ .*

1. (Length) *The length of a  $\mathcal{C}^1$  curve  $c : [a, b] \rightarrow \mathcal{M}$  is defined by  $L(c) = \int_0^1 \|\dot{c}(t)\| dt$ . This definition extends to piecewise  $\mathcal{C}^1$  curves. The length is independent from the parametrization of the curve.*
2. (Distance) *The Riemannian distance between  $x, y \in \mathcal{M}$  is defined by  $d(x, y) = \inf L(c)$  over all piecewise  $\mathcal{C}^1$  curves  $c : [0, 1] \rightarrow \mathcal{M}$  from  $x = c(0)$  to  $y = c(1)$ . If  $\mathcal{M}$  is connected, then the Riemannian distance is a distance (defining the topology of  $\mathcal{M}$ ), the Riemannian length and the metric length coincide on piecewise  $\mathcal{C}^1$  curves, and  $(\mathcal{M}, d)$  is a length space.*

Fortunately, the two notions of geodesics coincide: a piecewise  $\mathcal{C}^1$  curve is a geodesic for the Levi-Civita connection if and only if it is a geodesic for the Riemannian distance. Then, it is natural to ask what are the *globally* minimizing geodesics. A few concepts can be introduced to formalize this question: injectivity radius, cut time, cut locus, injection domain. Moreover, among the preimages  $v \in \text{Pre}_x(y)$  of the exponential map of the Levi-Civita connection, those which satisfy  $\|v\| = d(x, y)$  are called Riemannian logarithms. When there exists a unique Riemannian logarithm, the logarithm map can be defined.

DEFINITION 2.7 (Geodesics of a Riemannian metric). *Let  $(\mathcal{M}, g)$  be a connected Riemannian manifold. Let  $d$  be the Riemannian distance and  $\nabla$  be the Levi-Civita connection. We denote  $\mathcal{B}_x(\varepsilon) = \{v \in T_x\mathcal{M} \mid \|v\| < \varepsilon\} \subset T_x\mathcal{M}$  the centered open ball of  $T_x\mathcal{M}$  of radius  $\varepsilon > 0$ .*

1. (Geodesics) [31, Proposition 3.14] *A curve  $\gamma : I \rightarrow \mathcal{M}$  is a geodesic of  $(\mathcal{M}, d^g)$  (locally length-minimizing curve of constant speed) if and only if it is a geodesic of  $(\mathcal{M}, \nabla^g)$  (self-parallel curve). Such a curve is called a geodesic of  $(\mathcal{M}, g)$ .*
2. (Injectivity radius) [31] *The injectivity radius at  $x \in \mathcal{M}$  is defined by  $\text{inj}(x) = \sup \varepsilon$  over all  $\varepsilon > 0$  such that  $\exp_x$  is a diffeomorphism from  $\mathcal{B}_x(\varepsilon) \subset T_x\mathcal{M}$  to its image. The injectivity ball is  $\mathcal{B}_x(\text{inj}(x)) \subset T_x\mathcal{M}$ . For  $v \in T_x\mathcal{M}$  of norm 1 and  $t_0 \in [0, \text{inj}(x))$ , the map  $\gamma_{(x,v)} : t \in [0, t_0] \mapsto \exp_x(tv) \in \mathcal{M}$  is the unique geodesic between  $x$  and  $\gamma_{(x,v)}(t_0)$ . It is globally minimizing. The injectivity radius of  $\mathcal{M}$  is defined by  $\text{inj}(\mathcal{M}) = \inf_{x \in \mathcal{M}} \text{inj}(x)$ .*
3. (Cut time) *The cut time at  $x \in \mathcal{M}$  in the direction  $v \in T_x\mathcal{M}$ ,  $\|v\| = 1$ , is defined by  $t_{\text{cut}}(x, v) = \sup\{t \in I_{x,v} \mid d(x, \text{Exp}_x(tv)) = t\} \in (0, +\infty]$ . Note that the geodesic  $\gamma_{(x,v)}$  need not be minimizing on  $(-t_{\text{cut}}(x, -v), t_{\text{cut}}(x, v))$ .*
4. (Tangential cut locus) *The tangential cut locus at  $x$  is the set  $TCL(x) = \{t_{\text{cut}}(x, v)v \mid v \in T_x\mathcal{M}, \|v\| = 1, t_{\text{cut}}(x, v) < +\infty\}$ .*
5. (Injection domain) *The injection domain of the exponential map at  $x$  is the set  $\text{Inj}(x) = \{tv \mid t \in [0, t_{\text{cut}}(x, v)), v \in T_x\mathcal{M}, \|v\| = 1\} \subseteq \mathcal{D}_x$ . The injection ball is included in the injection domain.*
6. (Logarithms) *We call (Riemannian) logarithm of  $y \in \mathcal{M}$  from  $x \in \mathcal{M}$  a preimage  $v \in \text{Pre}_x(y) \subseteq T_x\mathcal{M}$  of  $y$  from  $x$  by the exponential map such that  $\|v\| = d(x, y)$ . We denote  $\text{Log}_x(y) \subseteq \text{Pre}_x(y)$  the set of logarithms of  $y$  from  $x$ . In particular, the geodesic  $\gamma_{x \rightarrow y}^v$  joining  $x$  to  $y$  with initial speed  $v \in \text{Log}_x$  is minimizing on  $[0, 1]$ .*



7. (*Logarithm map*) Denoting  $\mathcal{U}_x \subseteq \mathcal{M}$  the subset of points  $y$  such that there exists a unique logarithm of  $y$  from  $x$ , this defines a map  $\text{Log}_x : \mathcal{U}_x \rightarrow T_x\mathcal{M}$ . In particular,  $\mathcal{B}_x(\text{inj}(x)) \subseteq \mathcal{U}_x$  and  $\text{Log}_x : \text{Exp}_x(\mathcal{B}_x(\varepsilon)) \rightarrow \mathcal{B}_x(\varepsilon)$  is a diffeomorphism for all  $\varepsilon < \text{inj}(x)$ .

*Remark 2.8.* The cut time is usually defined in complete manifolds only [8]. Although the definition still holds, some basic results may fail in non-complete manifolds. For example, if  $I_{x,v} \neq \mathbb{R}$  and  $t_{\text{cut}}(x,v) < +\infty$ , then  $t_{\text{cut}}(x,v)$  need not belong to  $I_{x,v}$  and  $\text{Exp}_x(t_{\text{cut}}(x,v)v)$  need not be defined. Then the cut locus, which is the image of the tangential cut locus by the exponential map in complete manifolds, should be defined differently. One possible definition could simply forget the vectors  $v \in T_x\mathcal{M}$  such that  $t_{\text{cut}}(x,v) \notin I_{x,v}$  as well as the definition in complete manifolds forgets about the vectors  $v \in T_x\mathcal{M}$  such that  $t_{\text{cut}}(x,v) = +\infty$ . A maybe more satisfying definition could rely on the metric completion of the space: the cut point in direction  $v$  could be the limit of  $\text{Exp}_x(tv)$  when  $t$  tends to  $t_{\text{cut}}(x,v)$  if this limit exists.

Studying the geodesics of a space means (at least) determining precisely an expression of the geodesic  $\gamma_{(x,v)}$ , the maximal domain  $I_{x,v}$ , the injectivity radiuses  $\text{inj}(x)$  and  $\text{inj}(\mathcal{M})$ , the cut time  $t_{\text{cut}}(x,v)$ , the preimages of the exponential map  $\text{Pre}_x(y)$ , an expression of the geodesics  $\gamma_{x \rightarrow y}^v$ , the Riemannian logarithms  $\mathcal{L}og_x(y)$ , the logarithm map  $\text{Log}_x$  and its definition domain  $\mathcal{U}_x$ .

The goal of Sections 4 and 5 is to clarify and complete the knowledge on geodesics of the Bures-Wasserstein metric on the manifolds  $\text{Sym}^+(n)$  and  $\text{Sym}^+(n,k)$ .

## 2.2. Quotient spaces.

### 2.2.1. Quotient distance.

**DEFINITION 2.9** (Quotient distance). *Let  $(\mathcal{M}, d)$  be a metric space. Let  $G$  be a group acting isometrically on  $(\mathcal{M}, d)$ . We denote  $\mathcal{M}^0 = \mathcal{M}/G$  and we define the map  $d^0 : (Gx, Gy) \in \mathcal{M}^0 \times \mathcal{M}^0 \mapsto d(Gx, Gy) = \inf_{g \in G} d(gx, y) \in [0, +\infty)$ . If the orbits are closed, then  $d_0$  is a distance on  $\mathcal{M}^0$  called the quotient distance.*

*Remark 2.10.* There exists a more general notion of quotient pseudo-distance when the action is not isometric. In this work, the actions are isometric so we don't detail it.

*Remark 2.11.* The hypotheses of the definition are satisfied if the following conditions hold together:

- $G$  is a locally compact topological group (e.g. a Lie group),
- $(\mathcal{M}, d)$  is a metric space (e.g. a Riemannian manifold),
- the action of  $G$  on  $(\mathcal{M}, d)$  is continuous and isometric,
- the action is proper, i.e. (with the previous assumptions [5, III.4.4 Proposition 7]) for all  $x, y \in \mathcal{M}$ , there exist respective neighborhoods  $\mathcal{V}_x, \mathcal{V}_y$  such that the set  $K = \{g \in G | g\mathcal{V}_x \cap \mathcal{V}_y \neq \emptyset\}$  is relatively compact.

Indeed, let  $(g_n x)_{n \in \mathbb{N}}$  be a sequence in  $Gx$  tending to  $y \in \mathcal{M}$ . Then there exists  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$ ,  $g_n x \in \mathcal{V}_y$  so  $g_n \mathcal{V}_x \cap \mathcal{V}_y \neq \emptyset$ , i.e.  $g_n \in K$ . Since  $\mathcal{M}$  is a metric space and  $K$  is relatively compact, there exists a subsequence  $(g_{\varphi(n)})$  converging to  $g \in G$ . Since the action is continuous,  $g_{\varphi(n)}x$  converges to  $gx$ . Therefore,  $y = gx \in Gx$  and  $Gx$  is sequentially closed, hence closed.

**DEFINITION 2.12** (Registered points). *We say that  $x, y \in \mathcal{M}$  are registered points when  $d(x, y) = d^0(Gx, Gy)$ .*

**LEMMA 2.13** (Length in a quotient metric space). *We denote  $L$  the length on*

both  $\mathcal{M}$  and  $\mathcal{M}^0 = \mathcal{M}/G$ , and  $\pi : \mathcal{M} \rightarrow \mathcal{M}^0$  the canonical projection. For all curve  $c : [0, 1] \rightarrow \mathcal{M}$ , we have  $L(\pi \circ c) \leq L(c)$ . In particular, if  $x = c(0)$  and  $y = c(1)$  are registered and if  $c$  is minimizing, then  $\pi \circ c$  is minimizing and  $L(\pi \circ c) = L(c) = d(x, y) = d^0(\pi(x), \pi(y))$ .

*Proof.* By definition  $L(\pi \circ c) \leq \sup \sum_{k=0}^p \underbrace{d^0(\pi(c(t_k)), \pi(c(t_{k+1})))}_{\leq d(c(t_k), c(t_{k+1}))} \leq L(c)$ . If  $x = c(0)$  and  $y = c(1)$  are registered, then  $d(x, y) = d^0(\pi(x), \pi(y)) \leq L(\pi \circ c)$ . If  $c$  is minimizing, then  $d(x, y) = L(c) \geq L(\pi \circ c)$ . Thus if both hold, then  $L(\pi \circ c) = L(c) = d(x, y) = d^0(\pi(x), \pi(y))$  and  $\pi \circ c$  is minimizing.  $\square$

### 2.2.2. Quotient Riemannian metric.

DEFINITION 2.14 (Quotient metric). *Let  $(\mathcal{M}, g)$  be a Riemannian manifold. Let  $G$  be a Lie group acting smoothly, properly, freely and isometrically on  $(\mathcal{M}, g)$ . Then there exists a unique smooth structure on  $\mathcal{M}^0 = \mathcal{M}/G$  such that the quotient map  $\pi : \mathcal{M} \rightarrow \mathcal{M}^0$  is a submersion [19]. Thus, one can define for all  $x \in \mathcal{M}$ :*

1. (Vertical space)  $\mathcal{V}_x = T_x \mathcal{M}_x = \ker d_x \pi$  where  $\mathcal{M}_x = \pi^{-1}(x)$  is a submanifold of  $\mathcal{M}$ ,
2. (Horizontal space)  $\mathcal{H}_x = \mathcal{V}_x^\perp$  so that  $T_x \mathcal{M} = \mathcal{V}_x \oplus \mathcal{H}_x$ ,
3. (Horizontal lift)  $\#_x : T_{\pi(x)} \mathcal{M}^0 \rightarrow \mathcal{H}_x$  the inverse isomorphism of  $(d_x \pi)|_{\mathcal{H}_x} : \mathcal{H}_x \rightarrow T_{\pi(x)} \mathcal{M}^0$ ,
4. (Quotient metric)  $g_{\pi(x)}^0 : (v, w) \in T_{\pi(x)} \mathcal{M}^0 \times T_{\pi(x)} \mathcal{M}^0 \mapsto g_x(v_x^\#, w_x^\#) \in \mathbb{R}$ .

This is a particular case of a Riemannian submersion [30]. The Riemannian distance of the quotient metric  $g^0$  is the quotient distance of  $d$ , defined by  $d^0(\pi(x), \pi(y)) = d(Gx, Gy) = \inf_{g \in G} d(gx, y)$ .

The fundamental theorem on geodesics of a quotient metric is the following.

THEOREM 2.15 (Geodesics of a quotient metric). [30] *The projection of a horizontal geodesic is a horizontal geodesic and their lengths coincide on any segment. More precisely, let  $x \in \mathcal{M}$  and  $v \in T_{\pi(x)} \mathcal{M}^0$ . Then  $I_{x, v_x^\#} \subseteq I_{\pi(x), v}$  and for all  $t \in I_{x, v_x^\#}$ ,  $\text{Exp}_{\pi(x)}(tv) = \pi(\text{Exp}_x(tv_x^\#))$ .*

**2.2.3. Riemannian orbit spaces.** When the action of  $G$  on  $\mathcal{M}$  is not free, the orbit space  $\mathcal{M}/G$  is not a Riemannian manifold in general. This was studied in [1]. We briefly recall the main facts that we use in this work.

THEOREM 2.16 (Riemannian geometry of orbit spaces). [1] *Let  $(\mathcal{M}, g)$  be a connected complete Riemannian manifold. Let  $G$  be a Lie group acting smoothly, properly and isometrically on  $(\mathcal{M}, g)$ . We denote  $\pi : \mathcal{M} \rightarrow \mathcal{M}/G$  the canonical surjection. For a Lie subgroup  $H$  of  $G$ , we denote  $(H) = \{gHg^{-1} | g \in G\}$  the conjugacy class of  $H$  and  $\mathcal{M}_{(H)}$  the set of points  $x \in \mathcal{M}$  such that the stabilizer of  $x$ ,  $\text{Stab}(x) = \{g \in G | gx = x\}$ , belongs to  $(H)$ .*

1.  $\mathcal{M}_{(H)}$  is a smooth submanifold of  $\mathcal{M}$ .
2. We denote  $(\mathcal{M}/G)_{(H)} = \pi(\mathcal{M}_{(H)}) = \mathcal{M}_{(H)}/G$  the (isotropy) stratum of type  $(H)$ . Then  $\pi_{(H)} := \pi|_{\mathcal{M}_{(H)}} : \mathcal{M}_{(H)} \rightarrow (\mathcal{M}/G)_{(H)}$  is a smooth fiber bundle with fiber type  $G/H$ .
3. The isotropy strata form a partition of  $\mathcal{M}/G$ .
4.  $(\mathcal{M}/G, d^0)$  is a complete metric space and a length space.

The definition of the vertical space still holds while the horizontal space is replaced by the normal space [26, VI.29.2].

DEFINITION 2.17 (Vertical space, normal space). *We take the notations of Theorem 2.16. Let  $x \in \mathcal{M}_{(H)}$ . Thus  $\mathcal{M}_x = \pi^{-1}(\pi(x))$  is a submanifold of  $\mathcal{M}_{(H)}$ .*

1. (Vertical space)  $\mathcal{V}_x = T_x \mathcal{M}_x \subseteq T_x \mathcal{M}_{(H)} \subseteq T_x \mathcal{M}$ .
2. (Normal space)  $\mathcal{N}_x = \mathcal{V}_x^\perp \subseteq T_x \mathcal{M}$  so that  $T_x \mathcal{M} = \mathcal{V}_x \oplus \mathcal{N}_x$ .

Note that  $\mathcal{V}_x$  and  $\mathcal{N}_x$  need not have a constant dimension. We recall a result that we use later.

LEMMA 2.18 (Geodesics in a Riemannian orbit space). *[1, Lemma 3.5] We take the notation of Theorem 2.16. Given Lie subgroups  $H_1, H_2$  of  $G$ , we denote  $(H_1) \leq (H_2)$  if  $H_1$  is conjugated to a subgroup of  $(H_2)$ . Let  $\gamma : [0, 1] \rightarrow \mathcal{M}/G$  be a minimizing curve. For  $t \in [0, 1]$ , let  $(\mathcal{M}/G)_{(H_t)}$  denote the stratum of  $\gamma(t)$ . Then, for all  $t \in (0, 1)$ ,  $(H_t) \leq (H_0)$  and  $(H_t) \leq (H_1)$ .*

We are now well prepared to study the construction of the orbit space of covariance matrices (Section 3), the geodesics within each stratum (Sections 4 and 5) and the minimizing geodesics in the whole space (Section 6).

**3. Bures-Wasserstein geometry of covariance matrices.** We denote the set of symmetric positive semi-definite matrices or covariance matrices by  $\text{Cov}(n) = \{\Sigma \in \text{Sym}(n) \mid \Sigma \geq 0\} = \{XX^\top \mid X \in \text{Mat}(n)\} \subset \text{Mat}(n)$ . It is a complete metric subspace of the vector space of  $n \times n$  square matrices equipped with the Euclidean distance  $d^E(\Sigma, \Lambda) = \|\Sigma - \Lambda\|_2 = \text{tr}((\Sigma - \Lambda)^2)^{1/2}$ .

In this section, we recall that this set can also be described as the orbit space of the manifold  $\text{Mat}(n)$  of  $n \times n$  matrices quotiented on the right by the orthogonal group  $\text{O}(n)$ . We recall that the quotient topology coincides with the Euclidean topology and we recall the expression of the quotient distance known as the Bures-Wasserstein distance. We insist on the definition of the strata, which are investigated in the following sections.

**3.1. The quotient geometry of covariance matrices.** The **group action** of the orthogonal group  $\text{O}(n)$  on the vector space of square matrices  $\text{Mat}(n)$  is  $(X, U) \in \text{Mat}(n) \times \text{O}(n) \mapsto XU \in \text{Mat}(n)$ . It is smooth, proper and isometric for the Euclidean distance.

The **stabilizer** (or isotropy group) of a matrix  $X \in \text{Mat}(n)$  is  $\text{Stab}(X) = \{U \in \text{O}(n) \mid XU = X\}$ . If we denote  $k = \text{rk}(X)$ , it is well known that  $X$  is equivalent to the matrix  $J_k = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$ , i.e. there exist  $P, Q \in \text{GL}(n)$  such that

$X = PJ_kQ$ . Then it is clear that  $H_k := \text{Stab}(J_k) = \left\{ \begin{pmatrix} I_k & 0 \\ 0 & U \end{pmatrix} \mid U \in \text{O}(n-k) \right\}$  and

$\text{Stab}(X) = Q^{-1} \text{Stab}(J_k) Q$  with  $\dim \text{Stab}(X) = \dim \text{O}(n-k) = \frac{(n-k)(n-k-1)}{2}$ . Hence two matrices have conjugate stabilizers if and only if they have the same rank. Note that  $(H_k) \leq (H_l)$  if and only if  $k \geq l$ .

The **orbit** of  $X$  is  $\text{Orb}(X) = \{XU \mid U \in \text{O}(n)\} \simeq \text{O}(n)/\text{Stab}(X)$ . Its dimension is  $\dim \text{Orb}(X) = \dim \text{O}(n) - \dim \text{Stab}(X) = nk - \frac{k(k+1)}{2}$ . Given a result recalled in the introduction, we have  $\text{Orb}(X) = \{Y \in \text{Mat}(n) \mid YY^\top = XX^\top\}$ .

The **orbit space**  $\text{Mat}(n)/\text{O}(n) = \{\text{Orb}(X) \mid X \in \text{Mat}(n)\}$  is thus in bijection with the set of covariance matrices  $\text{Cov}(n) = \{XX^\top \mid X \in \text{Mat}(n)\}$  by the map  $\text{Orb}(X) \mapsto XX^\top$ .

The **orbit strata** of  $\text{Mat}(n)/\text{O}(n)$  are the sets of points that have conjugate stabilizers [1], i.e. that have the same rank here. This give a manifold structure to  $\text{Mat}(n)_{(H_k)} = \mathbb{R}_k^{n \times n}$ . The strata of  $\text{Mat}(n)/\text{O}(n)$  are  $\mathbb{R}_k^{n \times n}/\text{O}(n)$ , or equivalently the

strata of covariance matrices are the sets of symmetric positive semi-definite matrices of fixed rank  $\text{Sym}^+(n, k) = \text{Cov}(n) \cap \mathbb{R}_k^{n \times n}$ . The principal/regular stratum is the set of Symmetric Positive Definite (SPD) matrices  $\text{Sym}^+(n) = \text{Cov}(n) \cap \text{GL}(n)$ . Finally,  $\pi_k := \pi_{(H_k)} : \mathbb{R}_k^{n \times n} \rightarrow \text{Sym}^+(n, k)$  is a smooth fiber bundle with fiber type  $\text{St}(n, k) = \text{O}(n)/\text{O}(n-k)$ , in particular it is a submersion.

**3.2. The Bures-Wasserstein distance.** Since the group action is continuous, proper and isometric, the Euclidean distance descends to a distance on  $\text{Mat}(n)/\text{O}(n)$ . Via the bijection  $\text{Orb}(X) \in \text{Mat}(n)/\text{O}(n) \mapsto XX^\top \in \text{Cov}(n)$ , it is usually expressed as a distance on covariance matrices. It is known as the Bures-Wasserstein distance [9, 29, 3].

DEFINITION 3.1 (Bures-Wasserstein distance). [3] *The Bures-Wasserstein distance between  $\Sigma$  and  $\Lambda$  is defined by:*

$$(3.1) \quad d^{\text{BW}}(\Sigma, \Lambda) = \inf_{\substack{X, Y \in \text{Mat}(n) \\ XX^\top = \Sigma, YY^\top = \Lambda}} d^{\text{E}}(X, Y) = \inf_{R \in \text{O}(n)} d^{\text{E}}(\Sigma^{1/2}, \Lambda^{1/2}R)$$

$$(3.2) \quad = \text{tr}(\Sigma + \Lambda - 2(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2})^{1/2}.$$

If  $XX^\top = \Sigma$  and  $YY^\top = \Lambda$ , let  $R \in \text{O}(n)$  such that  $X^\top Y = (X^\top \Lambda X)^{1/2}R$ . Then  $d^{\text{E}}(X, YR^\top)^2 = \|YR^\top - X\|_2^2 = \text{tr}(XX^\top + YY^\top - 2X^\top YR^\top) = d^{\text{BW}}(\Sigma, \Lambda)^2$ .

As the quotient of a length space, the space of covariance matrices endowed with the Bures-Wasserstein metric is a length space. It is even a complete geodesic metric space [1, Proposition 3.1.(1)].

The following result seems elementary although we did not find a clear reference in the literature.

LEMMA 3.2 (Euclidean and Bures-Wasserstein topologies coincide). *The Euclidean distance  $d^{\text{E}}$  and the Bures-Wasserstein distance  $d^{\text{BW}}$  define the same topology on  $\text{Cov}(n)$ .*

See the proof of Lemma 3.2 in Appendix A.1.

**3.3. Topology, metric and smooth structure of the strata.** The set of symmetric positive definite (SPD) matrices  $\text{Sym}^+(n) = \{\Sigma \in \text{Sym}(n) \mid \text{sp}(\Sigma) \subset (0, +\infty)\}$  is an open set of the vector space of symmetric matrices, hence it has a natural structure of smooth manifold. This topology clearly coincides with the topology induced by  $(\text{Cov}(n), d^{\text{E}})$ , thus it also coincides with the topology induced by  $(\text{Cov}(n), d^{\text{BW}})$ .

The set  $\text{Sym}^+(n, k)$  is in bijection with  $\mathbb{R}_k^{n \times n}/\text{O}(n)$ . The Euclidean distance on  $\mathbb{R}_k^{n \times n}$  clearly descends to the Bures-Wasserstein distance on  $\text{Sym}^+(n, k)$ . Therefore, the quotient topology coincides with the Bures-Wasserstein topology induced by  $(\text{Cov}(n), d^{\text{BW}})$ , thus also with the Euclidean topology induced by  $(\text{Cov}(n), d^{\text{E}})$  and  $(\text{Sym}(n), d^{\text{E}})$ .

However, the Riemannian geometry is difficult to study via the submersion  $\pi_k : \mathbb{R}_k^{n \times n} \rightarrow \text{Sym}^+(n, k)$ . Indeed, it is the projection of a bundle of fiber  $\text{St}(n, k) \simeq \text{O}(n)/\text{O}(n-k)$  which is not a Lie group. Fortunately, the set  $\text{Sym}^+(n, k)$  is also in bijection with the quotient manifold  $\mathbb{R}_*^{n \times k}/\text{O}(k)$  [24, Proposition 2.1], where  $\mathbb{R}_*^{n \times k}$  is the open set of matrices of full rank in  $\mathbb{R}^{n \times k}$ . We recall this quotient geometry in Table 3.1. The quotient distance induced on  $\text{Sym}^+(n, k)$  is the Bures-Wasserstein distance [24, Proposition 5.1]. In particular, the quotient topology coincides with the previous ones. This bijection naturally provides a smooth structure on  $\text{Sym}^+(n, k)$ . Above all, the submersion  $\tilde{\pi}_k : \mathbb{R}_*^{n \times r} \rightarrow \text{Sym}^+(n, k)$  is the projection of a principal fiber bundle. Hence, it is much more convenient to study the Bures-Wasserstein

Riemannian geometry of  $\text{Sym}^+(n, k)$  via the Riemannian submersion  $\tilde{\pi}_k$ . This is exactly what is done in [24].

To summarize, the strata  $\text{Sym}^+(n, k)$  are smooth connected manifolds and the regular stratum  $\text{Sym}^+(n)$  is a dense open set in  $\text{Cov}(n)$ .

Set	$\text{Sym}^+(n, k) = \text{Cov}(n) \cap \mathbb{R}_k^{n \times n}$
Smooth manifold	$\mathbb{R}_*^{n \times k} / \text{O}(k)$
Group action	$\begin{cases} \mathbb{R}_*^{n \times k} \times \text{O}(k) & \longrightarrow \mathbb{R}_*^{n \times k} \\ (X, U) & \longmapsto XU \end{cases}$
Orbit	$\text{Orb}(X) = \{Y \in \mathbb{R}_*^{n \times k}   YY^\top = XX^\top\}$
Identification	$\begin{cases} \mathbb{R}_*^{n \times k} / \text{O}(k) & \longrightarrow \text{Sym}^+(n, k) \\ \text{Orb}(X) & \longmapsto XX^\top \end{cases}$
Submersion	$\pi_{\text{Sym}^+(n, k)} : \begin{cases} \mathbb{R}_*^{n \times k} & \longrightarrow \text{Sym}^+(n, k) \\ X & \longmapsto XX^\top \end{cases}$

Table 3.1: Smooth manifold structure of  $\text{Sym}^+(n, k)$ .

The Riemannian geometry of the principal stratum was extensively studied [36, 37, 23, 3, 41, 39] and the Riemannian geometry of the other strata was recently well detailed [25, 24]. However, there remain missing formulae and open questions about the geodesics in each stratum: injection domain, preimages, and explicit formulae of the horizontal lift, the exponential map and logarithms in the base space  $\text{Sym}^+(n, k)$  of the principal fiber bundle  $\mathbb{R}_*^{n \times k} \longrightarrow \text{Sym}^+(n, k)$ . We precisely answer these questions in Section 4 (full-rank matrices) and Section 5 (low-rank matrices). Furthermore, we contribute in Section 6 the minimizing geodesics for the Bures-Wasserstein *distance* between different strata and the condition of uniqueness of the geodesic between two points.

**4. Geodesics of the Bures-Wasserstein metric on  $\text{Sym}^+(n)$ .** In this section, we give complements and new results on the Bures-Wasserstein geodesics on SPD matrices  $\text{Sym}^+(n)$ . The quotient structure is well known, as well as the exponential map [23] and the injectivity radius [24]. The definition interval of the geodesic was implicitly described in [23] as the connected component of 0 in a subset of  $\mathbb{R}$  so we give it explicitly here. It was proved in [3] that there exists a preimage which is a logarithm. We prove the uniqueness of the preimage and the logarithm based on a result of [24] on  $\text{Sym}^+(n, k)$  applied for  $k = n$ . Moreover, we contribute the cut time, thus the injection domain. After Theorem 4.2 on Bures-Wasserstein geodesics, we show on an example that we already know some geodesics between degenerate matrices that cross the main stratum of SPD matrices.

**DEFINITION 4.1** (Bures-Wasserstein metric on  $\text{Sym}^+(n)$ ). [23, 3, 41] *The Bures-Wasserstein metric on  $\text{Sym}^+(n)$  is the quotient Riemannian metric induced by the submersion  $\pi : X \in \text{GL}(n) \longmapsto XX^\top \in \text{Sym}^+(n)$  and the Frobenius metric on  $\text{GL}(n)$ . Let  $X \in \text{GL}(n)$  such that  $XX^\top = \Sigma$  and let  $V \in T_\Sigma \text{Sym}^+(n) \equiv \text{Sym}(n)$ . The quotient operations are:*

1. (Vertical space)  $\mathcal{V}_X = \ker d_X \pi = X \text{Skew}(n)$ ,
2. (Horizontal space)  $\mathcal{H}_X = \text{Sym}(n) X$ ,
3. (Horizontal lift)  $V_X^\# = \mathcal{S}_\Sigma(V) X \in \mathcal{H}_X$ ,
4. (Bures-Wasserstein metric)  $g_\Sigma^{BW(n)}(V, V) = \text{tr}(\mathcal{S}_\Sigma(V) \Sigma \mathcal{S}_\Sigma(V))$ ,

where  $\mathcal{S}_\Sigma(V) \in \text{Sym}(n)$  is the unique solution of the Sylvester equation  $\Sigma \mathcal{S}_\Sigma(V) + \mathcal{S}_\Sigma(V) \Sigma = V$ .

**THEOREM 4.2** (Bures-Wasserstein geodesics on  $\text{Sym}^+(n)$ ). *Let  $\Sigma \in \text{Sym}^+(n)$ .*

1. (Exponential map) [23] *For all  $V \in T_\Sigma \text{Sym}^+(n) \equiv \text{Sym}(n)$ , the geodesic from  $\Sigma$  with initial speed  $V$  writes  $\gamma_{(\Sigma, V)}(t) = \Sigma + tV + t^2 \mathcal{S}_\Sigma(V) \Sigma \mathcal{S}_\Sigma(V) \in \text{Sym}^+(n)$ .*
2. (Definition interval) *Let  $\lambda_{\max} = \max \text{sp}(\mathcal{S}_\Sigma(V))$  and  $\lambda_{\min} = \min \text{sp}(\mathcal{S}_\Sigma(V))$ . The definition interval of the geodesic  $\gamma_{(\Sigma, V)}$  is the interval  $I_{\Sigma, V}$  defined by:*
  - $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, -\frac{1}{\lambda_{\min}})$  if  $\lambda_{\min} < 0 < \lambda_{\max}$ ,
  - $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_{\min}})$  if  $\lambda_{\min} < 0$  and  $\lambda_{\max} \leq 0$ ,
  - $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, +\infty)$  if  $\lambda_{\min} \geq 0$  and  $\lambda_{\max} > 0$ ,
  - $I_{\Sigma, V} = \mathbb{R}$  if  $\lambda_{\min} = \lambda_{\max} = 0$  (which only happens for  $V = 0$ ).
3. (Cut time) *The cut time is  $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$  if  $\lambda_{\min} < 0$  or  $+\infty$  otherwise. The geodesic  $\gamma_{(\Sigma, V)} : I_{\Sigma, V} \rightarrow \mathcal{M}$  is even minimizing on  $I_{\Sigma, V}$ .*
4. (Logarithm map) *For all  $\Lambda \in \text{Sym}^+(n)$ , there exists a unique preimage  $V \in \text{Pre}_\Sigma(\Lambda)$ . It writes  $V = 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}) - 2\Sigma$ , where we denote  $\text{sym}(M) = \frac{1}{2}(M + M^\top)$ . The geodesic joining  $\Sigma$  to  $\Lambda$  writes:*

$$(4.1) \quad \gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}).$$

Moreover, it is a logarithm:  $V \in \mathcal{L}og_x(y)$ . Thus the logarithm map is defined on  $\mathcal{U}_\Sigma = \text{Sym}^+(n)$  and it writes:

$$(4.2) \quad \text{Log}_\Sigma : \begin{cases} \text{Sym}^+(n) & \longrightarrow \\ \Lambda & \longmapsto 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}) - 2\Sigma \end{cases} .$$

See the proof of Theorem 4.2 in Appendix A.2.

*Remark 4.3.* The minimizing geodesic  $\gamma_{(\Sigma, V)} : I_{\Sigma, V} \rightarrow \text{Sym}^+(n)$  clearly has a limit at the finite boundaries of  $I_{\Sigma, V}$ . When  $I_{\Sigma, V}$  is bounded, we can define  $\Sigma_0 = \lim_{t \rightarrow -1/\lambda_{\max}} \gamma_{(\Sigma, V)}(t)$ ,  $\Sigma_1 = \lim_{t \rightarrow -1/\lambda_{\min}} \gamma_{(\Sigma, V)}(t)$  and the extended curve  $\gamma : \bar{I}_{\Sigma, V} = [-1/\lambda_{\max}, -1/\lambda_{\min}]$  by  $\gamma(t) = \gamma_{(\Sigma, V)}(t)$  for  $t \in I_{\Sigma, V}$ ,  $\gamma(-1/\lambda_{\max}) = \Sigma_0$  and  $\gamma(-1/\lambda_{\min}) = \Sigma_1$ . The curve  $\gamma$  is a minimizing geodesic on  $I_{\Sigma, V}$ . Thus, by Lemma 2.3 (continuity of the length), the curve  $\gamma$  is a minimizing geodesic on  $[-1/\lambda_{\max}, -1/\lambda_{\min}]$ . So we already have examples of minimizing geodesics between two degenerate matrices which pass through the principal stratum  $\text{Sym}^+(n)$ . For instance, we have  $\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$  and  $\Lambda = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$  which are linked by the geodesic

$$\text{Exp}_{I_2}(tV) = \begin{pmatrix} (1+t)^2 & 0 \\ 0 & (1-t)^2 \end{pmatrix} \text{ for } t \in (-1, 1) \text{ with } V = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Moreover, from the viewpoint of the Levi-Civita connection, the curve  $t \mapsto \text{Exp}_\Sigma(tV)$  is a geodesic (self-parallel curve) on each subinterval of the set  $J_{\Sigma, V} = \{t \in \mathbb{R} \mid \text{Exp}_\Sigma(tV) \in \text{Sym}^+(n)\} = \mathbb{R} \setminus \{t \in \mathbb{R} \mid -\frac{1}{t} \in \text{sp}(\mathcal{S}_\Sigma(V))\}$ , which is  $\mathbb{R}$  without a maximum of  $n$  points. Theorem 4.2 actually states that every geodesic is minimizing on its domain. Hence, the minimizing geodesic  $\gamma_{(\Sigma, V)} : I_{\Sigma, V} \rightarrow \text{Sym}^+(n)$  naturally extends to a curve  $\gamma_{(\Sigma, V)} : \mathbb{R} \rightarrow \text{Cov}(n)$  which is a minimizing geodesic on the segments delimited by two consecutive values in  $\mathbb{R} \setminus J_{\Sigma, V}$ .

**5. Geodesics of the Bures-Wasserstein metric on  $\text{Sym}^+(n, k)$ .** In this section, we give complements and new results on the Bures-Wasserstein geodesics on the manifold of PSD matrices of fixed rank  $k$ ,  $\text{Sym}^+(n, k)$ . They were mainly studied in

[24]. The formulae of the exponential map and its definition domain were kept implicit because they were formulated in function of horizontal vectors in the total space  $\mathbb{R}_*^{n \times k}$  of matrices of full-rank  $k$ . We compute the horizontal lift, which allows us to express the Bures-Wasserstein metric, the exponential map and the definition interval directly in function of the tangent vector. Moreover, the characterization of preimages of the exponential map given in [24] seems to omit the condition that a preimage  $v \in \mathcal{P}re_x(y)$  must satisfy  $1 \in I_{x,v}$ . Otherwise, the geodesic  $\gamma_{x,v} : I_{x,v} \rightarrow \mathcal{M}$  could leave the space before reaching  $y$ . Therefore, we characterize the set of preimages by relying on their work and considering this additional condition. We also give an explicit formula of the minimizing geodesic joining two points when it is unique and we specify the number of minimizing geodesics between two points otherwise. Finally, we compute the injection domain that was kept implicit in [24]. After Theorem 5.3, we precisely specify the novelty of our result with respect to the reference work [24]. Then, we give examples to illustrate the possible cases for the number of preimages and logarithms.

**DEFINITION 5.1** (Bures-Wasserstein metric on  $\text{Sym}^+(n, k)$ ). [24] *The Bures-Wasserstein metric on  $\text{Sym}^+(n, k)$  is the quotient Riemannian metric induced by the submersion  $\pi : X \in \mathbb{R}_*^{n \times k} \mapsto XX^\top \in \text{Sym}^+(n, k)$  and the Frobenius metric on  $\mathbb{R}^{n \times k}$ . Let  $X \in \mathbb{R}_*^{n \times k}$  such that  $XX^\top = \Sigma$  and let  $V \in T_\Sigma \text{Sym}^+(n, k)$ . The vertical and horizontal spaces are:*

1. (Vertical space)  $\mathcal{V}_X = \ker d_X \pi = X \text{Skew}(k)$ ,
2. (Horizontal space)  $\mathcal{H}_X = \{X(X^\top X)^{-1}F + X_\perp K, F \in \text{Sym}(k), K \in \text{Mat}(n - k, k)\}$  where  $X_\perp \in \text{Mat}(n, n - k)$  has orthonormal columns ( $X_\perp^\top X_\perp = I_{n-k}$ ) that are orthogonal to the columns of  $X$  ( $X^\top X_\perp = 0$ ).

**THEOREM 5.2** (Horizontal lift, tangent space, metric). *Let  $\Sigma \in \text{Sym}^+(n, k)$ , let  $X \in \mathbb{R}_*^{n \times k}$  such that  $\Sigma = XX^\top \in \text{Sym}^+(n, k)$  and let  $V \in T_\Sigma \text{Sym}^+(n, k)$ . Let  $\Sigma = UDU^\top$  be a singular value decomposition with  $D \in \text{Diag}^+(k)$  and  $U \in \text{St}(n, k)$ . We denote  $S = S_{\Sigma, V} = US_D(U^\top VU)U^\top$ , where  $\mathcal{S}_A(B)$  denotes the unique solution  $Z$  of the Sylvester equation  $AZ + ZA = B$ . Note that  $S_{\Sigma, V}$  and  $(I_n - UU^\top)$  are independent from the chosen decomposition.*

1. (Tangent space)  $T_\Sigma \text{Sym}^+(n, k) = \{V \in \text{Sym}(n) | X_\perp^\top V X_\perp = 0\}$ ,
2. (Horizontal lift)  $V_X^\# = X(X^\top X)^{-1} \mathcal{S}_{X^\top X}(X^\top V X) + X_\perp X_\perp^\top V X (X^\top X)^{-1}$ , where  $X_\perp X_\perp^\top = I_n - X(X^\top X)^{-1} X^\top$ ,
3. (Bures-Wasserstein metric)  $g_\Sigma^{BW(n, k)}(V, V) = \text{tr}(S_{\Sigma, V} \Sigma S_{\Sigma, V} + V \Sigma^{-1} V (I_n - UU^\top))$ .

See the proof of Theorem 5.2 in Appendix A.3.

**THEOREM 5.3** (Bures-Wasserstein geodesics on  $\text{Sym}^+(n, k)$ ). *Let  $\Sigma, \Lambda \in \text{Sym}^+(n, k)$  and  $X, Y \in \mathbb{R}_*^{n \times k}$  such that  $XX^\top = \Sigma$  and  $YY^\top = \Lambda$ . Let  $U \in \text{St}(n, k)$  and  $D \in \text{Diag}^+(k)$  such that  $\Sigma = UDU^\top$ .*

1. (Exponential map) *For all  $V \in T_\Sigma \text{Sym}^+(n, k)$ , the geodesic from  $\Sigma$  with initial speed  $V$  is  $\gamma_{(\Sigma, V)} : t \in I_{\Sigma, V} \mapsto \Sigma + tV + t^2 W_{\Sigma, V}$ , where  $W_{\Sigma, V} = S_{\Sigma, V} \Sigma S_{\Sigma, V} + S_{\Sigma, V} V (I_n - UU^\top) + (I_n - UU^\top) V S_{\Sigma, V} + (I_n - UU^\top) V \Sigma^{-1} V (I_n - UU^\top)$  and  $S_{\Sigma, V} = US_D(U^\top VU)U^\top$ .*
2. (Definition interval) *Let  $F_{X, V}^0 = \mathcal{S}_{X^\top X}((X^\top X)^{-1/2} X^\top V X (X^\top X)^{-1/2})$  and  $M_{X, V}^0 = (X^\top X)^{-3/2} X^\top V (I_n - X(X^\top X)^{-1} X^\top) V X (X^\top X)^{-3/2} \in \text{Sym}(n)$ . Let  $\mathcal{E}_{\Sigma, V} = \{\lambda \in \text{sp}(F_{X, V}^0) | \ker(\lambda I_k - F_{X, V}^0) \cap \ker(M_{X, V}^0) \neq \{0\}\} \subseteq \text{sp}(S_{\Sigma, V})$ . If  $\mathcal{E}_{\Sigma, V}$  is non-empty, then let  $\lambda_+ = \max \mathcal{E}_{\Sigma, V}$  and  $\lambda_- = \min \mathcal{E}_{\Sigma, V}$ . The definition interval of the geodesic  $\gamma_{(\Sigma, V)}$  is the interval  $I_{\Sigma, V}$  defined by:*

- $I_{\Sigma, V} = (-\frac{1}{\lambda_+}, -\frac{1}{\lambda_-})$  if  $\lambda_- < 0 < \lambda_+$ ,
- $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_-})$  if  $\lambda_- < 0$  and  $\lambda_+ \leq 0$ ,
- $I_{\Sigma, V} = (-\frac{1}{\lambda_+}, +\infty)$  if  $\lambda_- \geq 0$  and  $\lambda_+ > 0$ ,
- $I_{\Sigma, V} = \mathbb{R}$  if  $\mathcal{E}_{\Sigma, V}$  is empty.

Applying this to  $X = UD^{1/2}$  without loss of generality,  $F_{X, V}^0 = \mathcal{S}_D(U^\top VU)$  and  $M_{X, V}^0 = D^{-1}U^\top V(I_n - UU^\top)VUD^{-1}$  which is a bit more tractable to compute  $\mathcal{E}_{\Sigma, V}$ .

3. (Cut time) Let  $\lambda_{\max} = \max \text{sp}(F_{X, V}^0)$  and  $\lambda_{\min} = \min \text{sp}(F_{X, V}^0)$ . Note that if  $\mathcal{E}_{\Sigma, V} \neq \emptyset$ , then we have  $(\lambda_-, \lambda_+) \subseteq (\lambda_{\min}, \lambda_{\max})$ . The cut time is  $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$  if  $\lambda_{\min} < 0$  or  $+\infty$  otherwise. Symmetrically, we have  $t_{\text{cut}}(\Sigma, -V) = \frac{1}{\lambda_{\max}}$  if  $\lambda_{\max} > 0$  or  $+\infty$  otherwise.
4. (Preimages) We define the indexing set  $\mathcal{I}_{X, Y}^{\text{Pre}}$  by:

$$\begin{aligned} \mathcal{I}_{X, Y}^{\text{Pre}} &= \{R \in \text{O}(n) \mid H := X^\top Y R^\top \in \text{Sym}(n) \text{ and} \\ &\quad \forall \mu < 0, \ker(\mu I_k - (X^\top X)^{-1/2} H (X^\top X)^{-1/2}) \\ &\quad \cap \ker(\mu^2 I_k - (X^\top X)^{-1/2} R Y^\top Y R^\top (X^\top X)^{-1/2}) = \{0\}\}. \end{aligned}$$

For  $R \in \mathcal{I}_{X, Y}^{\text{Pre}}$ , we denote  $H = H_{X, Y, R} = X^\top Y R^\top$  so that  $X^\top Y = HR$ . Then the map  $R \in \mathcal{I}_{X, Y}^{\text{Pre}} \mapsto V = 2 \text{sym}(XRY^\top) - 2\Sigma \in \text{Pre}_\Sigma(\Lambda)$  is a bijection whose inverse is  $V \in \text{Pre}_\Sigma(\Lambda) \mapsto R = (Y^\top Y)^{-1} Y^\top (X + V_X^\#) \in \mathcal{I}_{X, Y}^{\text{Pre}}$ . The geodesic joining  $\Sigma$  to  $\Lambda$  parametrized by  $R \in \mathcal{I}_{X, Y}^{\text{Pre}}$  writes:

$$(5.1) \quad \forall t \in [0, 1], \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(XRY^\top).$$

5. (Logarithms) Let  $\mathcal{I}_{X, Y}^{\text{Log}} = \{R \in \text{O}(n) \mid H_{X, Y, R} = X^\top Y R^\top \in \text{Cov}(n)\} = \{R \in \text{O}(n) \mid H_{X, Y, R} = (X^\top \Lambda X)^{1/2}\} = \{R \in \text{O}(n) \mid X^\top Y = (X^\top \Lambda X)^{1/2} R\} \subseteq \mathcal{I}_{X, Y}^{\text{Pre}}$ . Then, the map  $R \in \mathcal{I}_{X, Y}^{\text{Log}} \mapsto V = 2 \text{sym}(XRY^\top) - 2\Sigma \in \text{Log}_\Sigma(\Lambda)$  is a bijection whose inverse is  $V \in \text{Log}_\Sigma(\Lambda) \mapsto R = (Y^\top Y)^{-1} Y^\top (X + V_X^\#) \in \mathcal{I}_{X, Y}^{\text{Log}}$ .
6. (Logarithm map) Let  $r = \text{rk}(\Sigma \Lambda) = \text{rk}(X^\top Y) = \text{rk}(H)$ .

- (a) If  $r = k$ , then there exists a unique logarithm of  $\Lambda$  from  $\Sigma$ . In this case, the minimizing geodesic joining  $\Sigma$  to  $\Lambda$  writes:

$$(5.2) \quad \gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2} ((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda).$$

- (b) If  $r = k - 1$ , then there exist exactly two logarithms of  $\Lambda$  from  $\Sigma$ .

- (c) If  $r < k - 1$ , then there is an infinity of logarithms of  $\Lambda$  from  $\Sigma$ .

Therefore, the logarithm map is defined on  $\mathcal{U}_\Sigma = \{\Lambda \in \text{Sym}^+(n, k) \mid \text{rk}(\Sigma \Lambda) = k\}$  and it writes  $\text{Log}_\Sigma : \Lambda \in \mathcal{U}_\Sigma \mapsto 2 \text{sym}(\Sigma^{1/2} ((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda) - 2\Sigma \in T_\Sigma \text{Sym}^+(n, k)$ .

See the proof of Theorem 5.3 in Appendix A.4.

*Remark 5.4.* Let us clarify our contributions in Theorem 5.3 with respect to the reference paper [24].

1. (Exponential map) The formula is new. Only the exponential map of a horizontal vector in the total space  $\mathbb{R}_*^{n \times k}$  was given in [24].
2. (Definition interval) The definition interval was formulated in the total space  $\mathbb{R}_*^{n \times k}$  in [24]. The novelty here is to formulate it in function of  $V$  thanks to the horizontal lift.



3. (Cut time) This is new.
4. (Preimages) Note that  $\mathcal{I}_{X,Y}^{Pre} \subseteq \mathcal{I}_{X,Y}^{Sol} := \{R \in O(n) | X^\top Y R^\top \in \text{Sym}(n)\}$ . The set  $\mathcal{I}_{X,Y}^{Sol}$  indexes the solutions of the logarithm equation  $\text{Exp}_\Sigma(V) = \Lambda$  without taking into account the condition  $1 \in I_{\Sigma,V}$ . The fact that there may exist  $R \in \mathcal{I}_{X,Y}^{Sol}$  that does not lead to a preimage seems to be forgotten in [24]. Indeed, the curve “ $\gamma^R$ ” may hit the boundary before reaching  $\Lambda$ . Therefore, among these candidate  $R$ 's such that  $X^\top Y R^\top \in \text{Sym}(n)$ , we specify the set of  $R$ 's that really define a geodesic from  $\Sigma$  to  $\Lambda$  in the manifold  $\text{Sym}^+(n, k)$ , based on the condition underlying the definition interval.
5. (Logarithms) It was stated in [24] that if  $R$  leads to a logarithm then  $H \geq 0$ , and that if  $H \geq 0$ , then  $R$  leads to a preimage which is additionally a logarithm. However, it is not stated clearly that in this case,  $H$  has to be equal to  $(X^\top \Lambda X)^{1/2}$ . It is important for the next point though. The expression of the logarithms is new, although very straightforward. It was not given in [24] because they prefer to work in the total space  $\mathbb{R}_*^{n \times k}$ .
6. (Logarithm map) It was stated that the logarithm is unique if and only if  $r = k$  in [24]. However, it was not stated that there are exactly two logarithms when  $r = k - 1$  and that there is an infinity of *logarithms* when  $r < k - 1$ . The expression of the minimizing geodesic when it is unique is also new.

In other words, we have a minor contribution on the reformulation in  $\text{Sym}^+(n, k)$  of results stated in the total space, and more important contributions on the injection domain (cut time), the expression of the minimizing geodesic when it is unique and the clarification between the three sets  $\mathcal{I}_{X,Y}^{Log} \subseteq \mathcal{I}_{X,Y}^{Pre} \subseteq \mathcal{I}_{X,Y}^{Sol} := \{R \in O(n) | X^\top Y R^\top \in \text{Sym}(n)\}$ . We give several examples below to illustrate the differences between these three sets.

*Remark 5.5.* The definitions of the sets  $\mathcal{E}_{\Sigma,V}$  and  $\mathcal{I}_{\Sigma,V}^{Pre}$  might seem intricate. It is difficult to simplify them though. Nevertheless, the set  $\mathcal{E}_{\Sigma,V}$  is easy to determine numerically. Analogously, it is easy to determine numerically if a candidate  $R \in \mathcal{I}_{X,Y}^{Sol}$  given in [24, Lemma 4.1] belongs to  $\mathcal{I}_{X,Y}^{Pre}$ .

EXAMPLES 5.6. Let us denote  $R_\theta^- = \begin{pmatrix} -\cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \in O(2) \cap \text{Sym}(2)$ .

1. Let  $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$  and  $\Lambda = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 4\Sigma$  in  $\text{Sym}^+(3, 2)$  with  $r = 2$ . Then, let  $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$  and  $Y = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}$ . Then  $XX^\top = \Sigma$ ,  $YY^\top = \Lambda$  and  $X^\top Y = 2I_2$ . Thus the candidate pairs for  $(H, R)$  are  $(2I_2, I_2)$ ,  $(-2I_2, -I_2)$  and  $(2R_\theta^-, R_\theta^-)$ . One can show that:
  - (a)  $R = I_2$  leads to the unique minimizing geodesic  $\gamma_{\Sigma \rightarrow \Lambda}(t) = (1+t)^2 \Sigma \in \text{Sym}^+(3, 2)$  for  $t \in (-1, +\infty) \supset [0, 1]$ ,
  - (b) there is no non-minimizing geodesic,
  - (c)  $R = -I_2$  or  $R = R_\theta^-$  lead to curves that hit  $\text{Sym}^+(3, 1)$  at  $t = \frac{1}{3} < 1$ , e.g.  $R = -I_2$  leads to the curve  $\gamma(t) = (1-3t)^2 \Sigma \in \text{Sym}^+(3, 2)$  only for  $t \in (-\infty, \frac{1}{3})$ .
2. Let  $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$  and  $\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$  in  $\text{Sym}^+(3, 2)$  with  $r = 2$  again. Then, let  $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$  and  $Y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$ . Then  $X^\top Y = I_2$  so the candidate pairs for  $(H, R)$  are  $(I_2, I_2)$ ,  $(-I_2, -I_2)$  and  $(R_\theta^-, R_\theta^-)$ . One can show that:
  - (a)  $R = I_2$  leads to the geodesic  $\gamma_{\Sigma \rightarrow \Lambda}^{I_2}(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & t & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$  for  $t \in \mathbb{R}$  which is minimizing on  $[-\frac{2}{1+\sqrt{2}}, \frac{2}{\sqrt{2}-1}] \supset [0, 1]$ ,

- (b)  $R = R_\theta^-$  for  $\theta \neq 0$  lead to non-minimizing geodesics, e.g.  $R = R_\pi^-$  leads to the curve  $\gamma_{\Sigma \rightarrow \Lambda}^{R_\pi^-}(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (1-2t)^2 & -t(1-2t) \\ 0 & -t(1-2t) & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$  for  $t \in [0, 1]$ ,
  - (c)  $R = R_0^-$  and  $R = -I_2$  lead to curves that hit  $\text{Sym}^+(3, 1)$  at  $t = \frac{1}{2} < 1$ , e.g.  $R = R_0^-$  leads to the curve  $\gamma(t) = \begin{pmatrix} (1-2t)^2 & 0 & 0 \\ 0 & 1 & t \\ 0 & t & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$  only for  $t \in (\infty, \frac{1}{2})$ .
3. Let  $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$  and  $\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  in  $\text{Sym}^+(3, 2)$  with  $r = 1$ . Then, let  $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  and  $Y = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$ . Then  $X^\top Y = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$  so the candidate values of  $R$  are  $\text{Diag}(\pm 1, \pm 1)$ . One can show that:
- (a)  $R_\pm^0 = \text{Diag}(1, \pm 1)$  lead to two minimizing geodesics whose expressions are  $\gamma_{\Sigma \rightarrow \Lambda}^{R_\pm^0}(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (1-t)^2 & \pm t(1-t) \\ 0 & \pm t(1-t) & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$  for  $t \in [0, 1]$ ,
  - (b) there is no non-minimizing geodesic,
  - (c)  $R_\pm^1 = \text{Diag}(-1, \pm 1)$  lead to curves that hit  $\text{Sym}^+(3, 1)$  at  $t = \frac{1}{2} < 1$ , namely  $\gamma^{R_\pm^1}(t) = \begin{pmatrix} (1-2t)^2 & 0 & 0 \\ 0 & (1-t)^2 & \pm t(1-t) \\ 0 & \pm t(1-t) & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$  only for  $t \in (-\infty, \frac{1}{2})$ .
4. Let  $\Sigma = \begin{pmatrix} I_2 & 0 \\ 0 & 0 \end{pmatrix}$  and  $\Lambda = \begin{pmatrix} 0 & 0 \\ 0 & I_2 \end{pmatrix}$  in  $\text{Sym}^+(4, 2)$  with  $r = 0$ . Then, let  $X = \begin{pmatrix} I_2 \\ 0 \end{pmatrix}$  and  $Y = \begin{pmatrix} 0 \\ I_2 \end{pmatrix}$ . Then  $X^\top Y = 0$  so every  $R \in \text{O}(2)$  is a candidate. One can show that any  $R \in \text{O}(2)$  leads to a minimizing geodesic  $\gamma_{\Sigma \rightarrow \Lambda}^R(t) = \begin{pmatrix} (1-t)^2 I_2 & t(1-t)R \\ t(1-t)R^\top & t^2 I_2 \end{pmatrix} \in \text{Sym}^+(4, 2)$  for  $t \in \mathbb{R}$ .

In the two last sections, we studied the geodesics and the minimizing geodesics within each stratum. In the next section, we turn to the study of the minimizing geodesic segments in the Bures-Wasserstein metric space  $(\text{Cov}(n), d^{\text{BW}})$ , that is between any two covariance matrices of any rank.

### 6. Minimizing geodesics of the Bures-Wasserstein distance on $\text{Cov}(n)$ .

In this section, we completely characterize the Bures-Wasserstein minimizing geodesic segments between any two covariance matrices. We show that they have constant rank on the interior of the segment and we give an explicit expression. Moreover, we show that the number of geodesics depends on the ranks of the extremities and we give this number in all cases. More precisely, we show that minimizing geodesics between  $\Sigma$  and  $\Lambda \in \text{Cov}(n)$  are parametrized by the closed unit ball of  $\mathbb{R}^{(k-r) \times (l-r)}$  for the spectral norm, where  $k, l, r$  are the respective ranks of  $\Sigma, \Lambda, \Sigma\Lambda$ . We also give the number of geodesics of minimal rank. Finally, we show that there exists a canonical geodesic with an expression that does not depend on the ranks of the extremities. This expression coincides with the formula in low rank when the minimizing geodesic is unique and with the formula in full rank. The proofs are deferred to the Supplementary Material.

**6.1. Characterization of minimizing geodesics.** The following lemma states that the rank of a minimizing geodesic segment is constant on the interior of the segment. Then, Theorem 6.2 characterizes the Bures-Wasserstein minimizing geodesic segments.

**LEMMA 6.1** (Rank of minimizing curve). *Let  $\gamma : [0, 1] \rightarrow \text{Cov}(n)$  be a minimizing curve from  $\Sigma$  to  $\Lambda$ . Then  $\gamma$  has constant rank  $p \geq \max(\text{rk}(\Sigma), \text{rk}(\Lambda))$  on  $(0, 1)$ .*

*Proof.* Let  $p = \max_{t \in [0,1]} \text{rk}(\gamma(t))$  and let  $t_0 \in [0,1]$  such that  $\text{rk}(t_0) = p$ . By Lemma 2.18, for all  $t \in (0, t_0) \cup (t_0, 1)$ ,  $(H_{\text{rk}(\gamma(t))}) \leq H_{\text{rk}(\gamma(t_0))}$  so  $\text{rk}(\gamma(t)) \geq \text{rk}(\gamma(t_0)) = p$  so  $\text{rk}(\gamma(t)) = p$ .  $\square$

**THEOREM 6.2** (Bures-Wasserstein minimizing geodesics in  $\text{Cov}(n)$ ). *Let  $\Sigma, \Lambda \in \text{Cov}(n)$  with  $\text{rk}(\Sigma) = k$  and  $\text{rk}(\Lambda) = l$ . Let  $X, Y \in \mathbb{R}^{n \times n}$  such that  $XX^\top = \Sigma$  and  $YY^\top = \Lambda$ . The two following statements are equivalent:*

- (i) *the curve  $\gamma : [0, 1] \rightarrow \text{Cov}(n)$  is a minimizing geodesic segment from  $\Sigma$  to  $\Lambda$ ,*
- (ii) *there exists  $R \in O(n)$  such that  $H_{X,Y,R} := X^\top Y R^\top \in \text{Cov}(n)$  and for all  $t \in [0, 1]$ ,  $\gamma(t) = \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(XRY^\top)$ .*

*Moreover,  $H_{X,Y,R} = (X^\top \Lambda X)^{1/2}$  and the minimizing geodesic  $\gamma_{\Sigma \rightarrow \Lambda}^R$  is of constant rank  $p \geq \max(k, l)$  on  $(0, 1)$ .*

See the proof of Theorem 6.2 in Appendix A.5.

**6.2. Number of minimizing geodesics.** In this section, we count the number of minimizing geodesic segments between two covariance matrices. We start with an elementary lemma.

**LEMMA 6.3** (Elementary algebra). *Let  $\Sigma, \Lambda \in \text{Cov}(n)$  with  $\text{rk}(\Sigma) = k$  and  $\text{rk}(\Lambda) = l$ . Let  $r = \text{rk}(\Sigma\Lambda)$ .*

- 1. *For all  $X, Y \in \text{Mat}(n)$  such that  $XX^\top = \Sigma$  and  $YY^\top = \Lambda$ ,  $r = \text{rk}(X^\top Y)$ .*
- 2. *We have  $l - r \leq n - k$ .*

See the proof of Lemma 6.3 in Appendix A.6.

**THEOREM 6.4** (Number of Bures-Wasserstein minimizing geodesic segments in  $\text{Cov}(n)$ ). *Let  $\Sigma, \Lambda \in \text{Cov}(n)$  with  $\text{rk}(\Sigma) = k$  and  $\text{rk}(\Lambda) = l$ . We assume that  $k \geq l$  without loss of generality. We denote  $r = \text{rk}(\Sigma\Lambda)$ . We have  $l - r \leq n - k$ .*

- 1. *There exists a bijection between the set of minimizing geodesics from  $\Sigma$  to  $\Lambda$  and the closed unit ball of  $\mathbb{R}^{(k-r) \times (l-r)}$  for the spectral norm  $\bar{\mathcal{B}}_S(0, 1) = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S \leq 1\} = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid 0 \leq R_0^\top R_0 \leq I_{l-r}\}$ .*
- 2. *The minimizing geodesic is unique if and only if  $r = l$ . This includes the cases  $k = n$ .*
- 3. *There is an infinite number of minimizing geodesics if and only if  $r < l$ .*
- 4. *The minimizing geodesics corresponding to the choices  $R_0 \in \text{St}(k-r, l-r)$  (including the empty matrix if  $r = l$ ) have rank exactly  $k$  on  $[0, 1]$  (on  $[0, 1]$  if  $l = k$ ). Note that  $\text{St}(k-r, l-r)$  is included in the unit sphere  $\mathcal{S}_S(0, 1) = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S = 1\}$ .*
- 5. *The minimizing geodesic corresponding to the choice  $R_0 = 0$  (or the empty matrix if  $r = l$ ) writes for all  $t \in [0, 1]$ :*

$$(6.1) \quad \gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda).$$

*If  $r = l$ , it has rank exactly  $k$  on  $[0, 1]$ .*

*The number of minimizing geodesic segments in  $\text{Sym}^+(n, k)$  and in  $\text{Cov}(n)$  is summarized in Table 6.1 with  $n \geq k \geq l \geq r$ .*

See the proof of Theorem 6.4 in Appendix A.7.

**7. Conclusion.** We have answered several open questions on geodesics of the Bures-Wasserstein distance on covariance matrices. Beyond geodesics, a very important element of Riemannian geometry is the curvature. We know that the space of covariance matrices with the Bures-Wasserstein distance is an Aleksandrov space of non-negative curvature [37] and we know the curvature tensor in each stratum [36, 37, 25]. However, we lack a comprehensive and global approach of the curvature

$\Sigma \in$	$\Lambda \in$	$r = \text{rk}(\Sigma\Lambda)$	Number of minimizing geodesics	
			in $\text{Sym}^+(n, k)$	in $\text{Cov}(n)$
$\text{Sym}^+(n)$	$\text{Sym}^+(n)$	$n$	1	1
$\text{Sym}^+(n)$	$\text{Sym}^+(n, k)$	$k$	1	1
$\text{Sym}^+(n, k)$	$\text{Sym}^+(n, k)$	$k$	1	1
		$k - 1$	2	$\infty$
		$< k - 1$	$\infty$	$\infty$
$\text{Sym}^+(n, k)$	$\text{Sym}^+(n, l)$	$l$	1	1
		$< l$	$\infty$	$\infty$

Table 6.1: Number of Bures-Wasserstein minimizing geodesics ( $n \geq k \geq l \geq r$ ).

of the whole metric space. In particular, what is the appropriate notion of curvature to use to go from a stratum to another?

In the community of geometric statistics, most of the stratified spaces that were studied from the viewpoint of geodesics or curvature are very singular (spiders, trees) or a bit complex to start with (BHV space, Wald space) [12]. Thus, the familiar example of the Bures-Wasserstein Riemannian orbit space appears to be a good basis to generalize concepts defined in Riemannian statistics. Indeed, after studying the geometry of these non-Riemannian spaces, what statistical tools should we define on them to generalize the Euclidean and Riemannian ones? This is probably the main question to investigate for the future.

**Acknowledgements.** This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant G-Statistics agreement No 786854). This work has been supported by the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01 and through the 3IA Côte d’Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. The authors warmly thank Anna Calissano for insightful discussions about the Bures-Wasserstein distance and orbit spaces.

**A. Appendix.** In this Appendix, we prove the main results of the paper. For readability, we recall the result before stating the proof.

**A.1. Lemma 3.2.** (Euclidean and Bures-Wasserstein topologies coincide) The Euclidean distance  $d^E$  and the Bures-Wasserstein distance  $d^{BW}$  define the same topology on  $\text{Cov}(n)$ .

*Proof of Lemma 3.2.* The map  $\pi : X \in \text{Mat}(n) \mapsto XX^T \in (\text{Cov}(n), d^E)$  is continuous so the quotient topology, i.e. the topology induced by the Bures-Wasserstein distance, is finer than the Euclidean topology. Conversely, let  $\mathcal{U}$  be an open set for the Bures-Wasserstein distance. Let  $\Sigma \in \mathcal{U}$ . Let  $\varepsilon > 0$  such that the Bures-Wasserstein ball  $\mathcal{B}^{BW}(\Sigma, \varepsilon)$  is included in  $\mathcal{U}$ . The set  $\mathcal{V} = \text{pow}_2(\mathcal{B}^E(\Sigma^{1/2}, \varepsilon))$  is open for the Euclidean distance because the map  $\text{pow}_2 : \Sigma \mapsto \Sigma^2$  is a homeomorphism of  $(\text{Cov}(n), d^E)$ . Moreover, if  $\Lambda \in \mathcal{V}$ , then  $d^{BW}(\Sigma, \Lambda) \leq d^E(\Sigma^{1/2}, \Lambda^{1/2}) \leq \varepsilon$  so  $\Lambda \in \mathcal{B}^{BW}(\Sigma, \varepsilon) \subseteq \mathcal{U}$ . So  $\mathcal{V} \subseteq \mathcal{U}$  is a Euclidean neighborhood of  $\Sigma$ , so  $\mathcal{U}$  is open for the Euclidean distance. Therefore the two topologies coincide.  $\square$

**A.2. Theorem 4.2.** (Bures-Wasserstein geodesics on  $\text{Sym}^+(n)$ )

Let  $\Sigma \in \text{Sym}^+(n)$ .

1. (Exponential map) [23] For all  $V \in T_\Sigma \text{Sym}^+(n) \equiv \text{Sym}(n)$ , the geodesic from  $\Sigma$  with initial speed  $V$  writes  $\gamma_{(\Sigma, V)}(t) = \Sigma + tV + t^2 \mathcal{S}_\Sigma(V) \Sigma \mathcal{S}_\Sigma(V) \in \text{Sym}^+(n)$ .
2. (Definition interval) Let  $\lambda_{\max} = \max \text{sp}(\mathcal{S}_\Sigma(V))$  and  $\lambda_{\min} = \min \text{sp}(\mathcal{S}_\Sigma(V))$ . The definition interval of the geodesic  $\gamma_{(\Sigma, V)}$  is the interval  $I_{\Sigma, V}$  defined by:
  - $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, -\frac{1}{\lambda_{\min}})$  if  $\lambda_{\min} < 0 < \lambda_{\max}$ ,
  - $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_{\min}})$  if  $\lambda_{\min} < 0$  and  $\lambda_{\max} \leq 0$ ,
  - $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, +\infty)$  if  $\lambda_{\min} \geq 0$  and  $\lambda_{\max} > 0$ ,
  - $I_{\Sigma, V} = \mathbb{R}$  if  $\lambda_{\min} = \lambda_{\max} = 0$  (which only happens for  $V = 0$ ).
3. (Cut time) The cut time is  $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$  if  $\lambda_{\min} < 0$  or  $+\infty$  otherwise. The geodesic  $\gamma_{(\Sigma, V)} : I_{\Sigma, V} \rightarrow \mathcal{M}$  is even minimizing on  $I_{\Sigma, V}$ .
4. (Logarithm map) For all  $\Lambda \in \text{Sym}^+(n)$ , there exists a unique preimage  $V \in \text{Pre}_\Sigma(\Lambda)$ . It writes  $V = 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}) - 2\Sigma$ , where we denote  $\text{sym}(M) = \frac{1}{2}(M + M^\top)$ . The geodesic joining  $\Sigma$  to  $\Lambda$  writes:

$$\gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}).$$

Moreover, it is a logarithm:  $V \in \mathcal{L}og_x(y)$ . Thus the logarithm map is defined on  $\mathcal{U}_\Sigma = \text{Sym}^+(n)$  and it writes:

$$\text{Log}_\Sigma : \begin{cases} \text{Sym}^+(n) & \longrightarrow & T_\Sigma \text{Sym}^+(n) \\ \Lambda & \longmapsto & 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}) - 2\Sigma \end{cases} .$$

*Proof of Theorem 4.2.* We prove statement 3 in the end because it requires statement 4.

1. (Exponential map) The expression of the exponential map comes from [23].
2. (Definition domain) The domain  $I_{\Sigma, V}$  is described in [23] as the connected component of 0 in  $J_{\Sigma, V} = \{t \in \mathbb{R} | I_n + t \mathcal{S}_\Sigma(V) \in \text{Sym}^+(n)\}$ . Since  $t \in J_{\Sigma, V}$  if and only if  $0 \notin \{1 + t\lambda | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\}$  if and only if  $t \notin \{-\frac{1}{\lambda} | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\}$ , we have  $\max(-\infty, 0] \cap \{-\frac{1}{\lambda} | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\} = -\frac{1}{\lambda_{\max}}$  if  $\lambda_{\max} > 0$  and  $\min[0, +\infty) \cap \{-\frac{1}{\lambda} | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\} = -\frac{1}{\lambda_{\min}}$  if  $\lambda_{\min} < 0$ . Therefore, we have the following cases:
  - if  $\lambda_{\min} < 0 < \lambda_{\max}$ , then  $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, -\frac{1}{\lambda_{\min}})$ ,
  - if  $\lambda_{\min} < 0$  and  $\lambda_{\max} \leq 0$ , then  $(-\infty, 0] \subseteq J_{\Sigma, V}$  so  $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_{\min}})$ ,
  - if  $\lambda_{\min} \geq 0$  and  $\lambda_{\max} > 0$ , then  $[0, +\infty) \subseteq J_{\Sigma, V}$  so  $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, +\infty)$ ,
  - if  $\lambda_{\min} \geq 0$  and  $\lambda_{\max} \leq 0$ , which means  $\lambda_{\min} = \lambda_{\max} = 0$ , then  $V = 0$  and  $I_{\Sigma, V} = J_{\Sigma, V} = \mathbb{R}$ .
4. (Logarithm map) The existence of a preimage  $V \in \text{Pre}_\Sigma(V)$  and even a logarithm  $V \in \mathcal{L}og_\Sigma(V)$  (because it satisfies  $\|V\| = d(\Sigma, \Lambda)$ ) is due to [3]. The geodesic joining  $\Sigma$  to  $\Lambda$  (Equation 4.1) is derived in [3] and it suffices to derive the expression at  $t = 0$  to compute  $V = \dot{\gamma}_{\Sigma \rightarrow \Lambda}(0) = 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}) - 2\Sigma$ .

The uniqueness of the preimage comes from [24, Proposition 4.4]. Indeed, it is stated that there exists a unique  $W \in \mathcal{H}_{\Sigma^{1/2}}$  such that:

- (a) for all  $t \in [0, 1]$ ,  $\Sigma^{1/2}(\Sigma^{1/2} + tW) \in \text{GL}(n)$ ,
- (b)  $(\Sigma^{1/2} + W)(\Sigma^{1/2} + W)^\top = \Lambda$ .

Therefore, there exists a unique  $V = d_{\Sigma^{1/2}} \pi(W) = \Sigma^{1/2} W^\top + W \Sigma^{1/2} \in T_\Sigma \text{Sym}^+(n)$  (and  $W = V_{\Sigma^{1/2}}^\#$ ) such that  $1 \in I_{\Sigma, V}$  (that is, for all  $t \in [0, 1]$ ,

$\text{Exp}_\Sigma(tV) \in \text{Sym}^+(n)$ ) and  $\text{Exp}_\Sigma(V) = \Lambda$ , i.e.  $V \in \mathcal{P}re_\Sigma(\Lambda)$ . Thus the logarithm map  $\text{Log}_\Sigma$  is defined on  $\mathcal{U}_\Sigma = \text{Sym}^+(n)$ .

3. (Cut time) Let us prove that  $\gamma_{(\Sigma, V)}$  is minimizing on  $I_{\Sigma, V}$ . This will prove in particular that  $t_{cut}(\Sigma, V) = \sup I_{\Sigma, V}$ . Let  $t, t' \in I_{\Sigma, V}$ ,  $t < 0 < t'$ , let  $\Lambda = \text{Exp}_\Sigma(tV)$  and  $\Lambda' = \text{Exp}_\Sigma(t'V)$ . Changing the base point of the geodesic, we have  $\Lambda' = \text{Exp}_\Lambda((t' - t)V')$  with  $V' = -\dot{\gamma}_{(\Sigma, V)}(t) \in T_\Lambda \text{Sym}^+(n)$ . Since for all  $s \in [0, 1]$ ,  $(1-s)t + st' \in I_{\Sigma, V}$  and  $\text{Exp}_\Lambda(s(t' - t)V') = \text{Exp}_\Sigma(((1-s)t + st')V) \in \text{Sym}^+(n)$ , we have  $1 \in I_{\Lambda, (t' - t)V'}$  so  $(t' - t)V' \in \mathcal{P}re_\Lambda(\Lambda')$ . By uniqueness of the preimage of  $\Lambda'$  from  $\Lambda$ ,  $\text{Log}_\Lambda(\Lambda') = (t' - t)V'$  and  $\gamma_{\Lambda, (t' - t)V'}$  is minimizing on  $[0, 1]$ . Equivalently,  $\gamma_{(\Sigma, V)}$  is minimizing on  $[t, t']$  so it is minimizing on  $I_{\Sigma, V}$ .  $\square$

**A.3. Theorem 5.2.** (Horizontal lift, tangent space, metric)

Let  $\Sigma \in \text{Sym}^+(n, k)$ , let  $X \in \mathbb{R}_*^{n \times k}$  such that  $\Sigma = XX^\top \in \text{Sym}^+(n, k)$  and let  $V \in T_\Sigma \text{Sym}^+(n, k)$ . Let  $\Sigma = UDU^\top$  be a singular value decomposition with  $D \in \text{Diag}^+(k)$  and  $U \in \text{St}(n, k)$ . We denote  $S = S_{\Sigma, V} = US_D(U^\top VU)U^\top$ , where  $\mathcal{S}_A(B)$  denotes the unique solution  $Z$  of the Sylvester equation  $AZ + ZA = B$ . Note that  $S_{\Sigma, V}$  and  $(I_n - UU^\top)$  are independent from the chosen decomposition.

1. (Tangent space)  $T_\Sigma \text{Sym}^+(n, k) = \{V \in \text{Sym}(n) | X_\perp^\top V X_\perp = 0\}$ ,
2. (Horizontal lift)  $V_X^\# = X(X^\top X)^{-1} \mathcal{S}_{X^\top X}(X^\top V X) + X_\perp X_\perp^\top V X (X^\top X)^{-1}$ , where  $X_\perp X_\perp^\top = I_n - X(X^\top X)^{-1} X^\top$ ,
3. (Bures-Wasserstein metric)  $g_\Sigma^{BW(n, k)}(V, V) = \text{tr}(S_{\Sigma, V} \Sigma S_{\Sigma, V} + V \Sigma^{-1} V (I_n - UU^\top))$ .

*Proof of Theorem 5.2.* 1&2. We prove the expression of the tangent space and the horizontal lift together. Let  $V \in T_\Sigma \text{Sym}^+(n, k)$ . The horizontal lift is defined by:

- (lift)  $V = d_X \pi(V^\#) = X(V^\#)^\top + V^\# X^\top$ ,
- (horizontal)  $V^\# = X(X^\top X)^{-1} F + X_\perp K$  where  $F \in \text{Sym}(r)$  and  $K \in \text{Mat}(n - r, r)$ .

When we plug the second equality in the first one and we multiply by  $X^\top$  on the left and  $X$  on the right, since  $X^\top X_\perp = 0$ , we get immediately  $X^\top V X = X^\top X F + F X^\top X$  so  $F = \mathcal{S}_{X^\top X}(X^\top V X)$ . By multiplying by  $X_\perp^\top$  on the left instead, we get  $X_\perp^\top V X = K X^\top X$  so  $K = X_\perp^\top V X (X^\top X)^{-1}$ . Since the matrix  $(X(X^\top X)^{-1/2}; X_\perp)$  is orthogonal, we have  $X(X^\top X)^{-1} X^\top + X_\perp X_\perp^\top = I_n$ .

We compute  $d_X \pi(V^\#)$  to check that it is equal to  $V$ :

$$\begin{aligned} d_X \pi(V^\#) &= X[(X^\top X)^{-1} F + F(X^\top X)^{-1}] X^\top \\ &\quad + 2 \text{sym}(X_\perp X_\perp^\top V X (X^\top X)^{-1} X^\top) \\ &= X(X^\top X)^{-1} X^\top V X (X^\top X)^{-1} X^\top \\ &\quad + 2 \text{sym}(X_\perp X_\perp^\top V (I_n - X_\perp X_\perp^\top)) \\ &= (I_n - X_\perp X_\perp^\top) V (I_n - X_\perp X_\perp^\top) \\ &\quad + X_\perp X_\perp^\top V + V X_\perp X_\perp^\top - 2 X_\perp X_\perp^\top V X_\perp X_\perp^\top \\ &= V - X_\perp X_\perp^\top V X_\perp X_\perp^\top. \end{aligned}$$

Thus,  $X_\perp X_\perp^\top V X_\perp X_\perp^\top = 0$  so  $X_\perp^\top V X_\perp = 0$ . Conversely, if  $X_\perp^\top V X_\perp = 0$ , then  $V$  is the image by  $d_X \pi$  of a horizontal vector so  $V \in T_\Sigma \text{Sym}^+(n, k)$ . Hence  $T_\Sigma \text{Sym}^+(n, k) = \{V \in \text{Sym}(n) | X_\perp^\top V X_\perp = 0\}$ .

3. The quotient metric is defined by  $g_{\Sigma}^{\text{BW}(n,k)}(V, V) = \text{tr}(V_{\Sigma}^{\#}(V_{\Sigma}^{\#})^{\top})$  so we only need to compute  $V^{\#}(V^{\#})^{\top}$  for any  $X$ , for example  $X = UD^{1/2}$ , and its trace.

$$\begin{aligned} V^{\#} &= UD^{1/2}D^{-1}\mathcal{S}_D(D^{1/2}U^{\top}VUD^{1/2}) \\ &\quad + (I_n - UD^{1/2}D^{-1}D^{1/2}U^{\top})VUD^{1/2}D^{-1} \\ &= US_D(U^{\top}VU)D^{1/2} + (I_n - UU^{\top})VUD^{-1/2} \\ &= SUD^{1/2} + (I_n - UU^{\top})VUD^{-1/2}, \\ V^{\#}(V^{\#})^{\top} &= SUDU^{\top}S + SV(I_n - UU^{\top}) + (I_n - UU^{\top})VS \\ &\quad + (I_n - UU^{\top})V\Sigma^{-1}V(I_n - UU^{\top}). \end{aligned}$$

$$\text{Hence } \text{tr}(V^{\#}(V^{\#})^{\top}) = \text{tr}(S\Sigma S + V\Sigma^{-1}V(I_n - UU^{\top})). \quad \square$$

**A.4. Theorem 5.3.** (Bures-Wasserstein geodesics on  $\text{Sym}^+(n, k)$ ) Let  $\Sigma, \Lambda \in \text{Sym}^+(n, k)$  and  $X, Y \in \mathbb{R}_*^{n \times k}$  such that  $XX^{\top} = \Sigma$  and  $YY^{\top} = \Lambda$ . Let  $U \in \text{St}(n, k)$  and  $D \in \text{Diag}^+(k)$  such that  $\Sigma = UDU^{\top}$ .

1. (Exponential map) For all  $V \in T_{\Sigma}\text{Sym}^+(n, k)$ , the geodesic from  $\Sigma$  with initial speed  $V$  is  $\gamma_{(\Sigma, V)} : t \in I_{\Sigma, V} \mapsto \Sigma + tV + t^2W_{\Sigma, V}$ , where  $W_{\Sigma, V} = S_{\Sigma, V}\Sigma S_{\Sigma, V} + S_{\Sigma, V}V(I_n - UU^{\top}) + (I_n - UU^{\top})VS_{\Sigma, V} + (I_n - UU^{\top})V\Sigma^+V(I_n - UU^{\top})$  and  $S_{\Sigma, V} = US_D(U^{\top}VU)U^{\top}$ .
2. (Definition interval) Let  $F_{X, V}^0 = S_{X^{\top}X}((X^{\top}X)^{-1/2}X^{\top}VX(X^{\top}X)^{-1/2})$  and  $M_{X, V}^0 = (X^{\top}X)^{-3/2}X^{\top}V(I_n - X(X^{\top}X)^{-1}X^{\top})VX(X^{\top}X)^{-3/2} \in \text{Sym}(n)$ . Let  $\mathcal{E}_{\Sigma, V} = \{\lambda \in \text{sp}(F_{X, V}^0) \mid \ker(\lambda I_k - F_{X, V}^0) \cap \ker(M_{X, V}^0) \neq \{0\}\} \subseteq \text{sp}(S_{\Sigma, V})$ . If  $\mathcal{E}_{\Sigma, V}$  is non-empty, then let  $\lambda_+ = \max \mathcal{E}_{\Sigma, V}$  and  $\lambda_- = \min \mathcal{E}_{\Sigma, V}$ . The definition interval of the geodesic  $\gamma_{(\Sigma, V)}$  is the interval  $I_{\Sigma, V}$  defined by:
  - $I_{\Sigma, V} = (-\frac{1}{\lambda_+}, -\frac{1}{\lambda_-})$  if  $\lambda_- < 0 < \lambda_+$ ,
  - $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_-})$  if  $\lambda_- < 0$  and  $\lambda_+ \leq 0$ ,
  - $I_{\Sigma, V} = (-\frac{1}{\lambda_+}, +\infty)$  if  $\lambda_- \geq 0$  and  $\lambda_+ > 0$ ,
  - $I_{\Sigma, V} = \mathbb{R}$  if  $\mathcal{E}_{\Sigma, V}$  is empty.

Applying this to  $X = UD^{1/2}$  without loss of generality,  $F_{X, V}^0 = \mathcal{S}_D(U^{\top}VU)$  and  $M_{X, V}^0 = D^{-1}U^{\top}V(I_n - UU^{\top})VUD^{-1}$  which is a bit more tractable to compute  $\mathcal{E}_{\Sigma, V}$ .

3. (Cut time) Let  $\lambda_{\max} = \max \text{sp}(F_{X, V}^0)$  and  $\lambda_{\min} = \min \text{sp}(F_{X, V}^0)$ . Note that if  $\mathcal{E}_{\Sigma, V} \neq \emptyset$ , then we have  $(\lambda_-, \lambda_+) \subseteq (\lambda_{\min}, \lambda_{\max})$ . The cut time is  $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$  if  $\lambda_{\min} < 0$  or  $+\infty$  otherwise. Symmetrically, we have  $t_{\text{cut}}(\Sigma, -V) = \frac{1}{\lambda_{\max}}$  if  $\lambda_{\max} > 0$  or  $+\infty$  otherwise.
4. (Preimages) We define the indexing set  $\mathcal{I}_{X, Y}^{\text{Pre}}$  by:

$$\begin{aligned} \mathcal{I}_{X, Y}^{\text{Pre}} &= \{R \in O(n) \mid H := X^{\top}YR^{\top} \in \text{Sym}(n) \text{ and} \\ &\quad \forall \mu < 0, \ker(\mu I_k - (X^{\top}X)^{-1/2}H(X^{\top}X)^{-1/2}) \\ &\quad \cap \ker(\mu^2 I_k - (X^{\top}X)^{-1/2}RY^{\top}YR^{\top}(X^{\top}X)^{-1/2}) = \{0\}\}. \end{aligned}$$

For  $R \in \mathcal{I}_{X, Y}^{\text{Pre}}$ , we denote  $H = H_{X, Y, R} = X^{\top}YR^{\top}$  so that  $X^{\top}Y = HR$ . Then the map  $R \in \mathcal{I}_{X, Y}^{\text{Pre}} \mapsto V = 2\text{sym}(XRY^{\top}) - 2\Sigma \in \text{Pre}_{\Sigma}(\Lambda)$  is a bijection whose inverse is  $V \in \text{Pre}_{\Sigma}(\Lambda) \mapsto R = (Y^{\top}Y)^{-1}Y^{\top}(X + V_{\Sigma}^{\#}) \in \mathcal{I}_{X, Y}^{\text{Pre}}$ .

The geodesic joining  $\Sigma$  to  $\Lambda$  parametrized by  $R \in \mathcal{I}_{X, Y}^{\text{Pre}}$  writes:

$$\forall t \in [0, 1], \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1-t)^2\Sigma + t^2\Lambda + 2t(1-t)\text{sym}(XRY^{\top}).$$

5. (Logarithms) Let  $\mathcal{I}_{X,Y}^{\mathcal{L}og} = \{R \in O(n) | H_{X,Y,R} = X^\top Y R^\top \in \text{Cov}(n)\} = \{R \in O(n) | H_{X,Y,R} = (X^\top \Lambda X)^{1/2}\} = \{R \in O(n) | X^\top Y = (X^\top \Lambda X)^{1/2} R\} \subseteq \mathcal{I}_{X,Y}^{\mathcal{P}re}$ . Then, the map  $R \in \mathcal{I}_{X,Y}^{\mathcal{L}og} \mapsto V = 2 \text{sym}(XRY^\top) - 2\Sigma \in \mathcal{L}og_\Sigma(\Lambda)$  is a bijection whose inverse is  $V \in \mathcal{L}og_\Sigma(\Lambda) \mapsto R = (Y^\top Y)^{-1} Y^\top (X + V_X^\#) \in \mathcal{I}_{X,Y}^{\mathcal{L}og}$ .
6. (Logarithm map) Let  $r = \text{rk}(\Sigma\Lambda) = \text{rk}(X^\top Y) = \text{rk}(H)$ .
- (a) If  $r = k$ , then there exists a unique logarithm of  $\Lambda$  from  $\Sigma$ . In this case, the minimizing geodesic joining  $\Sigma$  to  $\Lambda$  writes:
- (A.1)
- $$\gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2} ((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda).$$
- (b) If  $r = k - 1$ , then there exist exactly two logarithms of  $\Lambda$  from  $\Sigma$ .
- (c) If  $r < k - 1$ , then there is an infinity of logarithms of  $\Lambda$  from  $\Sigma$ .
- Therefore, the logarithm map is defined on  $\mathcal{U}_\Sigma = \{\Lambda \in \text{Sym}^+(n, k) | \text{rk}(\Sigma\Lambda) = k\}$  and it writes  $\text{Log}_\Sigma : \Lambda \in \mathcal{U}_\Sigma \mapsto 2 \text{sym}(\Sigma^{1/2} ((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda) - 2\Sigma \in T_\Sigma \text{Sym}^+(n, k)$ .

*Proof of Theorem 5.3.* We prove statement 3 in the end because it requires statements 4 and 5.

1. (Exponential map) By Theorem 2.15 [30], the exponential map is simply  $\text{Exp}_\Sigma(tV) = \pi(\text{Exp}_X(tV_X^\#)) = (X + tV_X^\#)(X + tV_X^\#)^\top = \Sigma + tV + V_X^\#(V_X^\#)^\top$ . The matrix  $W = V_X^\#(V_X^\#)^\top$  was already computed in the proof of Theorem 5.2.
2. (Definition domain) As in  $\text{Sym}^+(n)$ , let us first determine  $J_{\Sigma,V} = \{t \in \mathbb{R} | \text{rk}(\Sigma + tV + t^2 W_{\Sigma,V}) = k\}$ . According to [24, Proposition 3.2] applied to  $tV_X^\# = t[X(X^\top X)^{-1} F_{X,V} + X_\perp K_{X,V}]$  with  $F = F_{X,V} = \mathcal{S}_{X^\top X}(X^\top V X)$  and  $K = K_{X,V} = X_\perp^\top V X (X^\top X)^{-1}$ , we have  $t \in J_{\Sigma,V}$  if and only if  $\ker(I_k + t(X^\top X)^{-1} F) \cap \ker(K) = \{0\}$ . Let  $\mathcal{E}_{\Sigma,V} = \{\lambda \in \text{sp}((X^\top X)^{-1} F) | \ker(\lambda I_k - (X^\top X)^{-1} F) \cap \ker(K) \neq \{0\}\}$ . It is clear that  $J_{\Sigma,V} = \mathbb{R} \setminus \{-\frac{1}{\lambda} | \lambda \in \mathcal{E}_{\Sigma,V}\}$ . Then  $I_{\Sigma,V}$  is the connected component of 0 in  $J_{\Sigma,V}$ . Its computation is analogous to the one in the proof of Theorem 4.2.
- To get a condition in  $\mathcal{E}_{\Sigma,V}$  that directly depends on  $V$ , note that the condition rewrites  $\ker(\lambda I_k - (X^\top X)^{-1/2} F (X^\top X)^{-1/2}) \cap \ker(K (X^\top X)^{-1/2}) \neq \{0\}$  and  $\ker(K (X^\top X)^{-1/2}) = \ker((X^\top X)^{-1/2} K^\top K (X^\top X)^{-1/2})$  with  $K^\top K = (X^\top X)^{-1} X^\top V (I_n - X (X^\top X)^{-1} X^\top) V X (X^\top X)^{-1}$ . Note that  $\mathcal{E}_{\Sigma,V}$  is independent from the choice of  $X$  because  $F_{XR,V} = R^\top F_{X,V} R$  and  $K_{XR,V} = R^\top K_{X,V} R$  for all  $R \in O(n)$  so the condition does not depend on  $X$ .
4. (Preimages) In [24, Propositions 4.4 & 4.5], the solutions of the equation  $\text{Exp}_\Sigma(V) = \Lambda$  are not computed in the definition domain  $\mathcal{D}_\Sigma$  of the exponential map but in the wider set  $\{V \in T_\Sigma \text{Sym}^+(n, k) | \Sigma + V + W_{\Sigma,V} \in \text{Sym}^+(n, k)\}$ . Thus, the geodesic  $\gamma_{(\Sigma,V)}$  may leave the manifold  $\text{Sym}^+(n, k)$  before reaching  $\Lambda$ . Therefore, we complete their work with the additional condition  $1 \in I_{\Sigma,V}$  to characterize the preimages of  $\Lambda$  from  $\Sigma$ .

From [24], we know that preimages  $V$  necessarily satisfy  $V_X^\# = YR^\top - X$  with  $X^\top Y = HR$ ,  $H \in \text{Sym}(n)$ ,  $R \in O(n)$ . Thus,  $V = XRY^\top + YR^\top X^\top - 2XX^\top$  so  $X^\top V X = HX^\top X + X^\top XH - 2(X^\top X)^2$  so  $F = \mathcal{S}_{X^\top X}(X^\top V X) = H - X^\top X$ . Moreover,  $K = X_\perp^\top V X (X^\top X)^{-1} = X_\perp^\top YR^\top$ . Denoting  $A = (X^\top X)^{-1/2} H (X^\top X)^{-1/2}$  and  $B = (X^\top X)^{-1/2} RY^\top YR^\top (X^\top X)^{-1/2}$ , we have  $(X^\top X)^{-1/2} F (X^\top X)^{-1/2} = A - I_k$  and  $(X^\top X)^{-1/2} K^\top K (X^\top X)^{-1/2} = (X^\top X)^{-1/2} RY^\top X_\perp X_\perp^\top YR^\top (X^\top X)^{-1/2} = B - A^2$ .



We can now compute  $\mathcal{E}_{\Sigma, V}$  and  $I_{\Sigma, V}$ . For all  $\lambda \in \mathbb{R}$ , for all  $Z \in \mathbb{R}^n$ :

$$\begin{aligned} Z &\in \ker(\lambda I_k - S_{X, V}^0) \cap \ker(M_{X, V}^0) \\ &\iff Z \in \ker((\lambda + 1)I_k - A) \cap \ker(B - A^2) \\ &\iff AZ = (\lambda + 1)Z \text{ and } BZ = A^2Z \\ &\iff AZ = (\lambda + 1)Z \text{ and } BZ = (\lambda + 1)^2Z \\ &\iff Z \in \ker((\lambda + 1)I_k - A) \cap \ker((\lambda + 1)^2I_k - B). \end{aligned}$$

Therefore:

$$\begin{aligned} \mathcal{E}_{\Sigma, V} &= \{\lambda \in \text{sp}(A - I_k) \mid \ker((\lambda + 1)I_k - A) \cap \ker((\lambda + 1)^2I_k - B) \neq \{0\}\} \\ &= \{\mu - 1 \in \text{sp}(A - I_k) \mid \ker(\mu I_k - A) \cap \ker(\mu^2 I_k - B) \neq \{0\}\}. \end{aligned}$$

Thus, denoting  $\lambda_- = \min \mathcal{E}_{\Sigma, V}$ , the condition  $1 \in I_{\Sigma, V}$  rewrites:

$$\begin{aligned} 1 \in I_{\Sigma, V} &\iff \lambda_- \geq 0 \text{ or } -\frac{1}{\lambda_-} \geq 1 \iff \lambda_- \geq -1 \\ &\iff \forall \mu \in \text{sp}(A), \ker(\mu I_k - A) \cap \ker(\mu^2 I_k - B) \neq \{0\} \implies \mu \geq 0 \\ &\iff \forall \mu < 0, \ker(\mu I_k - A) \cap \ker(\mu^2 I_k - B) = \{0\}. \end{aligned}$$

To conclude, with the notations of statement 4,  $V = 2 \text{sym}(XRY^\top) - 2\Sigma \in \text{Pre}_\Sigma(\Lambda)$  if and only if  $R \in \mathcal{I}_{X, Y}^{\text{Pre}}$ .

5. (Logarithms) In [24], it is stated that the shortest vectors  $V = d\pi(V_X^\#)$  with  $V_X^\# = YR^\top - X$  are those for which  $(H, R)$  is a polar decomposition of  $X^\top Y$ , i.e.  $H \geq 0$ . Therefore, we necessarily have  $H = (X^\top \Lambda X)^{1/2}$ . It is well known that they even satisfy  $\|V\| = \|V^\#\| = d^{\text{BW}}(\Sigma, \Lambda)$  (see Definition 3.1). Moreover, if  $H \geq 0$ , the condition  $1 \in I_{\Sigma, V}$  is automatically satisfied, as stated in [24, Corollary 3.3 (5)]. So with the notations of statement 5, the logarithms are indexed by  $\mathcal{I}_{X, Y}^{\text{Log}}$ .
6. (Logarithm map) We have  $\text{rk}(\Sigma\Lambda) \leq \text{rk}(X^\top Y)$  since  $\Sigma\Lambda = X(X^\top Y)Y^\top$ . We also have  $X^\top Y = (X^\top X)^{-1}X^\top(\Sigma\Lambda)Y(Y^\top Y)^{-1}$  so  $\text{rk}(X^\top Y) \leq \text{rk}(\Sigma\Lambda)$ . Finally,  $\text{rk}(\Sigma\Lambda) = \text{rk}(X^\top Y) = \text{rk}(H)$ . We denote it  $r = \text{rk}(\Sigma\Lambda)$ .
  - (a) As stated in [24], if  $r = k$ , then there exists a unique logarithm of  $\Sigma$  from  $\Lambda$ . Moreover, we can compute an explicit expression. Indeed,  $R = H^{-1}X^\top Y$  so  $XRY^\top = XH^{-1}X^\top \Lambda = X(X^\top \Lambda X)^{-1/2}X^\top \Lambda$ . Since the choice of  $X$  is free, let us take  $X = UD^{1/2}$  where  $\Sigma = UDU^\top$  with  $U \in \text{St}(n, k)$  and  $D \in \text{Diag}^+(k)$ . Therefore:

$$\begin{aligned} XRY^\top &= UD^{1/2}(D^{1/2}U^\top \Lambda UD^{1/2})^{-1/2}D^{1/2}U^\top \Lambda \\ &= UD^{1/2}U^\top U((D^{1/2}U^\top \Lambda UD^{1/2})^{1/2})^{-1}U^\top UD^{1/2}U^\top \Lambda \\ &= \Sigma^{1/2}(U(D^{1/2}U^\top \Lambda UD^{1/2})^{1/2}U^\top)^{-\Sigma^{1/2}}\Lambda \\ &= \Sigma^{1/2}((UD^{1/2}U^\top \Lambda UD^{1/2}U^\top)^{1/2})^{-\Sigma^{1/2}}\Lambda \\ &= \Sigma^{1/2}((\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2})^{-\Sigma^{1/2}}\Lambda. \end{aligned}$$

So the unique minimizing geodesic joining  $\Sigma$  to  $\Lambda$  writes:

$$\gamma_{(\Sigma, \Lambda)}(t) = (1-t)^2\Sigma + t^2\Lambda + 2t(1-t)\text{sym}(\Sigma^{1/2}((\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2})^{-\Sigma^{1/2}}\Lambda).$$

- (b) If  $r = k - 1$ , without loss of generality, let us assume that  $X^\top Y \in \text{Diag}(k)$ ,  $X^\top Y = \text{Diag}(d_1, \dots, d_{k-1}, 0)$ . Then  $H \geq 0$  and  $H^2 = (X^\top Y)^2$  imposes that  $H = \text{Diag}(|d_1|, \dots, |d_{k-1}|, 0)$ . Therefore, there are only two matrices  $R_\pm \in O(n)$  defined by  $R_\pm = \text{Diag}(\text{sgn}(d_1), \dots, \text{sgn}(d_{k-1}), \pm 1)$  such that  $X^\top Y = HR_\pm$ . Thus there are exactly two logarithms of  $\Lambda$  from  $\Sigma$ .
- (c) If  $r < k - 1$ , similarly we can assume without loss of generality that  $X^\top Y = \text{Diag}(d_1, \dots, d_r, 0, \dots, 0)$ . Then  $H = \text{Diag}(|d_1|, \dots, |d_r|, 0, \dots, 0)$  and  $R = \text{Diag}(\varepsilon, R_0)$  is a block-diagonal matrix with  $R_0 \in O(k - r)$  and  $\varepsilon = \text{Diag}(\text{sgn}(d_1), \dots, \text{sgn}(d_r)) \in \text{Diag}(r)$ . Thus there is an infinity of logarithms of  $\Lambda$  from  $\Sigma$ .

Thus the logarithm map is defined on  $\mathcal{U}_\Sigma = \{\Lambda \in \text{Sym}^+(n, k) \mid \text{rk}(\Sigma\Lambda) = k\}$ , as stated in [24].

3. (Cut time) Let  $t \in I_{\Sigma, V} \cap \mathbb{R}_+$ . Let  $\Lambda = \gamma_{(\Sigma, V)}(t)$ ,  $Y \in \mathbb{R}_*^{n \times k}$  such that  $YY^\top = \Lambda$ ,  $(H, R) \in \text{Sym}(k) \times O(k)$  such that  $X^\top Y = HR$  and  $V_X^\# = YR^\top - X$ . Then,  $X^\top X + tX^\top V_X^\# = X^\top YR^\top = H$ . Besides,  $X^\top X + tX^\top V^\# = (X^\top X)^{1/2}(I_n + tS_{X, V}^0)(X^\top X)^{1/2}$ . Let  $\lambda_{\min} = \min \text{sp}(S_{X, V}^0) = \min \text{sp}(S_{\Sigma, V})$ . Therefore,  $\gamma_{(\Sigma, V)}$  is minimizing on  $[0, t]$  if and only if  $H \geq 0$  if and only if  $I_n + tS_{X, V}^0 \geq 0$  if and only if  $1 + t\lambda_{\min} \geq 0$ . If  $\lambda_{\min} \geq 0$ , the condition is empty so  $t_{\text{cut}}(\Sigma, V) = +\infty$ . If  $\lambda_{\min} < 0$ , the condition writes  $t \leq -\frac{1}{\lambda_{\min}}$  so  $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$ .  $\square$

**A.5. Theorem 6.2.** (Bures-Wasserstein minimizing geodesics in  $\text{Cov}(n)$ ) Let  $\Sigma, \Lambda \in \text{Cov}(n)$  with  $\text{rk}(\Sigma) = k$  and  $\text{rk}(\Lambda) = l$ . Let  $X, Y \in \mathbb{R}^{n \times n}$  such that  $XX^\top = \Sigma$  and  $YY^\top = \Lambda$ . The two following statements are equivalent:

- (i) the curve  $\gamma : [0, 1] \rightarrow \text{Cov}(n)$  is a minimizing geodesic segment from  $\Sigma$  to  $\Lambda$ ,  
 (ii) there exists  $R \in O(n)$  such that  $H_{X, Y, R} := X^\top YR^\top \in \text{Cov}(n)$  and for all  $t \in [0, 1]$ ,  $\gamma(t) = \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1 - t)^2 \Sigma + t^2 \Lambda + 2t(1 - t) \text{sym}(XRY^\top)$ .

Moreover,  $H_{X, Y, R} = (X^\top \Lambda X)^{1/2}$  and the minimizing geodesic  $\gamma_{\Sigma \rightarrow \Lambda}^R$  is of constant rank  $p \geq \max(k, l)$  on  $(0, 1)$ .

*Proof of Theorem 6.2.*

(Necessity) Let  $\gamma : [0, 1] \rightarrow \text{Cov}(n)$  be a minimizing geodesic segment from  $\Sigma = \gamma(0)$  to  $\Lambda = \gamma(1)$ . Let  $p = \max_{t \in [0, 1]} \text{rk}(\gamma(t))$ . By Lemma 6.1,  $\gamma$  is of constant rank  $p \geq \max(k, l)$  on  $(0, 1)$ . In other words,  $\gamma|_{(0, 1)} : (0, 1) \rightarrow \text{Sym}^+(n, p)$  is a minimizing geodesic of  $\text{Sym}^+(n, p)$ . Let  $c_0 : (0, 1) \rightarrow \mathbb{R}_*^{n \times p}$  be a horizontal lift of  $\gamma|_{(0, 1)}$ . Necessarily,  $c_0(t) = (1 - t)X_0 + tY_0$  with  $X_0, Y_0 \in \mathbb{R}^{n \times p}$  with  $X_0X_0^\top = \Sigma$  and  $Y_0Y_0^\top = \Lambda$  since  $(X_0, \Sigma) = \lim_{t \rightarrow 0} (c_0(t), \gamma(t))$  and  $(Y_0, \Lambda) = \lim_{t \rightarrow 1} (c_0(t), \gamma(t))$ . Let us show that  $X_0^\top Y_0 \in \text{Cov}(p)$ .

For all  $[a, b] \subset (0, 1)$ , the tangent vectors  $V_{c_0(a)}^\# = (b - a)(Y_0 - X_0) \in T_{c_0(a)} \mathbb{R}_*^{n \times p} = \mathbb{R}^{n \times p}$  and  $V = c_0(a)(V_{c_0(a)}^\#)^\top + V_{c_0(a)}^\# c_0(a)^\top \in \mathcal{L}og_{\gamma(a)}(\gamma(b)) \subset T_{\gamma(a)} \text{Sym}^+(n, p)$  uniquely determine a pair of matrices  $R_{a, b} \in \mathcal{I}_{c_0(a), c_0(b)}^{\mathcal{L}og}$  and  $H_{a, b} = H_{c_0(a), c_0(b), R_{a, b}} = c_0(a)^\top c_0(b) \in \text{Cov}(p)$ . We compute  $H_{a, b}$ :

$$\begin{aligned} H_{a, b} &= c_0(a)^\top c_0(b) = [(1 - a)X_0 + aY_0]^\top [(1 - b)X_0 + bY_0] \\ &= (1 - a)(1 - b)X_0^\top X_0 + abY_0^\top Y_0 + (1 - a)bX_0^\top Y_0 + a(1 - b)Y_0^\top X_0. \end{aligned}$$

Therefore,  $\lim_{\substack{a \rightarrow 0 \\ b \rightarrow 1}} H_{a, b} = X_0^\top Y_0$  so  $X_0^\top Y_0 \in \text{Cov}(p)$ .

Since  $[X_0 \ 0][X_0 \ 0]^\top = \Sigma$  and  $[Y_0 \ 0][Y_0 \ 0]^\top = \Lambda$ , there exist  $P, Q \in O(n)$  such that  $X = [X_0 \ 0]P$  and  $Y = [Y_0 \ 0]Q$ . Thus the curve  $c : t \in [0, 1] \mapsto (1-t)X + tYR^\top \in \text{Mat}(n)$  with  $R = P^\top Q \in O(n)$  satisfies  $\gamma(t) = c(t)c(t)^\top$  for all  $t \in [0, 1]$ . Indeed, it is equal to  $c_0(t)c_0(t)^\top$  on  $(0, 1)$  and the equality is clear for  $t \in \{0, 1\}$ . Moreover,  $H_{X,Y,R} = X^\top YR^\top = U^\top \text{Diag}(X_0^\top Y_0, 0)U \in \text{Cov}(n)$  so  $H_{X,Y,R} = ((X^\top Y)(X^\top Y)^\top)^{1/2} = (X^\top \Lambda X)^{1/2}$ . Finally,  $\gamma(t) = c(t)c(t)^\top = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(XRY^\top)$ .

(Sufficiency) Let  $\gamma_{(\Sigma, \Lambda)}^R(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(XRY^\top)$  with  $H = H_{X,Y,R} = X^\top YR^\top \in \text{Cov}(n)$  and let us prove that it is a minimizing geodesic segment. We define  $W = YR^\top - X$  and  $c(t) = X + tW = (1-t)X + tYR^\top$  for  $t \in [0, 1]$ . The curve  $c$  is a geodesic of  $\text{Mat}(n)$  such that  $c(t)c(t)^\top = \gamma_{(\Sigma, \Lambda)}^R(t)$  for all  $t \in [0, 1]$ . Moreover,  $L(c) = \|W\| = \text{tr}(XX^\top + YY^\top - 2X^\top YR^\top)^{1/2} = \text{tr}(\Sigma + \Lambda - 2H)^{1/2}$ . Let  $Q \in O(n)$  such that  $X = \Sigma^{1/2}Q$ . Since  $H \geq 0$ ,  $H = (X^\top \Lambda X)^{1/2} = Q^\top (\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} Q$ . Therefore,  $L(c) = \|YR^\top - X\| = d^{\text{BW}}(\Sigma, \Lambda)$ . In other words,  $c : [0, 1] \rightarrow \text{Cov}(n)$  is a minimizing curve between two registered points  $X$  and  $YR^\top$  so by Lemma 2.13, its projection  $\gamma : [0, 1] \rightarrow \text{Cov}(n)$  is a minimizing curve and  $L(\gamma) = L(c) = d^{\text{BW}}(\Sigma, \Lambda)$ .

By Lemma 6.1 again,  $\gamma$  has constant rank  $p \geq \max(k, l)$  on  $(0, 1)$  so  $\gamma|_{(0,1)} : (0, 1) \rightarrow \text{Sym}^+(n, p)$  is a minimizing curve of  $\text{Sym}^+(n, p)$ . Since  $c|_{(0,1)}$  has constant speed, so does  $\gamma|_{(0,1)}$ . By continuity of the length,  $\gamma$  has constant speed on  $[0, 1]$  so  $\gamma : [0, 1] \rightarrow \text{Cov}(n)$  is a minimizing geodesic segment.  $\square$

**A.6. Lemma 6.3.** (Elementary algebra) Let  $\Sigma, \Lambda \in \text{Cov}(n)$  with  $\text{rk}(\Sigma) = k$  and  $\text{rk}(\Lambda) = l$ . Let  $r = \text{rk}(\Sigma\Lambda)$ .

1. For all  $X, Y \in \text{Mat}(n)$  such that  $XX^\top = \Sigma$  and  $YY^\top = \Lambda$ ,  $r = \text{rk}(X^\top Y)$ .
2. We have  $l - r \leq n - k$ .

*Proof of Lemma 6.3.* 1. Let  $X_0 \in \mathbb{R}_*^{n \times k}$  and  $Y_0 \in \mathbb{R}_*^{n \times l}$  such that  $X_0 X_0^\top = \Sigma$  and  $Y_0 Y_0^\top = \Lambda$ . Thus there exist  $P, Q \in O(n)$  such that  $X = [X_0 \ 0]P$  and  $Y = [Y_0 \ 0]Q$ . Since  $\Sigma\Lambda = X(X^\top Y)Y^\top$ , we have  $r \leq \text{rk}(X^\top Y) = \text{rk}(X_0^\top Y_0)$ . Since  $X_0^\top Y_0 = (X_0^\top X_0)^{-1} X_0^\top \Sigma \Lambda Y_0 (Y_0^\top Y_0)^{-1}$ , we have  $\text{rk}(X_0^\top Y_0) \leq \text{rk}(\Sigma\Lambda)$ . Finally,  $r = \text{rk}(X_0^\top Y_0) = \text{rk}(X^\top Y)$ .

2. Let  $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be linear endomorphisms respectively represented by  $\Sigma$  and  $\Lambda$  is the canonical basis. From the rank-nullity theorem applied to the linear map  $f|_{\text{im}(g)} : \text{im}(g) \rightarrow \mathbb{R}^n$ , i.e. the restriction of  $f$  to  $\text{im}(g)$ , and since  $\text{im}(f|_{\text{im}(g)}) = \text{im}(f \circ g)$  and  $\ker(f|_{\text{im}(g)}) \subseteq \ker(f)$ , we have  $\text{rk}(g) = \text{rk}(f|_{\text{im}(g)}) + \dim \ker(f|_{\text{im}(g)}) \leq \text{rk}(f \circ g) + \dim(\ker f) = \text{rk}(f \circ g) + n - \text{rk}(f)$ . This writes  $l - r \leq n - k$ .  $\square$

**A.7. Theorem 6.4.** (Number of Bures-Wasserstein minimizing geodesic segments in  $\text{Cov}(n)$ ) Let  $\Sigma, \Lambda \in \text{Cov}(n)$  with  $\text{rk}(\Sigma) = k$  and  $\text{rk}(\Lambda) = l$ . We assume that  $k \geq l$  without loss of generality. We denote  $r = \text{rk}(\Sigma\Lambda)$ . We have  $l - r \leq n - k$ .

1. There exists a bijection between the set of minimizing geodesics from  $\Sigma$  to  $\Lambda$  and the closed unit ball of  $\mathbb{R}^{(k-r) \times (l-r)}$  for the spectral norm  $\mathcal{B}_S(0, 1) = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S \leq 1\} = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid 0 \leq R_0^\top R_0 \leq I_{l-r}\}$ .
2. The minimizing geodesic is unique if and only if  $r = l$ . This includes the cases  $k = n$ .
3. There is an infinite number of minimizing geodesics if and only if  $r < l$ .
4. The minimizing geodesics corresponding to the choices  $R_0 \in \text{St}(k-r, l-r)$  (including the empty matrix if  $r = l$ ) have rank exactly  $k$  on  $[0, 1)$  (on  $[0, 1]$  if  $l = k$ ). Note that  $\text{St}(k-r, l-r)$  is included in the unit sphere  $\mathcal{S}_S(0, 1) =$

$$\{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S = 1\}.$$

5. The minimizing geodesic corresponding to the choice  $R_0 = 0$  (or the empty matrix if  $r = l$ ) writes for all  $t \in [0, 1]$ :

$$\gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \operatorname{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda).$$

If  $r = l$ , it has rank exactly  $k$  on  $[0, 1]$ .

The number of minimizing geodesic segments in  $\operatorname{Sym}^+(n, k)$  and in  $\operatorname{Cov}(n)$  is summarized in Table 6.1 with  $n \geq k \geq l \geq r$ .

$\Sigma \in$	$\Lambda \in$	$r = \operatorname{rk}(\Sigma \Lambda)$	Number of minimizing geodesics	
			in $\operatorname{Sym}^+(n, k)$	in $\operatorname{Cov}(n)$
$\operatorname{Sym}^+(n)$	$\operatorname{Sym}^+(n)$	$n$	1	1
$\operatorname{Sym}^+(n)$	$\operatorname{Sym}^+(n, k)$	$k$	1	1
$\operatorname{Sym}^+(n, k)$	$\operatorname{Sym}^+(n, k)$	$k$	1	1
		$k-1$	2	$\infty$
		$< k-1$	$\infty$	$\infty$
$\operatorname{Sym}^+(n, k)$	$\operatorname{Sym}^+(n, l)$	$l$	1	1
		$< l$	$\infty$	$\infty$

Table 6.1: Number of Bures-Wasserstein minimizing geodesics ( $n \geq k \geq l \geq r$ )

*Proof of Theorem 6.4.* 1. Without loss of generality, let us choose  $X, Y \in \operatorname{Mat}(n)$  with an convenient form. We choose  $X_0 \in \mathbb{R}_*^{n \times k}$  and  $Y_0 \in \mathbb{R}_*^{n \times l}$  as in the proof of the previous lemma. Given a singular value decomposition of  $X_0^\top Y_0 = U_k D V_l^\top$  with  $U_k \in \operatorname{O}(k)$ ,  $V_l \in \operatorname{O}(l)$  and  $D = \operatorname{Diag}(D_r, 0)$  with  $D_r \in \operatorname{Diag}^+(r)$ , we define  $P = \operatorname{Diag}(U_k, I_{n-k}) \in \operatorname{O}(n)$  and  $Q = \operatorname{Diag}(V_l, I_{n-l}) \in \operatorname{O}(n)$ . We choose  $X = [X_0 \ 0]P = [X_k \ 0]$  with  $X_k = X_0 U_k \in \mathbb{R}_*^{n \times k}$  and  $Y = [Y_0 \ 0]Q = [Y_l \ 0]$  with  $Y_l = Y_0 V_l \in \mathbb{R}_*^{n \times l}$ . Therefore we have  $XX^\top = \Sigma$ ,  $YY^\top = \Lambda$ ,  $X = [X_k \ 0]$ ,  $Y = [Y_l \ 0]$  and  $X^\top Y = \operatorname{Diag}(D_r, 0)$ . We denote  $X = [X_r \ X_{k-r} \ 0]$  and  $Y = [Y_r \ Y_{l-r} \ 0]$  with  $X_r, Y_r \in \mathbb{R}_*^{n \times r}$ ,  $X_{k-r} \in \mathbb{R}_*^{n \times (k-r)}$  and  $Y_{l-r} \in \mathbb{R}_*^{n \times (l-r)}$ .

Necessarily,  $H = (X^\top Y Y^\top X)^{1/2} = \operatorname{Diag}(D_r, 0)$ . The possible  $R \in \operatorname{O}(n)$  such that  $HR = X^\top Y$  are  $R = \operatorname{Diag}(I_r, R_{n-r})$  with  $R_{n-r} = [R_{l-r} \ R_{n-l}] \in \operatorname{O}(n-r)$  where  $R_{l-r} \in \operatorname{St}(n-r, l-r)$  and  $R_{n-l} \in \operatorname{St}(n-r, n-l)$ . We denote  $R_{l-r} = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix}$  with  $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$  and  $R_1 \in \mathbb{R}^{(n-k) \times (l-r)}$  with  $R_0^\top R_0 + R_1^\top R_1 = I_{l-r}$ . Note that both  $R_0$  and  $R_1$  have more rows than columns.

A simple calculus gives  $XRY^\top = X_r Y_r^\top + X_{k-r} R_0 Y_{l-r}^\top$ . Given  $R, R'$  satisfying  $HR = HR' = X^\top Y$ , we have  $XRY^\top = XR'Y^\top$  if and only if  $R_0 = R'_0$  since  $X_{k-r}^\top X_{k-r}$  and  $Y_{l-r}^\top Y_{l-r}$  are invertible. We even have  $\operatorname{sym}(XRY^\top) = \operatorname{sym}(XR'Y^\top)$  if and only if  $R_0 = R'_0$ . Indeed if  $\operatorname{sym}(XRY^\top) = \operatorname{sym}(XR'Y^\top)$ , then  $X_{k-r}^\top (R_0 - R'_0) Y_{l-r}^\top = Y_{l-r}^\top (R'_0 - R_0)^\top X_{k-r}^\top$ . Since  $X_{k-r}^\top Y_{l-r} = 0$ , it suffices to multiply on the left by  $X_{k-r}^\top$  and on the right by  $Y_{l-r}$  to conclude that  $R_0 = R'_0$ . Thus there is a bijection between minimizing geodesic segments and submatrices  $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$  of  $R_{l-r} = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix} \in \operatorname{St}(n-r, l-r)$ . Since  $R_1$  has more rows than columns, any  $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$  such that  $R_0^\top R_0 \leq I_{l-r}$  can be completed by an appropriate  $R_1 = \begin{pmatrix} (I_{l-r} - R_0^\top R_0)^{1/2} \\ \mathbf{0}_{n-k-(l-r), l-r} \end{pmatrix}$ .

Therefore, the minimizing geodesic segments are in bijection with the matrices

- $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$  such that  $R_0^\top R_0 \leq I_{l-r}$ , that is the closed unit ball for the spectral norm  $\mathcal{B}_S(0, 1)$ .
2. When  $r = l$ , the component  $Y_{l-r}$  of  $Y$  is the empty matrix. In other words, the dependence of the minimizing geodesic on  $R_0$  vanishes so the minimizing geodesic is unique. In particular when  $k = n$ ,  $r = \text{rk}(\Sigma\Lambda) = \text{rk}(\Lambda) = l$ .
  3. On the contrary, when  $r < l$  (thus  $n > k \geq l$ ), there is an infinite number of convenient  $R_0$ 's. For example,  $R_0 = \begin{pmatrix} \cos \theta & \mathbf{0}_{1, l-r-1} \\ \mathbf{0}_{k-r-1, 1} & \mathbf{0}_{k-r-1, l-r-1} \end{pmatrix}$  and  $R_1 = \begin{pmatrix} \sin \theta & \mathbf{0}_{1, l-r-1} \\ \mathbf{0}_{l-r-1, 1} & I_{l-r-1} \\ \mathbf{0}_{n-k-(l-r), 1} & \mathbf{0}_{n-k-(l-r), l-r-1} \end{pmatrix}$ .
  4. Since  $R_0$  has more rows than columns,  $R_1$  may be null which means than  $R_0 \in \text{St}(k-r, l-r)$ , that is  $R_0^\top R_0 = I_{l-r}$ . A simple calculus shows than  $YR^\top = [Y_r \ Y_{l-r} R_0^\top \ \mathbf{0}_{n-k}]$ . Therefore, the curve  $c(t) = (1-t)X + tYR^\top$  has its  $n-k$  columns identically null so it has rank less than  $k$ . But it also has rank at least  $k$  because  $\text{rk}(\Sigma) = \text{rk}(X) = k$ . So  $c$  and  $\gamma_{\Sigma \rightarrow \Lambda}^{R_0}$  are of rank exactly  $k$  on  $[0, 1)$  (and on  $[0, 1]$  if  $l = \text{rk}(\Lambda) = k$ ).
  5. At the other extremity, there is  $R_0 = 0$  (and  $R_1 \in \text{St}(n-k, l-r)$ ). For  $r = l$ , it corresponds to the empty matrix. In this case,  $XRY^\top = X_r Y_r^\top$ . Inspired by the case of the unique geodesic in  $\text{Sym}^+(n, k)$  (with  $k = l = r$ ), we notice that  $XH^-X^\top Y Y^\top = XH^-HY^\top = X \text{Diag}(I_r, 0) Y^\top = X_r Y_r^\top = XRY^\top$ . Therefore, denoting  $X = UD^{1/2}V^\top$  with  $U, V \in \text{St}(n, k)$  and  $D \in \text{Diag}^+(k)$ , we have:

$$\begin{aligned}
XRY^\top &= XH^-X^\top \Lambda \\
&= UD^{1/2}V^\top ((VD^{1/2}U^\top \Lambda UD^{1/2}V^\top)^{1/2})^{-1} VD^{1/2}U^\top \\
&= UD^{1/2}((D^{1/2}U^\top \Lambda UD^{1/2})^{1/2})^{-1} D^{1/2}U^\top \\
&= UD^{1/2}U^\top ((UD^{1/2}U^\top \Lambda UD^{1/2}U^\top)^{1/2})^{-1} UD^{1/2}U^\top \\
&= \Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda.
\end{aligned}$$

Thus the minimizing geodesic writes:

$$\gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^{-1} \Sigma^{1/2} \Lambda).$$

When the geodesic is unique, i.e. when  $r = l$ , i.e. when  $R_0$  is the empty matrix, it has rank exactly  $k$  on  $[0, 1)$ .  $\square$

#### REFERENCES

- [1] D. ALEKSEEVSKY, A. KRIEGL, M. LOSIK, AND P. W. MICHOR, *The Riemannian geometry of orbit spaces. The metric, geodesics, and integrable systems*, Publicationes Mathematicae, 62 (2001), pp. 1–30.
- [2] V. ARSIGNY, P. FILLARD, X. PENNEC, AND N. AYACHE, *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*, Magnetic Resonance in Medicine, 56 (2006), pp. 411–421.
- [3] R. BHATIA, T. JAIN, AND Y. LIM, *On the Bures–Wasserstein distance between positive definite matrices*, Expositiones Mathematicae, 37 (2019), pp. 165–191.
- [4] L. J. BILLERA, S. P. HOLMES, AND K. VOGTMANN, *Geometry of the Space of Phylogenetic Trees*, Advances in Applied Mathematics, 27 (2001), pp. 733–767.
- [5] N. BOURBAKI, *General Topology: Chapters 1–4*, Springer, Berlin, Heidelberg, 1971.

- [6] M. R. BRIDSON AND A. HAEFLIGER, *Metric Spaces of Non-Positive Curvature*, vol. 319 of Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.
- [7] A. CALISSANO, A. FERAGEN, AND S. VANTINI, *Populations of Unlabeled Networks: Graph Space Geometry and Geodesic Principal Components*, 2020. MOX Report.
- [8] M. P. DO CARMO, *Riemannian Geometry*, Mathematics: Theory and Applications, Birkhäuser, 1992.
- [9] D. DOWSON AND B. LANDAU, *The Fréchet distance between multivariate normal distributions*, Journal of Multivariate Analysis, 12 (1982), pp. 450–455.
- [10] I. L. DRYDEN, A. KOLOYDENKO, AND D. ZHOU, *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, The Annals of Applied Statistics, 3 (2009), pp. 1102–1123.
- [11] A. FERAGEN, F. LAUZE, P. LO, M. DE BRUIJNE, AND M. NIELSEN, *Geometries on spaces of treelike shapes*, in Proceedings of the 10th Asian conference on Computer vision - Volume Part II, ACCV'10, Berlin, Heidelberg, Nov. 2010, Springer-Verlag, pp. 160–173.
- [12] A. FERAGEN AND T. NYE, *Chapter 8. Statistics on stratified spaces*, in Riemannian Geometric Statistics in Medical Image Analysis, X. Pennec, S. Sommer, and T. Fletcher, eds., Academic Press, Jan. 2020, pp. 299–342.
- [13] P. FILLARD, X. PENNEC, V. ARSIGNY, AND N. AYACHE, *Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics.*, IEEE Transactions on Medical Imaging, 26 (2007), pp. 1472–82.
- [14] P. T. FLETCHER AND S. JOSHI, *Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data*, Signal Processing, 87 (2007), p. 250–262.
- [15] M. K. GARBA, T. M. W. NYE, J. LUEG, AND S. F. HUCKEMANN, *Information geometry for phylogenetic trees*, Journal of Mathematical Biology, 82 (2021), p. 19.
- [16] M. HÀ QUANG, *Alpha procrustes metrics between positive definite operators: a unifying formulation for the bures-wasserstein and log-euclidean/log-hilbert-schmidt metrics*, Linear Algebra and its Applications, 636 (2022), pp. 25–68.
- [17] M. HÀ QUANG, M. SAN BIAGIO, AND V. MURINO, *Log-hilbert-schmidt metric between positive definite operators on hilbert spaces*, Advances in neural information processing systems, 27 (2014).
- [18] D. G. KENDALL, *Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces*, Bulletin of the London Mathematical Society, 16 (1984), pp. 81–121.
- [19] J. M. LEE, *Introduction to Smooth Manifolds*, Springer, New York, NY, 2nd ed., 2012.
- [20] C. LENGLET, M. ROUSSON, R. DERICHE, AND O. FAUGERAS, *Statistics on the Manifold of Multivariate Normal Distributions: Theory and Application to Diffusion Tensor MRI Processing*, Journal of Mathematical Imaging and Vision, 25 (2006), pp. 423–444.
- [21] P. LI, Q. WANG, H. ZENG, AND L. ZHANG, *Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (2017), pp. 803–817.
- [22] Z. LIN, *Riemannian Geometry of Symmetric Positive Definite Matrices via Cholesky Decomposition*, SIAM Journal on Matrix Analysis and Applications, 40 (2019), pp. 1353–1370.
- [23] L. MALAGÒ, L. MONTRUCCHIO, AND G. PISTONE, *Wasserstein Riemannian geometry of Gaussian densities*, Information Geometry, 1 (2018), pp. 137–179.
- [24] E. MASSART AND P.-A. ABSIL, *Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices*, SIAM Journal on Matrix Analysis and Applications, 41 (2020), pp. 171–198.
- [25] E. MASSART, J. 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 GSI 2019 - 4th conference on Geometric Science of Information, F. Nielsen and F. Barbaresco, eds., vol. 11712 of Lecture Notes in Computer Science, Toulouse, France, 2019, Springer International Publishing, pp. 739–748.
- [26] P. W. MICHOR, *Topics in Differential Geometry*, vol. 93 of Graduate Studies in Mathematics, American Mathematical Soc., 2008.
- [27] P. W. MICHOR, D. PETZ, AND A. ANDAI, *The Curvature of the Bogoliubov-Kubo-Mori Scalar Product on Matrices*, Infinite Dimensional Analysis, Quantum Probability and Related Topics, 3 (2000), pp. 1–14.
- [28] M. MOAKHER, *A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices*, SIAM Journal on Matrix Analysis and Applications, 26 (2005), pp. 735–747.
- [29] I. OLKIN AND F. PUKELSHEIM, *The distance between two random vectors with given dispersion matrices*, Linear Algebra and its Applications, 48 (1982), pp. 257–263.

- [30] B. O'NEILL, *The fundamental equations of a submersion.*, Michigan Mathematical Journal, 13 (1966), pp. 459–469.
- [31] F. PAULIN, *Groupes et géométries*, 2014.
- [32] X. PENNEC, P. FILLARD, AND N. AYACHE, *A Riemannian Framework for Tensor Computing*, International Journal of Computer Vision, 66 (2006), pp. 41–66.
- [33] D. PETZ AND G. TOTH, *The Bogoliubov inner product in quantum statistics*, Letters in Mathematical Physics, 27 (1993), pp. 205–216.
- [34] C. L. SIEGEL, *Symplectic Geometry*, American J. of Mathematics, 65 (1943), pp. 1–86.
- [35] L. T. SKOVGAARD, *A Riemannian Geometry of the Multivariate Normal Model*, Scandinavian Journal of Statistics, 11 (1984), pp. 211–223.
- [36] A. TAKATSU, *On Wasserstein geometry of Gaussian measures*, in Probabilistic Approach to Geometry, M. Kotani, M. Hino, and T. Kumagai, eds., vol. 57 of Advanced Studies in Pure Mathematics, Kyoto University, Japan, 2010, Mathematical Society of Japan, pp. 463–472.
- [37] A. TAKATSU, *Wasserstein geometry of Gaussian measures*, Osaka Journal of Mathematics, 48 (2011), pp. 1005–1026.
- [38] Y. THANWERDAS AND X. PENNEC, *Geodesics and Curvature of the Quotient-Affine Metrics on Full-Rank Correlation Matrices*, in Proceedings of GSI 2021 - 5th conference on Geometric Science of Information, vol. 12829 of Lecture Notes in Computer Science, Paris, France, July 2021, Springer International Publishing, pp. 93–102.
- [39] Y. THANWERDAS AND X. PENNEC,  *$O(n)$ -invariant Riemannian metrics on SPD matrices*. Preprint, Sept. 2022.
- [40] Y. THANWERDAS AND X. PENNEC, *Theoretically and computationally convenient geometries on full-rank correlation matrices*. Preprint, Jan. 2022.
- [41] J. VAN OOSTRUM, *Bures-Wasserstein Geometry*, 01 2020. ArXiv e-prints, submitted.