Geometric Characteristics of Wasserstein Metric on SPD(n)

Yihao Luo, Shiqiang Zhang, Yueqi Cao, Huafei Sun

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, P. R. China

Abstract: Wasserstein distance, especially among symmetric positive-definite matrices, has broad and deep influences on development of artificial intelligence (AI) and other branches of computer science. A natural idea is to describe the geometry of $SPD(n)$ as a Riemannian manifold endowed with the Wasserstein metric. In this paper, by involving the fiber bundle, we obtain explicit expressions for some locally geometric quantities, including geodesics, exponential maps, the Riemannian connection, Jacobi fields and curvatures. Furthermore, we discuss the behaviour of geodesics and prove that the manifold is globally geodesic convex with non-negative curvatures but no conjugate pair and cut locus. According to arithmetic estimates, we find curvatures can be controlled by the minimal eigenvalue.

Key Words: Symmetric positive-definite matrix; Wasserstein distance; geodesic; connection; curvature

1 Introduction

1.1 Background

The importance of symmetric positive-definite matrices [12, 21] is beyond words. Symmetric positive-definite matrices have wide usage in many fields of information science, such as stability analysis of signal processing, linear stationary systems, optimal control strategies, and imaging analysis [5, 11]. Due to the importance of covariance matrices, analyses about symmetric positive matrices induces series of more profound discussions for multivariate statistics[14]. Instead of

1

considering a single matrix, contemporary scientists tend to comprehend the global structure of the set consisting of all $n \times n$ symmetric positive-definite matrices. This set is known as $SPD(n)$.

The set $SPD(n)$ can be endowed with various structures. $SPD(n)$ has a natural differential structure to be a manifold since it can be regarded as an open subset in $\mathbb{R}^{\frac{n \times (n+1)}{2}}$. It is also a quotient space of the general linear group $GL(n)$ by the orthogonal group $\mathcal{O}(n)$ [17], denoted as $SPD(n) \cong \frac{GL(n)}{\mathcal{O}(n)}$. Further, in the sense of matrix exponential and matrix logarithm [12], $SPD(n)$ can be regard as a Lie group [18]. The group multiplication is defined by $Y_1 * Y_2 = \exp(\log Y_1 + \log Y_2)$ [15]. Various structures bring abundant geometric details. And from another sight, the family of n-dimensional Gaussian distributions can be equipped with distances and divergences. These measurements can be naturally carried onto $SPD(n)$. Therefore, there are some widely used measurements defined on $SPD(n)$, including Kullback-Leibler (KL) divergence, Jensen-Shannon (JS) divergence, and Wasserstein distance [16]. These measurements all contain deep physical background meanings and are extensively used in AI and data science. To balance above two ideas, scientists are trying to find some Riemannian metrics on $SPD(n)$. Series of metrics have been involved. The most traditional Euclidean metric implies the Euclidean distance among Gaussian distributions. X. Pennec, P. Fillard, et al. [17] defined the affine-invariant Riemannian metric. V. Arsigny, P. Fillard, et al. [15] showed the Lie group structure admitting a bi-invariant metric, which is called the Log-Euclidean metric. And most encouragingly, by constructing a principle bundle, M. Wong et al. [22] and S. Zhang et al. [8] gave a new Riemannian metric on $SPD(n)$ whose geodesic distance is equivalent to Wasserstein-2 distance.

Wasserstein distance [16] plays a key role in recent development of information science. This distance, called Earth-moving distance as well, describes the minimal price or energy required to move one distribution to another along an underlying manifold. Wasserstein distance is actually a family of distances depending on different definitions of energy:

$$W_p(Y_1, Y_2) = \inf_{\gamma \sim \Pi(Y_1, Y_2)} \left(E_{(x,y) \sim \gamma}[\|x - y\|^p] \right)^{\frac{1}{p}}. \tag{1.1}$$

These distances, especially Wasserstein-2 distance W_2 (p is chosen as 2 in followings), naturally reflect the metric structure of underlying manifold. Wasserstein distance has its deep and profound physical meanings. And compared to KL divergence and JS divergence, Wasserstein distance can measure the difference between two distributions with small overlapped supports. Therefore

Wasserstein distance has been regarded as an essential tool to access more advanced or explainable AI. For instance, A. Martin et al. solved the problems about the robustness of nets training by introducing Wasserstein distance into Generative Adversarial Networks (GAN) [2]. Even so, the general Wasserstein distance can not be expressed explicitly, unless on some specific distribution families. Fortunately, the distance among Gaussian distributions on $\mathbb{R}^n$ has a beautiful explicit form [9]

$$W(Y_1, Y_2) = \|\mu_1 - \mu_2\| + \left(\text{tr}[\Sigma_1 + \Sigma_2 - 2(\Sigma_1\Sigma_2)^{\frac{1}{2}}]\right)^{\frac{1}{2}}, \tag{1.2}$$

where μ, Σ are expectation and covariance of distribution Y. Because Wasserstein distance is originally defined by a low bound of energy, whether a real moving process among Gaussian distribution exists is a question omitted usually. We will show a geometric solution for this question in Section 2.

When the expectation fixed, [6] proved the distance can be geometrized into a Riemannian metric on $SPD(n)$, called Wasserstein metric g_w, which can also be constructed via regarding $GL(n)$ as a fiber bundle on $SPD(n)$. In other words, Wasserstein distance is actually the geodesic distance based on this new metric constructed by the fiber bundle (or Riemannian submersion). This amazing result is proved by [3], but some analytic details have been omitted. In this paper, from an original viewpoint , we reprove the facts and fill those essential blanks up. About the Wasserstein geometry, some geometric quantities have been calculated in some way [3, 4, 8], relying on conventional and complicated geometric computing. Due to the fiber bundle structure [19, 20], we present succinct expressions for geodesics, the exponential and the Riemannian connection. In the following, we give more geometric quantities, including Jacobi fields and curvatures. One of our main goals is to present this manifold more completely.

1.2 Notation

In the whole paper, we adopt conventional notations in algebra and geometry. Riemannian manifolds are denote as pairs of '(manifold, metric)'. For example, our main study object is $(SPD(n), g_w)$, meaning $SPD(n)$ endowed with Wasserstein metric. $\mathbb{R}^n$ is the n-dimensional Euclidean space. $Sym(n)$ means the set of $n \times n$ symmetric matrices. $T_X M$ is conventionally the tangent space of M at a point X. Λ always represents a diagonal $n \times n$ matrix. For an $n \times n$ matrix Y, $\lambda(Y)$ or $\lambda_i(Y)$ means a eigenvalue or the i-th eigenvalue of Y, respectively. Then $x_i \sim \lambda_i$ is a

pair of eigenvalues and one of its associated eigenvectors. The components of matrix A will always be noted as $[A_{ij}]$. And identity matrix is denoted as I. In this paper, we tends to express points on manifolds by X, Y while vector fields by A, B.

1.3 Overview

The paper is organized as follows. In the section 2, we introduce some basic knowledge of the Riemannian manifold $(SPD(n), g_W)$, including some definitions and a important Algorithm. We will consider the symmetry of the $(SPD(n), g_W)$ as well. In the section 3, we focus on geodesic. we prove the geodesic convexity and give the exponential and Wasserstein radius, which allows us to depict the manifold in some extent. In the section 4, we express the Riemannian connection explicitly by comparing the total space and base space of the bundle structure. Section 5 is about Jacobi fields. In this part, we prove the non-existence of conjugate pair and cut locus, which shows the uniqueness of geodesic. Finally, in the section 6, we study curvatures. We obtain explicit expressions for the Riemannian curvature, sectional curvatures and the scalar curvature. Furthermore, we give the estimate of sectional curvatures, showing that the curvatures can be controlled by the minimal eigenvalue of points on $(SPD(n), g_W)$.

2 Preliminary

2.1 Algebra about Sylvester Equation

Sylvester equation is one of the most classical matrix equations. The solution of this equation plays a key role to understand the geometry of $(SPD(n), g_w)$. First of all, we shall list series of properties of this equation, which are utilized in the following arguments frequently. Then we recall an useful algorithm to calculate the explicit form of the solution. This algorithm paves a way for calculating and estimating the Wasserstein curvatures. Here we denote the solution of Sylvester equation with Γ, which means:

$$Y\Gamma_Y A + \Gamma_Y AY = A, \qquad \forall Y \in SPD(n). \tag{2.3}$$

Our discussion does not contain the well-posedness of Sylvester equation. From geometric aspects, we can ensure the existence and uniqueness of the solution in the particular case involved

in this paper. More details about Sylvester equation are claimed in [13]. Here we list some basic properties which will be used frequently in the following discussions.

Proposition 2.1.

1. $\Gamma_Y(A + kB) = \Gamma_Y A + k\Gamma_Y B, \ \forall k \in \mathbb{R}.$

2. $\Gamma_{kY} A = \frac{1}{k}\Gamma_Y A, \ \forall k \in \mathbb{R}.$

3. $\Gamma_{Y+\sigma} A = \Gamma_Y A - \Gamma_Y(\sigma\Gamma_{Y+\sigma}A + \Gamma_{Y+\sigma}A\sigma).$

4. $\Gamma_Y(YA) = Y\Gamma_Y A \ , \ \Gamma_Y(AY) = \Gamma_Y AY.$

5. $\Gamma_{Y^{-1}} A = \Gamma_Y(YAY) = Y\Gamma_Y AY.$

6. $\Gamma_{QYQ^{-1}} QAQ^{-1} = Q\Gamma_Y AQ^{-1}, \ \forall Q \in GL(n).$

Proposition 2.1 will determine the geometry on $(SPD(n), g_W)$ and be involved into every calculation repetitively and alternatively. We omit the proof since these properties are easy-checked.

Then we recall an algorithm to solve this kind of Sylvester equations, which offers an explicit expression of the solution. And this expression only depends on the eigenvalue decomposition. More details can be found in [8].

Algorithm 1 Solution of Sylvester equation

Input: $Y \in SPD(n)$, $A \in Sym(n)$
Output: $\Gamma_Y A$
Eigenvalue decomposition: $Y = Q\Lambda Q^T$, where $Q \in \mathcal{O}(n)$, $\Lambda := \text{diag}[\lambda_1...\lambda_n]$
$C := [C_{ij}] = Q^T AQ$
$E := [E_{ij}] = [\frac{C_{ij}}{\lambda_i + \lambda_j}]$
return $\Gamma_Y A = QEQ^T$

The algorithm will be simplified if Y is specially a diagonal matrix Λ. This algorithm will be used frequently in the following passage, which helps us to comprehend the geometry of $SPD(n)$.

2.2 Wasserstein Metric

In this part, we will recall some concepts about Riemannian geometry and fiber bundles, and then introduce the Wasserstein metric on $SPD(n)$.

Definition 2.1. For a principal bundle $\left(\tilde{M}, \tilde{g}\right)$ with the structure group G and projection σ onto the underlying manifold M, any vector $A \in T_X \tilde{M}$ is **horizontal**, if and only if $\tilde{g}\left(\tilde{A}, \tilde{V}\right) = 0$, for all $\tilde{V} \in T_X \tilde{M}$ such that $d\sigma\left(\tilde{V}\right) = 0$. We say $A \perp G[X]$, where $G[X]$ is the orbit of X under the group action. And if $\mathrm{d}\sigma\left(\tilde{A}\right) = A \in T_{\sigma(X)} M$, we call $\tilde{A}$ a **lift** of A.

Lemma 2.2. *The general linear group with Euclidean metric* $(GL(n), g_E)$ *and* $\sigma(X) = X^T X$, *where* $g_E(A, B) := \mathrm{tr}\left(A^T B\right)$, *for any* $A, B \in T_X GL(n)$, *is a trivial principal bundles on* $SPD(n)$, *with orthogonal group* $\mathcal{O}(n)$ *as the structure group.*

This lemma establishes for two facts [8]: $SPD(n) \cong \frac{GL(n)}{\mathcal{O}(n)}$ and g_E keeps invariant under the group action of $\mathcal{O}(n)$.

Lemma 2.3. *For any* $X \in (GL(n), g_E)$, $Y := \sigma(X) = X^T X$, *and any* $W \in T_Y SPD(n)$, *there is a unique* $\widetilde{W}$ *to be the horizontal lift of* W *at* $T_X GL(n)$, *satisfying*

$$\widetilde{W} = X \Gamma_Y W. \tag{2.4}$$

Proof. By definition, horizontal lift subjects to two conditions

$$\begin{cases} \widetilde{W}^T X + X^T \widetilde{W} = W \quad \left(\mathrm{d}\sigma\left(\widetilde{W}\right) = W\right), \\ \widetilde{W} \perp \mathcal{O}(n)[X] \Rightarrow X^{-1}\widetilde{W} = \widetilde{W}^T X^{-T}. \end{cases}$$

Thus, we have

$$X^T X X^{-1} \widetilde{W} + X^{-1} \widetilde{W} X^T X = W,$$

which is equivalent to (2.4). $\qquad\square$

Definition 2.2. For any $Y \in SPD(n)$, $V, W \in T_Y SPD(n)$, we define

$$g_W|_Y (V, W) = \mathrm{tr}\left(\Gamma_Y V Y \Gamma_Y W\right) = \frac{1}{2}\mathrm{tr}\left(\Gamma_Y V W\right). \tag{2.5}$$

g_W is a symmetric and non-degenerated bilinear tensor fields on $SPD(n)$, we call g_W as Wasserstein metric.

Lemma 2.4. *The projection* $\sigma : (GL(n), g_W) \to (SPD(n), g_E)$ *is a Riemannian submersion* [7],

which means dσ *is surjective and*

$$g_E\left(\tilde{A}, \tilde{B}\right) = g_W\left(\mathrm{d}\sigma\left(\tilde{A}\right), \mathrm{d}\sigma\left(\tilde{B}\right)\right) = g_W\left(A, B\right). \tag{2.6}$$

2.3 Symmetry

Now we study the symmetry of $(SPD\,(n)\,, g_W)$. We consider the invariability of Wasserstein metric under a special group action, and we claim that the orthogonal group $\mathcal{O}\,(n)$ is isomorphic to a subgroup of the isometry group $ISO\,(SPD\,(n)\,, g_W)$.

Definition 2.3. The orthogonal action $\Psi : \mathcal{O}\,(n) \times SPD\,(n) \to SPD\,(n)$ is defined by

$$\Psi_O\,(Y) = OYO^T, \quad \forall O \in \mathcal{O}\,(n)\,. \tag{2.7}$$

Lemma 2.5. *The orthogonal group $\mathcal{O}\,(n)$ is isomorphic to a subgroup of the isometric group of* $(SPD\,(n)\,, g_W)$, *as*

$$\{\Psi_O\,(Y)\}_{O \in \mathcal{O}} \vartriangleleft ISO|g_W. \tag{2.8}$$

Proof. First, we check that Ψ is a group action of $\mathcal{O}\,(n)$,

$$\Psi_I Y = Y = \mathrm{id}\,(Y)\,,$$

$$\Psi_{O_1 O_2^{-1}} Y = O_1 O_2^T Y O_2 O_1^T = \Psi_{O_1} \circ \Psi_{O_2^{-1}} Y.$$

Then we show that the tangent maps of these actions $\mathrm{d}\Psi_O$ are isometric,

$$\langle \mathrm{d}\Psi_O\,(A)\,, \mathrm{d}\Psi_O\,(B)\rangle|_{OYO^T}$$
$$= \frac{1}{2}\mathrm{tr}\left(\Gamma_{OYO^T}\left(OAO^T\right)OBO^T\right)$$
$$= \frac{1}{2}\mathrm{tr}\left(B\Gamma_Y A\right) = \langle A, B\rangle|_Y.$$

Thus, we prove the invariability of Wasserstein metric. □

We have such a fact that for an n-dimensional Riemannian manifold, the dimension of isometry group achieves maximum if and only if it has constant sectional curvature. Therefore, in section 5, we will show that $(SPD\,(n)\,, g_W)$ has no constant sectional curvature, which means its

7

symmetry degree less than the highest. The famous interval theorem [1] about isometric group shows the nonexistence of isometric groups with dimension between $\frac{m(m-1)}{2} + 1$ and $\frac{m(m+1)}{2}$, for any m-dimensional Riemannian manifold, where $m \neq 4$. On the other hand, (2.8) shows that the dimension of Wasserstein isometric group is higher than the dimension of $\mathcal{O}(n)$. Therefore, by $\dim(SPD(n)) = \frac{n^2+n}{2} \neq 4$ and $\dim(\mathcal{O}(n)) = \frac{n^2-n}{2}$, we have the following result.

Corollary 2.6. $(SPD(n), g_W)$ *has its symmetry degree controlled by*

$$\frac{1}{2}(n-1)n \leq \dim(ISO|g_W) \leq \frac{1}{8}(n-1)n(n+1)(n+2)+1. \tag{2.9}$$

According to Lemma 2.5, when we tend to study local geometric characteristics, we only need to consider the sorted diagonal matrices as the representational elements under the orthogonal action rather than all general points on $SPD(n)$. Therefore, some pointwise quantities, such as the scalar curvature and the bounds of sectional curvatures, depend only on eigenvalues.

3 Geodesic

Although prior results [4] presented some expressions of geodesics and the Riemannian exponential, the well-posedness and uniqueness of geodesic equations have not been considered. Additionally, some questions, such as the domain of the exponential, extensions of geodesics, and the existence of the minimal geodesic jointing two arbitrary points, are waiting to be answered. In this section, we shall prove that $(SPD(n), g_W)$ is geodesic convex [19]. Further, we calculate the maximal length for each geodesic, and then give the results about Wasserstein radius. Meanwhile, we will reprove some known results [4], including the expression of geodesics and the exponential, from an original viewpoint.

3.1 Geodesic Convexity

Wasserstein metric is not complete (we will see it later), which takes lots of trouble for researching the geometry of $(SPD(n), g_W)$. However, we tend to show that the whole Riemannian manifold $(SPD(n), g_W)$ is geodesic convex, which means that we can always find the minimal geodesic jointing any two points. To some extent, geodesic convexity may make up for the incompleteness.

Definition 3.1. A curve $\gamma(t)$ in the bundle $\left(\widetilde{M}, g\right)$ is said to be level, if and only if, $\dot{\gamma}(t)$ is horizontal in $T_{\gamma(t)}\widetilde{M}$ for any t.

Theorem 3.1. *For any* $Y_1, Y_2 \in SPD(n)$, *and* $Y_1^{\frac{1}{2}}$ *as the fixed lift of* Y_1, *there exists* $\widetilde{Y_2} \in GL(n)$ *as a lift of* Y_2,

$$\widetilde{Y_2} = Y_1^{-\frac{1}{2}}(Y_1 Y_2)^{\frac{1}{2}}, \tag{3.10}$$

such that the line segment $\gamma(t) = t\widetilde{Y_2} + (1-t)Y_1^{\frac{1}{2}}$, $t \in [0,1]$ *is **level** and **non-degenerated**,*

Before its proof presented, we show that Theorem 3.1 brings some geometrical and physical facts.

Corollary 3.2. *(**geodesic convexity**)* $(SPD(n), g_W)$ *is a geodesic convex Riemannian manifold. Between any two points* $Y_1, Y_2 \in SPD(n)$, *there exists a minimal Wasserstein geodesic*

$$\gamma(t) = (1-t)^2 Y_1 + t(1-t)[(Y_1 Y_2)^{\frac{1}{2}} + (Y_2 Y_1)^{\frac{1}{2}}] + t^2 Y_2, \ t \in [0,1], \tag{3.11}$$

where $\gamma(t)$ *lies on* $SPD(n)$ *strictly.*

Let us recall the definition of Wasserstein distance (1.1). The lower energy bound does not ensure the existence of a real path to "earth-move" one distribution to another with the minimal price. However, by Theorem 3.1 and corollary 3.2, we can actually ensure the existence of the path among Gaussian distributions with zero expectation .

To prove Theorem 3.1, we shall give two lemmas at first.

Lemma 3.3. *For any two matrices* $X_1, X_2 \in GL(n)$, $l(t) = tX_2 + (1-t)X_1$, $t \in [0,1]$. $\dot{l}(t)$ *is horizontal if and only if* $\det[l(t)] > 0$ *and* $\dot{l}(0)$ *is horizontal at* $T_{X_1}GL(n)$.

Proof. According to (2.2), $\dot{l}(0) = X_2 - X_1$ is horizontal at $T_{X_1}GL(n)$

$$\Leftrightarrow X_1^{-1}(X_2 - X_1) \in Sym(n) \Leftrightarrow X_1^{-1}X_2 \in Sym(n)$$
$$\Leftrightarrow X_2^{-1}X_1 \in Sym(n) \Leftrightarrow X_2^{-1}(X_2 - X_1) \in Sym(n).$$

Thus, $\dot{l}(1)$ is also horizontal at $T_{X_2}GL(n)$. For any s such that $l(s) \in GL(n)$, we constrict the line segment by $l_s(t) := l(st)$, and the lemma can be induced by horizontal of $\dot{l}_s(1)$. $\square$

Lemma 3.3 converts the levelness of a line segment into the horizontal of its initial vector, which brings quite convenience for the following proof.

Lemma 3.4. *For any $Y_1, Y_2 \in SPD(n)$, there exists $P = Y_1^{-\frac{1}{2}}(Y_1Y_2)^{\frac{1}{2}}Y_2^{-\frac{1}{2}} \in \mathcal{O}(n)$ such that $\gamma(t) = tPY_2^{\frac{1}{2}} + (1-t)Y_1^{\frac{1}{2}}$. When $\det(\gamma(t)) > 0$, $\dot{\gamma}(t)$ keeps horizontal at $T_{\gamma(t)}GL(n)$.*

Proof. First, we have

$$
\begin{aligned}
P^T P &= Y_2^{-\frac{1}{2}}(Y_2Y_1)^{\frac{1}{2}}Y_1^{-1}(Y_1Y_2)^{\frac{1}{2}}Y_2^{-\frac{1}{2}} \\
&= Y_2^{-\frac{1}{2}}Y_1^{-1}(Y_1Y_2)^{\frac{1}{2}}(Y_1Y_2)^{\frac{1}{2}}Y_2^{-\frac{1}{2}} \\
&= I.
\end{aligned}
\tag{3.12}
$$

Then we only need to check that $Y_1^{-\frac{1}{2}}\dot{\gamma}(0) \in Sym(n)$. Due to Lemma 3.3, we have

$$
\begin{aligned}
Y_1^{-\frac{1}{2}}\dot{\gamma} &= Y_1^{-1}\left((Y_1Y_2)^{\frac{1}{2}} - I\right) \in Sym(n) \\
&\Leftrightarrow Y_1^{-1}(Y_1Y_2)^{\frac{1}{2}} \in Sym(n) \Leftrightarrow Y_1^{-1}(Y_1Y_2)^{\frac{1}{2}} = (Y_2Y_1)^{\frac{1}{2}}Y_1^{-1},
\end{aligned}
\tag{3.13}
$$

which proves the lemma. $\qquad\square$

Remark. In fact, finding the orthogonal matrix P is equivalent to solving an Riccati equation. One can see [4] for details.

To prove Theorem 3.1, the last step is to clarify the non-degeneration.

Proof. (Theorem 3.1) Because of the symmetry from (2.8), it suffices to prove the case $Y_1 = \Lambda$. We shall prove $\det[\gamma(t)] > 0, \forall\, t \in [0,1]$. We have

$$
\begin{aligned}
D(t) &:= \det[\gamma(t)] \\
&= \det[t\Lambda^{-\frac{1}{2}}(\Lambda Y_2)^{\frac{1}{2}} + (1-t)\Lambda^{\frac{1}{2}}] \\
&= \det[t\Lambda^{-\frac{1}{2}}(\Lambda Y_2)^{\frac{1}{2}} + (1-t)\Lambda^{\frac{1}{2}}] \\
&= (1-t)^n \det\left(\Lambda^{\frac{1}{2}}\right)\det[I + \frac{t}{(1-t)}\Lambda^{-1}(\Lambda Y_2)^{\frac{1}{2}}].
\end{aligned}
\tag{3.14}
$$

Let $E(t) := \det[I + \frac{t}{(1-t)}\Lambda^{-1}(\Lambda Y_2)^{\frac{1}{2}}]$. Since $D(0), D(1) > 0$ and $\lim_{t\to 1^-}(1-t)^n = 0^+$, we see that $D(t) > 0 \Longleftrightarrow E(t) > 0$, $\forall t \in (0,1)$. Let $M := \Lambda^{-1}(\Lambda Y_2)^{\frac{1}{2}}$. By choosing pairs $\{\bar{x}_i \sim \lambda_i\}$ of

10

$(\Lambda Y_2)^{\frac{1}{2}}$ as an orthonormal basis in $\mathbb{R}^n$, we have

$$\lambda_i > 0 \implies \bar{x}_i^T M \bar{x}_i = \lambda_i \bar{x}_i^T \Lambda^{-1} \bar{x}_i > 0.$$

Thus M is positive definite. Then $E(t) = \det\left(I + \frac{t}{(1-t)} M\right) > 0$, for any $t \in [0,1]$. To sum up, the line segment $\gamma(t)$ is non-degenerated.

Combining with Lemma 3.4, the proof is done. $\qquad\square$

3.2 Exponential and Radius

Following Lemma 3.3, we can directly write down the Wasserstein logarithm on $SPD(n)$,

$$\log_{Y_1} Y_2 = \mathrm{d}\sigma\big|_{Y_1^{\frac{1}{2}}} \dot{\gamma}(0) = (Y_1 Y_2)^{\frac{1}{2}} + (Y_2 Y_1)^{\frac{1}{2}} - 2Y_1. \tag{3.15}$$

By solving the inverse problem of above equation, we gain the expression of the **Wasserstein exponential**.

Theorem 3.5. *In a small open ball $\mathcal{B}(0,\varepsilon)$, $\varepsilon > 0$ in $T_Y SPD(n) \cong \mathbb{R}^{\frac{1}{2}n(n+1)}$, the Wasserstein exponential at Y, $\exp_Y : \mathcal{B}(0,\varepsilon) \to SPD(n)$ is explicitly defined by*

$$\exp_Y A = Y + A + \Gamma_Y A Y \Gamma_Y A. \tag{3.16}$$

Proof. By choosing the normal coordinates [19] at Y, there always exist neighborhoods where $\exp_Y$ is well-defined. From (3.15), given $\exp_Y A$ well defined, then it satisfies

$$(Y \exp_Y A)^{\frac{1}{2}} + (\exp_Y AY)^{\frac{1}{2}} = A + 2Y.$$

This equation can convert to the Sylvester equation and we can express its solution as

$$\begin{aligned}
& YY^{-1}(Y \exp_Y A)^{\frac{1}{2}} + Y^{-1}Y(\exp_Y AY)^{\frac{1}{2}} Y^{-1}Y = A + 2Y \\
\Leftrightarrow\; & Y[Y^{-1}(Y \exp_Y A)^{\frac{1}{2}}] + [Y^{-1}(Y \exp_Y A)^{\frac{1}{2}}]Y = A + 2Y \\
\Leftrightarrow\; & Y\Gamma_Y(A + 2Y) = (Y \exp_Y A)^{\frac{1}{2}}.
\end{aligned}$$

Therefore, we have

$$\exp_Y A = \Gamma_Y \left(A + 2Y\right) Y \Gamma_Y \left(A + 2Y\right)$$

$$= \left(\Gamma_Y A + I\right) Y \left(\Gamma_Y A + I\right)$$

$$= Y + A + \Gamma_Y A Y \Gamma_Y A,$$

which finishes this proof. $\qquad\square$

Remark. We call the first two terms $Y + A$ as the Euclidean exponential, and the last term $\Gamma_Y A Y \Gamma_Y A$ as the Wasserstein correction for this bend manifold.

Corollary 3.6. *The geodesic equations with initial conditions $\gamma(0), \dot{\gamma}(0)$ on $\left(SPD\left(n\right), g_W\right)$ has the following explicit solution*

$$\gamma\left(t\right) = \gamma\left(0\right) + t\dot{\gamma}\left(0\right) + t^2 \Gamma_{\gamma(0)} \dot{\gamma}\left(0\right) \gamma\left(0\right) \Gamma_{\gamma(0)} \dot{\gamma}\left(0\right), \ t \in \left(-\varepsilon, \varepsilon\right). \tag{3.17}$$
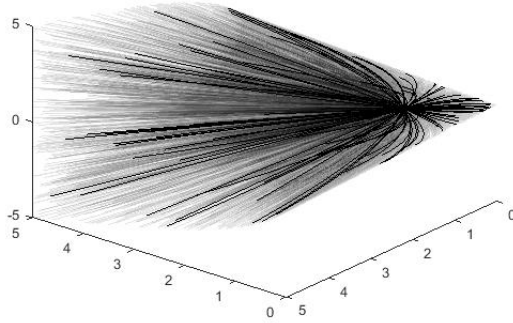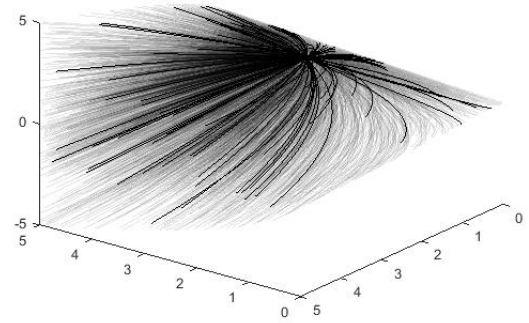

Figure 3-1 geodesics starting from [1,0;0,1]   Figure 3-2 geodesics starting from [3,1;1,1]

Subsequently, the next natural question is the maximal length of the extension of a geodesic. This question is equivalent to the largest domain of the exponential. We still focus on diagonal matrices.

Theorem 3.7. *For any $Y \in SPD\left(n\right)$ and $A \in T_Y SPD\left(n\right)$, $\exp_Y\left(tA\right) : [0, \varepsilon) \to SPD\left(n\right)$ is*

well-defined if and only if

$$\varepsilon_{max} = \begin{cases} -\frac{1}{\lambda_{min}}, & \text{if } \lambda_{min} < 0, \\ +\infty, & \text{if } \lambda_{min} \geq 0, \end{cases} \tag{3.18}$$

where λ_{min} is the minimal eigenvalue of $\Gamma_Y A$.

Proof. Evidently, $\varepsilon_{max} = \min\{s > 0 | \det[\exp_Y(sA) = 0]\}$. By (3.2), we have

$$\det[\exp_Y(sA)] = \det(Y)\det^2(I + s\Gamma_Y A) = 0$$

$$\Leftrightarrow \det\left(\Gamma_Y A - \frac{-1}{s}I\right) = 0$$

$$\Leftrightarrow s = -\frac{1}{\lambda(\Gamma_Y A)},$$

where $\lambda(\Gamma_Y A)$ is the eigenvalue of $\Gamma_Y A$. Thus $\varepsilon_{max} = \min\{-\frac{1}{\lambda(\Gamma_Y A)} > 0\}$. $\qquad\square$

Corollary 3.8. *Wasserstein metric g_W on $SPD(n)$ is incomplete.*

Corollary 3.8 can be directly obtained from Hopf-Rinow theorem [20].

Theorem 3.7 and the next theorem help us to comprehend the size of (SPD, g_W) from sense of each point.

Figure 3-1 and Figure 3-2 show geodesics starting from different origins on $SPD(2)$. From this group of pictures, we can observe the outline of the manifold and some behaviors of geodesics.

Definition 3.2. For any $Y \in SPD(n)$, call $r(Y)$ the Wasserstein **radius** of $SPD(n)$ at Y, if and only if $\exp_Y$ well-defined in all $\mathcal{B}(0, \varepsilon)$, $0 \leq \varepsilon \leq r(Y)$.

Theorem 3.9. *The Wasserstein radius $r(Y) : SPD(n) \to (0, +\infty)$ can be given by*

$$r(Y) = \sqrt{\lambda_{min}(Y)/2}. \tag{3.19}$$

The function $r(Y)$ is continuous.

Proof. Again, with (2.8), we have $r(Y) = r(\Lambda)$ if Y have diagonalization $Y = Q\Lambda Q^T$. Thus we still focus on $r(\Lambda)$. Following the definition and discussions around (3.18),

$$r(\Lambda) := \inf\{\varepsilon_{max}(V) | V \in T_\Lambda SPD(n), \|V\|_{g_W} = 1\}. \tag{3.20}$$

13

Equivalently, (3.20) has its dual expression, according to (3.2), we have

$$r\left(\Lambda\right) = \inf\{\|V\|_{g_W} | V \in T_\Lambda SPD\left(n\right), \det\left(I + \Gamma_\Lambda V\right)\}, \tag{3.21}$$

$$\Leftrightarrow 2r^2\left(\Lambda\right) = \inf\{\text{tr}\left(V\Gamma_\Lambda V\right) | \Gamma_\Lambda V \text{ has eigenvalue } \eta_k = -1\}. \tag{3.22}$$

We tend to solve (3.22). To avoid misconception, we denote $\{\lambda_i > 0\}$ as eigenvalues of Λ, and η_j as eigenvalues of $\Gamma_\Lambda V$. We choose the pairs $\{x_j \sim \eta_j\}$ of $\Gamma_\Lambda V$ as an orthonormal basis of $\mathbb{R}^n$. Then we consider the inner product $g_W\left(V, V\right)$, and by calculation we get

$$
\begin{aligned}
\text{tr}\left(V\Gamma_\Lambda V\right) &= \sum_{i=1}^n x_i{}^T V\Gamma_\Lambda V x_i \\
&= \sum_{i=1}^n \eta_i x_i{}^T V x_i = \sum_{i=1}^n \eta_i x_i{}^T \left(\Lambda\Gamma_\Lambda V + \Gamma_\Lambda V\Lambda\right) x_i \\
&= \sum_{i=1}^n \eta_i(\Gamma_\Lambda V x_i)^T \Lambda x_i + \eta_i x_i{}^T V\Gamma_\Lambda V x_i = \sum_{i=1}^n \eta_i^2 x_i{}^T \Lambda x_i.
\end{aligned}
$$

Because Λ is positive and $\eta_k = -1$, $\text{tr}\left(V\Gamma_\Lambda V\right) \geq \eta_k^2 x_k{}^T \Lambda x_k$. The equality $\text{tr}\left(V\Gamma_\Lambda V\right) = x_k{}^T \Lambda x_k$ holds if and only if $\eta_j = 0, \forall j \neq k$. In these cases, by Algorithm 1, we have

$$\left(\Gamma_\Lambda V\right)_{ij} = \frac{V_{ij}}{\lambda_i + \lambda_j} = \begin{cases} \eta_k = -1, & \text{if } i = j = k, \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$V_{ij} = -2\delta_{ik}\delta_{ik}\lambda_k; \ \left(x_k\right)_i = \delta_{ki},$$

and thus

$$\text{tr}\left(V\Gamma_\Lambda V\right) = x_k{}^T \Lambda x_k = \lambda_k.$$

Especially, we obtain λ_{min} as $\min\{\text{tr}\left(V\Gamma_\Lambda V\right)\}$, when $\Lambda_k = \lambda_{min}$, and $V_{ij} = -2\delta_{ik}\delta_{jk}\lambda_{min}$. The tangent vector V is called the speed-degenerated direction. The function $\lambda_{min}\left(Y\right)$ is certainly continuous, hence $r(Y)$ is also continuous. $\square$

Due to the geodesic convexity, the radius actually defines the wasserstein distance of a point on $SPD\left(n\right)$ to the 'boundary' of the manifold. It also measures the degenerated degree of a positive

symmetric matrix by $\sqrt{\lambda_{min}}$. Figure 3-3 shows three maximal geodisical balls with different centers on $SPD\left(2\right)$. From the viewpoint of $\mathbb{R}^3$, the three balls have different sizes in the sense of Euclidean distance, but on $\left(SPD\left(2\right), g_W\right)$ all of them have the radius of $\frac{\sqrt{2}}{2}$.

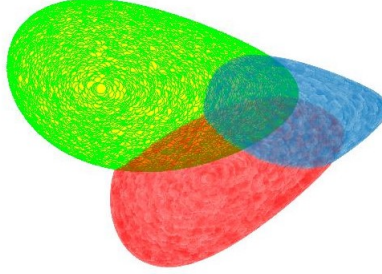Figure 3-3 three geodisical balls with same radius

4   Connection

In this section, we will study the Riemannian connection of $\left(SPD\left(n\right), g_W\right)$, called **Wasserstein connection**. The flatness of $\left(GL\left(n\right), g_E\right)$ and the structure of the Riemannian submersion will take series of convenience to our work.

During computation, we denote both tensor actions of g_W on $SPD\left(n\right)$ and g_E on $GL\left(n\right)$ by $\langle,.\rangle$. Then we denote the Euclidean connection as D, while the Wasserstein connection as ∇.

The main idea to express the Wasserstein connection is to compute the level decomposition of the Euclidean covariant derivative of lifted vector fields. In a word, we shall prove:

Theorem 4.1. *The Euclidean connection is a lift of the Wasserstein connection. For any smooth vector fields A and B on $SPD\left(n\right)$, and $\widetilde{A}$ and $\widetilde{B}$ are their level lifts, respectively, then the following equation holds*

$$\mathrm{d}\sigma|_X\left(D_{\widetilde{A}}\widetilde{B}\right) = \nabla_A B. \tag{4.23}$$

Before proving Theorem 4.1, we shall prove a key lemma which points the relation between the Lie-brackets on the total space and base space.

Lemma 4.2. *The level lift of vector fields commutes with Lie-brackets,*

$$d\sigma|_X[\widetilde{A}, \widetilde{B}] = [A, B]. \tag{4.24}$$

Proof. (Lemma 4.2) On the flat manifold $(GL(n), g_E)$, the connection equals to the usual derivative in the Euclidean space. Therefore, we have

$$
\begin{aligned}
D_{\widetilde{A}}\widetilde{B} &= \lim_{t\to 0} \frac{1}{t}[\widetilde{B}\left(X + \widetilde{A}t\right) - \widetilde{B}(X)] \\
&= \lim_{t\to 0} \frac{1}{t}[\left(X + \widetilde{A}t\right)\Gamma_{Y+At}B(Y + At) - X\Gamma_Y B(Y)] \\
&= \lim_{t\to 0} \frac{X}{t}[\Gamma_{Y+At}B(Y + At) - \Gamma_Y B(Y)] + X\Gamma_Y A\Gamma_Y B \\
&= X[\Gamma_Y dB(A) - \Gamma_Y(A\Gamma_Y B + \Gamma_Y BA) + \Gamma_Y A\Gamma_Y B].
\end{aligned} \tag{4.25}
$$

Putting $(\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A)$ into (2.3), we get

$$
\begin{aligned}
&\Gamma_Y(A\Gamma_Y B + \Gamma_Y BA - B\Gamma_Y A - \Gamma_Y AB) \\
&= (\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A) - 2\Gamma_Y(\Gamma_Y A\Gamma_Y BY - \Gamma_Y B\Gamma_Y AY).
\end{aligned} \tag{4.26}
$$

Then we compute the Lie-bracket

$$
\begin{aligned}
[\widetilde{A}, \widetilde{B}] &:= D_{\widetilde{A}}\widetilde{B} - D_{\widetilde{B}}\widetilde{A} \\
&= X[\Gamma_Y(dB(A) - dA(B))] + X[\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A] \\
&\quad - X\Gamma_Y(A\Gamma_Y B + \Gamma_Y BA - B\Gamma_Y A - \Gamma_Y AB) \\
&= \widetilde{[A, B]} + 2X\Gamma_Y(\Gamma_Y A\Gamma_Y BY - \Gamma_Y B\Gamma_Y AY),
\end{aligned} \tag{4.27}
$$

where the last equality in (4.27) comes from (4.26).

Finally we show the second term in (4.27) is vertical. In fact, we have

$$
\begin{aligned}
&d\sigma|_X(X\Gamma_Y(\Gamma_Y A\Gamma_Y BY - \Gamma_Y B\Gamma_Y AY)) \\
&= Y\Gamma_Y(\Gamma_Y A\Gamma_Y BY - \Gamma_Y B\Gamma_Y AY) + \Gamma_Y(Y\Gamma_Y B\Gamma_Y A - Y\Gamma_Y A\Gamma_Y B)Y \\
&= Y\Gamma_Y(\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A)Y + Y\Gamma_Y(\Gamma_Y B\Gamma_Y A - \Gamma_Y A\Gamma_Y B)Y \\
&= \Gamma_{Y^{-1}}[\Gamma_Y A\Gamma_Y BY - \Gamma_Y B\Gamma_Y AY + \Gamma_Y B\Gamma_Y AY - \Gamma_Y A\Gamma_Y BY] = 0.
\end{aligned} \tag{4.28}
$$

Thus the proof for Lemma 4.2 has been done. $\qquad\square$

By Lemma 4.2, the proof for Theorem 4.1 is clarified.

Proof. (Theorem 4.1) For any smooth vector field C on $SPD\,(n)$, and its level lift $\widetilde{C}$, we have

$$
\begin{aligned}
\langle \nabla_A B, C \rangle &= \langle \widetilde{\nabla_A B}, \widetilde{C} \rangle \\
&= \frac{1}{2} \left(A\langle B, C \rangle + B\langle C, A \rangle - C\langle A, B \rangle + \langle C, [A, B] \rangle + \langle B, [C, A] \rangle - \langle A, [B, C] \rangle \right) \\
&= \frac{1}{2} \left(\widetilde{A}\langle \widetilde{B}, \widetilde{C} \rangle + \widetilde{B}\langle \widetilde{C}, \widetilde{A} \rangle - \widetilde{C}\langle \widetilde{A}, \widetilde{B} \rangle + \langle \widetilde{C}, \widetilde{[A, B]} \rangle + \langle \widetilde{B}, \widetilde{[C, A]} \rangle - \langle \widetilde{A}, \widetilde{[B, C]} \rangle \right) \\
&= \frac{1}{2} \left(\widetilde{A}\langle \widetilde{B}, \widetilde{C} \rangle + \widetilde{B}\langle \widetilde{C}, \widetilde{A} \rangle - \widetilde{C}\langle \widetilde{A}, \widetilde{B} \rangle + \langle \widetilde{C}, [\widetilde{A}, \widetilde{B}] \rangle + \langle \widetilde{B}, [\widetilde{C}, \widetilde{A}] \rangle - \langle \widetilde{A}, [\widetilde{B}, \widetilde{C}] \rangle \right) \\
&= \langle D_{\widetilde{A}} \widetilde{B}, \widetilde{C} \rangle.
\end{aligned}
$$

By the arbitrariness of C, theorem 4.1 is proved. This proof implies that $d\sigma|_X \left(D_{\widetilde{A}} \widetilde{B} \right)$ is independent on X chosen among a fixed fiber. $\qquad\square$

Theorem 4.1 has a direct corollary which is one of essential results in this paper.

Corollary 4.3. *Wasserstein connection has an explicit expression:*

$$
\nabla_A B = dB\,(A) - \Gamma_Y A Y \Gamma_Y B - \Gamma_Y B Y \Gamma_Y A. \tag{4.29}
$$

Proof. From Theorem 4.1 and (4.27), we have

$$
\begin{aligned}
\nabla_A B &= d\sigma|_X \left(D_{\widetilde{A}} \widetilde{B} \right) \\
&= X^T D_{\widetilde{A}} \widetilde{B} + D_{\widetilde{A}}^T \widetilde{B} X \\
&= dB\,(A) - (A\Gamma_Y B + \Gamma_Y B A) + Y\Gamma_Y A \Gamma_Y B + \Gamma_Y B \Gamma_Y A Y \\
&= dB\,(A) - \Gamma_Y A Y \Gamma_Y B - \Gamma_Y B Y \Gamma_Y A.
\end{aligned}
$$

The linearity, Leibnitz's law and symmetry of Wasserstein connection are easy-checked from the expression. $\qquad\square$

Definition 4.1. The horizontal component of lifted covariant derivative of B over A is a vector

17

field in $GL(n)$ whose value at X is defined by

$$\mathcal{T}_X(A, B) := D_{\widetilde{A}}\widetilde{B} - \widetilde{\nabla_A B}. \tag{4.30}$$

The vector field is denoted as $\mathcal{T}(A, B)$.

Theorem 4.4. $\mathcal{T}_X(,)$ *is a antisymmetric tensor:* $T_Y SPD(n) \otimes T_Y SPD(n) \to T_X GL(n)$, *and it satisfies*

$$\mathcal{T}_X(A, B) = X\Gamma_Y(\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A)Y. \tag{4.31}$$

Proof. For any $V \in T_X GL(n)$, V has the orthogonal decomposition as

$$V = V_T + V_H := X^{-T}F + XS,$$

where S is a symmetric matrix, and F is antisymmetric. Hence,

$$V_T = X^{-T}\Gamma_{Y^{-1}}\left(X^{-1}V - V^T X^{-T}\right). \tag{4.32}$$

Combining (4.25) with (4.32), we have

$$\begin{aligned}
\mathcal{T}_X(A, B) &= X^{-T}\Gamma_{Y^{-1}}(\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A) \\
&= X\Gamma_Y(\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A)Y.
\end{aligned} \tag{4.33}$$

(4.33) shows that $\mathcal{T}_X(A, B)$ depends only on X and the vectors on $T_Y SPD(n)$. Meanwhile the multi-linearity and $\mathcal{T}_X(A, B) = -\mathcal{T}_X(B, A)$ are easy-checked. $\qquad\square$

Recalling (4.27), we can also find that

$$[\widetilde{A}, \widetilde{B}] = \widetilde{[A, B]} + 2\mathcal{T}(A, B).$$

In the following parts, we will show the tensor $\mathcal{T}(A, B)$ takes a fundamental role for computing curvature.

5   Jacobi Field

As the Wasserstein exponential is given in section 2.2 with the explicit form, we can clearly understand the behaviours of the geodesics. In this part, we will study the Wasserstein Jacobi fields on $SPD(n)$.

Come straight to the point. Jacobi fields can be directly constructed by the geodisical variation via the exponential.

Theorem 5.1. *Along a geodesic $\gamma(t)$ with $\gamma(0) = Y \in SPD(n)$, $\dot{\gamma}(0) = A \in T_Y SPD(n)$, there exists a unique normal Jacobi vector field $J(t)$ with initial conditions $J(0) = 0, \nabla_{\dot{\gamma}(0)} J(t) = B \in T_Y SPD(n)$, where $\langle A, B \rangle|_Y = 0$. We have*

$$J(t) = tB + t^2 \left(\Gamma_Y A Y \Gamma_Y B + \Gamma_Y B Y \Gamma_Y A \right). \tag{5.34}$$

As in [7] $J(t)$ is constructed by

$$J(t) := \left. \frac{\partial}{\partial s} \right|_{s=0} \exp_Y t(A + sB). \tag{5.35}$$

Substituting (3.16) into (5.35), Theorem 5.1 comes from direct computation.

Theorem 5.2. *There exists **no conjugate pair** on $(SPD(n), g_W)$.*

Proof. We prove by contradiction. Suppose that there exists a Jacobi field $J(t) \not\equiv 0$, lying on the geodesic $\gamma(t)$, where $\gamma(0) = Y$, $J(0) = 0$, and $p > 0$ such that $J(p) = 0$. We still denote $\nabla_{\dot{\gamma}(0)} J(t) := B$, $\dot{\gamma}(0) := A$. Thus, we have

$$J(p) = pB + p^2 \left(\Gamma_Y A Y \Gamma_Y B + \Gamma_Y B Y \Gamma_Y A \right) = 0,$$

and thus we have

$$- p \left(\Gamma_Y A Y \Gamma_Y B + \Gamma_Y B Y \Gamma_Y A \right) = B = Y \Gamma_Y B + \Gamma_Y B Y$$

$$\Leftrightarrow \Gamma_Y B Y (I + p \Gamma_Y A) + (I + p \Gamma_Y A) Y \Gamma_Y B = 0.$$

According to discussions around (3.2), we know that $(I + p \Gamma_Y A)$ is positive as long as $\gamma(p)$ is well-defined. Due to $B \neq 0$, and $\Gamma_Y B \neq 0$, we assume $\lambda \neq 0$ is an eigenvalue of $\Gamma_Y B$, and one of

19

its standard eigenvectors is denoted as x. Subsequently,

$$x^T \Gamma_Y BY (I + p\Gamma_Y A) x + x^T (I + p\Gamma_Y A) Y\Gamma_Y Bx$$

$$= \lambda x^T Y (I + p\Gamma_Y A) x + \lambda x^T (I + p\Gamma_Y A) Yx$$

$$= 2\lambda x^T Y (I + p\Gamma_Y A) x = 0,$$

which contradicts to $\|x\| = 1$ and the positive definiteness of $Y(I + p\Gamma_Y A)$. □

Theorem 5.3. *There exists a unique geodesic jointing any two points. Geometrically, there is no cut locus on any geodesic.*

Proof. From [8] we see that the shortest distance among two fibers in bundle equals to the length of the geodesics, which informs the geodesics defined as Corollary 3.2 are always the shortest. For any $Y_1, Y_2 \in SPD(n)$, $P_1, P_2 \in \mathcal{O}(n)$, the distance from the fiber $\sigma^{-1}(Y_1)$ to the fiber $\sigma^{-1}(Y_1)$ is defined as

$$\inf_{P_1, P_2 \in \mathcal{O}(n)} \left\| P_1 Y_1^{\frac{1}{2}} - P_2 Y_2^{\frac{1}{2}} \right\|.$$

From Lemma 3.3 and Lemma 3.4, due to the compactness of the structure group $\mathcal{O}(n)$, the distance will be achieved by the length of a level and non-degenerated line segment. Therefore, we have

$$\inf_{P_1, P_2 \in \mathcal{O}(n)} \left\| P_1 Y_1^{\frac{1}{2}} - P_2 Y_2^{\frac{1}{2}} \right\|$$

$$= \min_{P_1, P_2} \left(\operatorname{tr} \left(P_2 Y_2^{\frac{1}{2}} - P_1 Y_1^{\frac{1}{2}} \right)^T \left(P_2 Y_2^{\frac{1}{2}} - P_1 Y_1^{\frac{1}{2}} \right) \right)^{\frac{1}{2}}$$

$$= \min_{P_1, P_2} \left(\operatorname{tr}(Y_1) + \operatorname{tr}(Y_2) - 2\operatorname{tr} \left(P_2^T P_1 Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}} \right) \right)^{\frac{1}{2}}$$

$$= \left(\operatorname{tr}(Y_1) + \operatorname{tr}(Y_2) - 2 \max_{P_1, P_2} \operatorname{tr} \left(P_2^T P_1 Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}} \right) \right)^{\frac{1}{2}}.$$

By the results of matrix analysis [12], we see that for $P \in \mathcal{O}(n)$, $\operatorname{tr} \left(P_1 Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}} P_2 \right)$ achieves the maximum if and only if $P_1 Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}} P_2$ is orthogonal diagonalizable. In fact, P in Lemma 3.4 maximizes $\operatorname{tr} \left(P Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}} \right)$, where $P = Y_1^{-\frac{1}{2}} (Y_1 Y_2)^{\frac{1}{2}} Y_2^{-\frac{1}{2}}$ and $\max_P \left(\operatorname{tr} \left(P Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}} \right) \right) = \operatorname{tr}(Y_1 Y_2)^{\frac{1}{2}}$. Thus the minimal geodesic distance of $(SPD(n), g_W)$ exactly equals to the Wasserstein distance.

Then a basic geometric observation [7] claims that if a shortest geodesic can be extended, it must

be unique. Therefore, the uniqueness of the geodesic comes from the minimality proved above. □

Remark. The non-existence of the conjugate pair actually implies the non-existence of the cut locus in our cases.

Up to the present, we have proved both the **existence** and **uniqueness** of Wasserstein geodesic.

As well known that the behaviours of Jacobi fields are controlled by curvatures. Since Jacobi fields on $(SPD(n), g_W)$ is completely known, the Jacobi equation actually provides a method to calculate curvatures. Unfortunately, the computing is extremely complicated. We have to find another way for curvatures. One knows that the non-negative curvature always means geodesics convergent and brings conjugate pairs. But $(SPD(n), g_W)$ is wired since it admits non-negative curvatures but without any conjugate pair.

6 Curvature

Our ultimate aim is to understand the Riemannian curvature of $(SPD(n), g_W)$. We denote the Euclidean curvature on bundle (null entirely) as $\widetilde{R}$, and the Wasserstein (Riemannian) curvature on $(SPD(n), g_W)$ as R.

6.1 Riemannian Curvature Tensor

Theorem 6.1. *For any $Y \in SPD(n)$, and A, B are smooth vector fields on $SPD(n)$, the Wasserstein curvature tensor $R(A, B, A, B) := \langle R_{AB}A, B \rangle_Y$ at Y has an explicit expression*

$$R(A, B, A, B) = 3\mathrm{tr}\left(\Gamma_Y A Y \Gamma_Y \left(\Gamma_Y A \Gamma_Y B - \Gamma_Y B \Gamma_Y A\right) Y \Gamma_Y B\right). \tag{6.36}$$

Proof. (Theorem 6.1) We attempt to compare $R|_Y(A, B, A, B)$ with $\widetilde{R}|_X\left(\widetilde{A}, \widetilde{B}, \widetilde{A}, \widetilde{B}\right)$ as follows

$$
\begin{aligned}
\widetilde{R}\left(\widetilde{A}, \widetilde{B}, \widetilde{A}, \widetilde{B}\right) :&= \langle \widetilde{R}_{\widetilde{A}\widetilde{B}}\widetilde{A}, \widetilde{B}\rangle \\
&= D_{[\widetilde{A},\widetilde{B}]}\widetilde{A}, \widetilde{B}\rangle - \langle D_{\widetilde{A}}D_{\widetilde{B}}\widetilde{A}, \widetilde{B}\rangle + \langle D_{\widetilde{A}}D_{\widetilde{B}}\widetilde{A}, \widetilde{B}\rangle \\
&= \langle D_{\widetilde{[A,B]}}\widetilde{A}, \widetilde{B}\rangle + \langle D_{2\mathcal{T}(A,B)}\widetilde{A}, \widetilde{B}\rangle - \langle D_{\widetilde{A}}\widetilde{\nabla_B A}, \widetilde{B}\rangle \\
&\quad - \langle D_{\widetilde{A}}\mathcal{T}(B,A), \widetilde{B}\rangle + \langle D_{\widetilde{B}}\widetilde{\nabla_A A}, \widetilde{B}\rangle + \langle D_{\widetilde{B}}\mathcal{T}(A,A), \widetilde{B}\rangle \\
&= \langle \nabla_{[A,B]}A, B\rangle + 2\langle D_{\mathcal{T}(A,B)}\widetilde{A}, \widetilde{B}\rangle - \langle \nabla_A \nabla_B A, B\rangle \\
&\quad + \langle \mathcal{T}(B,A), \mathcal{T}(A,B)\rangle + \langle \nabla_B \nabla_A A, B\rangle - \langle \mathcal{T}(A,A), \mathcal{T}(B,B)\rangle.
\end{aligned}
$$

By the definition of the usual derivative, we have

$$
D_{\mathcal{T}(A,B)}\widetilde{A} = \lim_{t\to 0}\frac{1}{t}[(X + \mathcal{T}(A,B)\,t)\,\Gamma_Y A - X\Gamma_Y A] = \mathcal{T}(A,B)\,\Gamma_Y A. \tag{6.37}
$$

By Lemma 4.2 and (4.28), we have $\langle [\mathcal{T}(A,B), \widetilde{A}], \widetilde{B}\rangle = \langle \mathcal{T}(A,B), \widetilde{B}\rangle = 0$, and then we have

$$
\langle D_{2\mathcal{T}(A,B)}\widetilde{A}, \widetilde{B}\rangle = -2\langle \mathcal{T}(A,B), D_{\widetilde{A}}\widetilde{B}\rangle = -2\langle \mathcal{T}(A,B), \mathcal{T}(A,B)\rangle. \tag{6.38}
$$

Therefore, by putting (6.37) and (6.38) into (6.1) and the anti-symmetry of $\mathcal{T}(,)$, we get

$$
\begin{aligned}
\widetilde{R}\left(\widetilde{A}, \widetilde{B}, \widetilde{A}, \widetilde{B}\right) &= R(A,B,A,B) + 3\langle \mathcal{T}(A,B)\,\Gamma_Y A, X\Gamma_Y B\rangle \\
&= R(A,B,A,B) - 3\langle \mathcal{T}(A,B), \mathcal{T}(A,B)\rangle.
\end{aligned} \tag{6.39}
$$

Then, by the first equality from (6.39) and $\widetilde{R} \equiv 0$, we obtain the explicit expression for the Wasserstein curvature

$$
\begin{aligned}
R(A,B,A,B) &= -3\langle \mathcal{T}(A,B)\,\Gamma_Y A, X\Gamma_Y B\rangle \\
&= 3\mathrm{tr}\left(\Gamma_Y AY\Gamma_Y \left(\Gamma_Y A\Gamma_Y B - \Gamma_Y B\Gamma_Y A\right)Y\Gamma_Y B\right).
\end{aligned} \tag{6.40}
$$

$\square$

Noticing $\|\mathcal{T}(A,B)\|^2 \geq 0$, from above theorem we can obtain

Corollary 6.2. $(SPD(n), g_W)$ *has non-negative curvatures, namely*

$$R(A, B, A, B) \geq 0. \tag{6.41}$$

To solve the Sylvester equation by using Algorithm 1, we can simplify the expression. We give the sectional curvature K of the section span$\{A(Y), B(Y)\}$

$$
\begin{aligned}
K|_Y(A, B) &= \frac{R(A, B, A, B)}{\langle A, A \rangle \langle B, B \rangle - \langle A, B \rangle^2} \\
&= 12 \frac{\operatorname{tr}(E_A \Lambda \Gamma_\Lambda [E_A, E_B] \Lambda E_B)}{\operatorname{tr}(E_A C_A) \operatorname{tr}(E_B C_B) - \operatorname{tr}^2(E_A C_B)},
\end{aligned}
\tag{6.42}
$$

where we used the same donations as Algorithm 1. This formula will take some convenience for computing in computer programs.

We can also obverse that the sectional curvature conforms to the inverse ratio law

$$K|_{k\Lambda}(A, B) = \frac{1}{k} K|_\Lambda(A, B), \quad \forall k \in \mathbb{R}. \tag{6.43}$$

These results conform with our visualized views of $(SPD(n), g_W)$ presented in Figure 3-1 and Figure 3-2, where the manifold tends to be flat when k increases.

6.2 Estimate

Conventionally, we consider each fixed diagonal matrix Λ on $(SPD(n), g_W)$. Define $\{S^{p,q}\}$ as

$$S^{p,q} = [S_{ij}^{p,q}], \quad S_{ij}^{p,q} = \delta_i^p \delta_j^q + \delta_i^q \delta_j^p, \tag{6.44}$$

where the superscripts p, q mark the nonzero elements and δ is the Kronecker delta. In fact, $\{S^{p,q} | 1 \leq p \leq q \leq n\}$ is a basis naturally. For simplicity, we sometimes sign $S^{p,q}$, $S^{r,t}$ with S_1 and S_2. By this way, we can express the curvature under this basis.

Before that, we introduce a useful sign, which indicates how much symmetric pairs in the Cartesian product of two script-sets.

Definition 6.1. For any two finite script-sets $\eta_1 = \{\alpha_i | \alpha_i \in \mathbb{N}\}$, $\eta_2 = \{\beta_j | \beta_j \in \mathbb{N}\}$, the co-equal

number τ is defined by

$$\tau\left(\eta_1, \eta_2\right) = \sum_i \sum_j \delta_{\alpha_i}^{\beta_j}. \tag{6.45}$$

Theorem 6.3. *For any sorted diagonal matrix $\Lambda = \mathrm{diag}[\lambda_1, s, \lambda_n] \in SPD\,(n)$, Wasserstein sectional curvature satisfies*

$$K|_\Lambda\left(S_1, S_2\right) = \frac{3\left(1 + \delta^{rt}\right)\lambda_p \lambda_t}{\left(\lambda_p + \lambda_r\right)\left(\lambda_r + \lambda_t\right)\left(\lambda_p + \lambda_t\right)}. \tag{6.46}$$

Proof. Firstly, consider the matrix multiplication

$$\left(S_1 S_2\right)_{ij} = \sum_{k=1}^n S_{ik}^{p,q} S_{kj}^{r,t} = \delta_i^p \delta^{qt} \delta_j^r + \delta_i^p \delta^{qr} \delta_j^t + \delta_i^q \delta^{pt} \delta_j^r + \delta_i^q \delta^{pr} \delta_j^t.$$

Subsequently, combining it with Algorithm 1, we have

$$E_{S_1} E_{S_2} \neq 0 \iff S_1 S_2 \neq 0 \iff \tau\left(\{p, q\}, \{r, t\}\right) \geq 1,$$

where the $E_S = \left[\frac{S_{ij}}{\lambda_i + \lambda_j}\right] = \Gamma_\Lambda S$, since Λ is diagonal (recalling Algorithm 1 for details).

According to the anti-symmetry of curvature tensor, we get

$$\{p, q\} = \{r, t\} \implies R|_\Lambda\left(S_1, S_2, S_1, S_2\right) = 0.$$

Thus, without loss of generality, the non-vanishing curvature appears only in the particular case:

$$p \neq q = r, p \neq t.$$

In this case, we have

$$\left(S_1 S_2\right)_{ij} = \left(1 + \delta^{rt}\right)\delta_i^p \delta_j^t = \begin{cases} \delta_i^p \delta_j^t, & \text{if } r \neq t, \\ 2\delta_i^p \delta_j^t, & \text{if } r = t, \end{cases}$$

and

$$\left(E_{S_1} E_{S_2}\right)_{ij} = \frac{1 + \delta^{rt}}{\left(\lambda_p + \lambda_r\right)\left(\lambda_r + \lambda_t\right)}\delta_i^p \delta_j^t. \tag{6.47}$$

24

Then, we obtain

$$\left(\Gamma_\Lambda[E_{S_1}, E_{S_2}]\right)_{ij} = \frac{\left(1 + \delta^{rt}\right)\left(\lambda_p + \lambda_t\right)^{-1}}{\left(\lambda_p + \lambda_r\right)\left(\lambda_r + \lambda_t\right)}\left(\delta_i^p \delta_j^t - \delta_j^p \delta_i^t\right)$$

and

$$\sum_{k=1}^n \left(\Gamma_\Lambda[E_{S_1}, E_{S_2}]\right)_{ik}\left(\Lambda E_{S_2} E_{S_1}\Lambda\right)_{kj}$$

$$= \sum_{k=1}^n \frac{\left(1 + \delta^{rt}\right)\left(\lambda_p + \lambda_t\right)^{-1}}{\left(\lambda_p + \lambda_r\right)\left(\lambda_r + \lambda_t\right)}\left(\delta_i^p \delta_k^t - \delta_k^p \delta_i^t\right)\frac{\lambda_p \lambda_t \left(1 + \delta^{rt}\right)}{\left(\lambda_p + \lambda_r\right)\left(\lambda_r + \lambda_t\right)}\delta_k^t \delta_j^p$$

$$= \frac{\left(1 + \delta^{rt}\right)^2 \lambda_p \lambda_t}{\left(\lambda_p + \lambda_r\right)^2 \left(\lambda_r + \lambda_t\right)^2 \left(\lambda_p + \lambda_t\right)}\delta_i^p \delta_j^p.$$

Thus we gain the curvature

$$R|_\Lambda\left(S_1, S_2, S_1, S_2\right) = 3\mathrm{tr}\left(E_{S_1}\Lambda\Gamma_\Lambda[E_{S_1}, E_{S_2}]\Lambda E_{S_2}\right)$$

$$= \frac{3\left(1 + \delta^{rt}\right)^2 \lambda_p \lambda_t}{\left(\lambda_p + \lambda_r\right)^2 \left(\lambda_r + \lambda_t\right)^2 \left(\lambda_p + \lambda_t\right)}. \tag{6.48}$$

On the other hand, by (6.47) we have

$$S_1 \neq S_2 \implies \langle S_1, S_2\rangle = \mathrm{tr}\left(E_{S_1}\Lambda E_{S_2}\right) = 0,$$

and

$$\langle S_2, S_2\rangle = \frac{1}{2}\mathrm{tr}\left(E_{S_2} S_2\right) = \frac{1 + \delta^{rt}}{\left(\lambda_r + \lambda_t\right)}.$$

Then, we can get the area of sectional parallelogram as

$$\langle S_1, S_1\rangle\langle S_2, S_2\rangle - \langle S_1, S_2\rangle^2 = \frac{1 + \delta^{rt}}{\left(\lambda_p + \lambda_r\right)\left(\lambda_r + \lambda_t\right)}. \tag{6.49}$$

From (6.48) and (6.49), the proof is done consequently. $\qquad\square$

With the above expansion for sectional curvatures, we can easily gain the estimate, which means that any sectional curvature at a fixed point $Y \in SPD\left(n\right)$ can be controlled by the minimal eigenvalue λ_{min} of Y.

25

Theorem 6.4. *For any $Y \in SPD(n)$, there exists an orthonormal basis $\{e_k\}$ for $T_Y SPD(n)$ such that for any $e_{k_1}, e_{k_2} \in \{e_k\}$*

$$0 < K_Y(e_{k_1}, e_{k_2}) < \frac{3}{\lambda_{min}(Y)}. \tag{6.50}$$

Proof. As Y is diagonizable, there exists a $\Lambda \in SPD(n)$, and $A', B' \in T_\Lambda SPD(n)$ such that $K_Y(A, B) = K_\Lambda(A', B')$ from Lemma 2.5. Therefore, by Theorem 6.3, we have

$$K\left(S^{p,q}, S^{r,t}\right) = \begin{cases} \frac{3\lambda_p \lambda_t}{(\lambda_p + \lambda_r)(\lambda_r + \lambda_t)(\lambda_p + \lambda_t)}, & p \neq q = r \neq t, \ p \neq t, \\ \frac{3\lambda_p}{(\lambda_p + \lambda_t)^2}, & p \neq q = r = t, \\ 0, & else. \end{cases}$$

From the positive definiteness of Y, we see that

$$\lambda_i > 0 \Longrightarrow K\left(S^{p,q}, S^{r,t}\right) < \frac{3}{\lambda_p + \lambda_t},$$

which gives the desired estimate. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

It is remarkable that, sectional curvatures are controlled by the secondly minimal eigenvalue, which implies that the curvature will seldom explode even on a domain almost degenerated. Only when the matrices degenerate at over two dimensions will the curvatures be very large. This phoneme ensures the curvature information make sense in most applications. Some examples for this phoneme can be observed in the next subsection.

6.3   Scalar Curvature

In the last part of this section, we calculate the scaler curvature directly.

Theorem 6.5. *For any $Y \in SPD(n)$, the scalar curvature satisfies*

$$\rho(Y) = 3\text{tr}\left(2\Lambda U U^T + \Lambda U^T U + \Lambda U \left(U + U^T\right) \Lambda \left(U + U^T\right)\right), \tag{6.51}$$

where $\Lambda = \text{diag}[\lambda_1, s, \lambda_n]$ is a diagonalization of Y, and $U := [\frac{1}{\lambda_i + \lambda_j}]_{i<j}$ is a upper triangle matrix. The subscripts of eigenvalue λ must keep consistent when constructing Λ and U.

Proof. By summation, we get

$$\rho\left(\Lambda\right) = \sum_{p=1}^{n}\sum_{r=p}^{n}\sum_{t=1}^{n} K\left(S^{p,r}, S^{r,t}\right)$$

$$= \sum_{p=1}^{n}\left(\sum_{r=p+1}^{n}\sum_{t=1}^{n,t\neq p} K\left(S^{p,r}, S^{r,t}\right) + \sum_{t=1}^{n,t\neq p} K\left(S^{t,p}, S^{p,p}\right)\right)$$

$$= 3\sum_{p=1}^{n}\left(\sum_{r=p+1}^{n}\sum_{t=1}^{n,t\neq p} \frac{\left(1+\delta^{rt}\right)\lambda_p\lambda_t}{\left(\lambda_p+\lambda_r\right)\left(\lambda_r+\lambda_t\right)\left(\lambda_p+\lambda_t\right)} + \sum_{t=1}^{n,t\neq p} \frac{\lambda_t}{\left(\lambda_t+\lambda_p\right)^2}\right)$$

$$= 3\sum_{p=1}^{n}\left(\sum_{r=p}^{n}\sum_{t=1}^{n} \frac{\left(1-\delta^{pr}\right)\left(1+\delta^{rt}\right)\left(1-\delta^{pt}\right)\lambda_p\lambda_t}{\left(\lambda_p+\lambda_r\right)\left(\lambda_r+\lambda_t\right)\left(\lambda_p+\lambda_t\right)} + \sum_{t=1}^{n} \frac{\left(1-\delta^{pt}\right)\lambda_t}{\left(\lambda_t+\lambda_p\right)^2}\right)$$

$$= 3\sum_{p=1}^{n}\sum_{r=p}^{n} \frac{\lambda_p\left(1+\delta^{pr}\right)}{\lambda_p+\lambda_r} \sum_{t=1}^{n} \frac{1+\delta^{rt}}{\lambda_r+\lambda_t} \frac{\lambda_t\left(1-\delta^{tp}\right)}{\lambda_t+\lambda_p}.$$

Using the definitions before, we see that $\rho\left(\Lambda\right)$ can be rewritten as the desired trace. $\square$

Figure 6-4 presents some examples for scalar curvatures on $\left(SPD\left(2\right), g_W\right)$, which shows our argument in the last part of section 6.2.

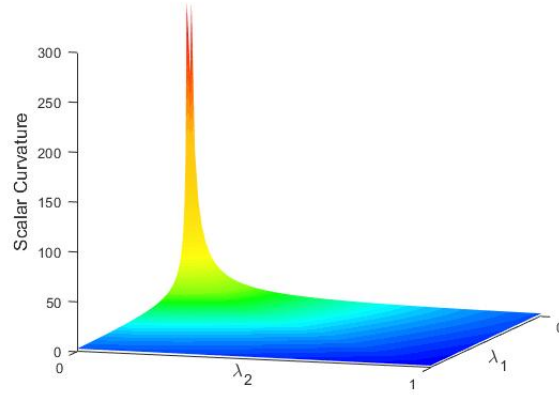
Figure 6-4 Scalar curvatures on SPD(2)

7   Conclusion and Further Work

In this paper, we study some geometric characteristics of $\left(SPD\left(n\right), g_W\right)$, including geodesics, the connection, Jacobi fields and curvatures. From this viewpoint, we have gained a basic compre-

hension for the Wasserstein geometry on $SPD\left(n\right)$, which shows that the Wasserstein distance has both deep application potentials and the mathematical elegance. While there still are some open problems left, such as the exact degree of symmetry and more precious bounds of curvatures, some ideas remain to be tested and generalized. Based on the results in the paper, we have designed a series of algorithms for data processing, and further relevant works are being carried out. The new results will be presented in our future works.

References

[1] S. Kobayashi. Transformation Groups in Differential Geometry. Classics in Mathematics, 1932.

[2] A. Martin, C. Soumith, B. Leon. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, 2017.

[3] T. Asuka. On Wasserstein Geometry of Gaussian Measures. Probabil Approach to Geometry. 2010, 1, 463-472.

[4] L. Malag, L. Montrucchio, G. Pistone. Wasserstein Riemannian geometry of Gaussian densities. Information Geometry. 2018, 1, 137-179.

[5] J. H. Manton. A Primer on Stochastic Differential Geometry for Signal Processing. IEEE Journal of Selected Topics in Signal Processing. 2013, 4(7), 681-699.

[6] M. Gelbrich. On a Formula for the L2 Wasserstein Metric between Measures on Euclidean and Hilbert Spaces. Mathematische Nachrichten. 2015, 1(147), 185-203.

[7] M. P. do Carmo. Riemannian Geometry. Boston, 1992.

[8] S. Zhang , Y. Cao, W. Li , et al. A New Riemannian Structure in SPD(n). IEEE International Conference on Signal, Information and Data Processing(ICSIDP), 2019.

[9] C. R. Givens, R. M. Shortt. A Class of Wasserstein Metrics for Probability Distributions. Michigan Mathematical Journal. 1984, 2(31), 231-240.

[10] J. M. Lee. Introduction to Smooth Manifolds. Springer Berlin, 2003.

[11] R. Hosseini, S. Sra. Matrix Manifold Optimization for Gaussian Mixtures. International Conference on Neural Information Processing Systems. MIT Press, 2015.

[12] R. Bhatia. Matrix Ananlysis. Graduate Texts in Mathematics, 1997.

[13] A. J. B. Ward. A General Analysis of Sylvester's Matrix Equation. International Journal of Mathematical Education in Science and Technology, 1991.

[14] E. Masry. Multivariate Probability Density Estimation for Associated Processes: Strong Consistency and Rates. Statistics & Probability Letters. 2002, 2(58), 205-219.

[15] V. Arsigny, P. Fillard, X. Pennec, N. Ayache. Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors. Magnetic Resonance in Medicine. 2010, 2(56), 411-421.

[16] Y. Rubner, C. Tomasi, L. J. Guibas. A Metric for Distributions with Applications to Image Databases. International Conference on Computer Vision, IEEE Computer Society. 1998.

[17] X. Pennec, P. Fillard, N. Ayache. A Riemannian Framework for Tensor Computing. International Journal of Computer Vision. 2006, 1(66), 41-66.

[18] B. C. Hall. Lie Groups, Lie Algebras, and Representations. Springer Berlin, 2015.

[19] J. M. Lee. Introduction to Smooth Manifolds. Springer Berlin, 2003.

[20] W. M. Boothby. An Introduction to Differentiable Manifolds and Riemannian Geometry (Second edition). Academic Press, 1986.

[21] R. Bhatia. Positive Definite Matrices. Princeton University Press, 2009.

[22] Y. Li, K. Wong. Riemannian Distances for Signal Classification by Power Spectral Density. IEEE Journal of Selected Topics in Signal Processing. 2013, 7, 655-669.