A New Riemannian Structure in SPD(n)

Shiqiang Zhang[1], Yueqi Cao[1], Wenyu Li[2], Fangjia Yan[1], Yihao Luo[1], Huafei Sun[1*],
[1]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China
[2]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[*]Email address: huafeisun@bit.edu.cn

Symmetric positive definite matrix has a widely application in many branches of science and engineering, a key issue is about how to find the 'nice' distance between two elements on $SPD(n)$ consisting of symmetric positive definite matrices. By using the fiber bundle and constructing a Riemannian submersion from the general linear group $GL(n)$ to $SPD(n)$, we develop the closed-form expressions of a new Riemannian metric and associated Riemannian distance in $SPD(n)$. This metric and distance can be used to obtain a more precise result compared with the widely used Euclidean distance, which is usually used in signal classification.

***Index Terms*—SPD(n), signal classification, Riemannian metric, fiber bundle, Riemannian submersion**

I. Introduction

Symmetric positive definite matrices [14] play an important role in many brunches of science and engineering, such as stability analysis of signal processing [6] [15], linear stationary systems, optimal control strategies, nuclear magnetic resonance imaging analysis, multivariate probability statistics [17] [3]. It is worthy mentioning that the covariant matrix of a Gaussian distribution is a symmetric positive definite matrix, and the power spectral density (PSD) matrix of a real signal is also a symmetric positive definite matrix, which is often used as a feature to facilitate classification [22]. Instead of considering a single matrix, investigating the geometric structure of the manifold $SPD(n)$ consists of all symmetric positive definite matrices is more natural and applicable.

A differentiable manifold is a set endowed a topology structure, which defines open sets and continuous maps [20]. A Riemannian manifold is a differentiable manifold with a Riemannian metric, which allows us to use the tools of calculus-differentiation, integration, etc. [4]. In application we focus on how to get the minimal distance between two symmetric positive definite matrices, where minimal distance depends on the structure of $SPD(n)$. A common idea is regarding $SPD(n)$ as a subset of $\mathbb{R}^{n^2}$ and using the Euclidean inner product. In this case, the minimal curve joints two points in $SPD(n)$ is straight line segment. However, this line segment does not totally fall to $SPD(n)$, which means Euclidean distance is not the 'minimal distance' we want. Since the incompleteness of Euclidean inner product, many researchers try to find a better geometric structure for $SPD(n)$. For example, Pennec et al. [21] defined the affine-invariant Riemannian metric, Arsigny et al. [18] constructed a Lie group structure which admits a bi-invariant metric-Log-Euclidean metric, and more recently, Wong et al. [22] applied a new distance measure to the classification of electroencephalogram (EEG) signals for the determination of sleep states and the results are highly encouraging. However, although this distance measure has given some highly encouraging results, few researchers have addressed that whether this distance measure is a metric or not. Moreover, geodesic distance [19] is more suitable with the metric defined in $SPD(n)$, and whether this distance measure is geodesic distance or not is important.

In this paper, we firstly clarify that this is a metric and develop closed-form expressions of this metric. After that we define a distance in $SPD(n)$ by using the length of piece-wise smooth curve, which is coincident to the 'distance measure' defined by Wong et al. and the Wasserstein distance of Gaussian densities with 0-mean [7]. Moreover, we calculate the natural gradient of $SPD(n)$ with this metric, which is useful in optimization on manifold. Compared with other metrics, this metric builds a bridge from a 'simple' space $GL(n)$ to $SPD(n)$ and allows us develop more properties of $SPD(n)$ by this approach. Furthermore, this idea may be extended to other cases, which provides a new way to study complicated manifolds.

II. A New Riemannian Structure on SPD(n)

A. Riemannian metric

In this part we are going to clarify that the "distance measure" by Wong et al. is a metric. Moreover, the closed-form of this metric can be obtained by fiber bundle and Riemannian submersion [12].

For any $A \in SPD(n)$, consider the orthogonal decomposition of A [11]:

$$A = Q\Lambda Q^T = Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q^T = (Q\Lambda^{\frac{1}{2}}Q^T)^2, \tag{1}$$

where $Q \in O(n)$ is an orthogonal matrix, and $\Lambda = diag(\lambda_1, ..., \lambda_n)$, where $\lambda_i > 0$ are the eigenvalues of A.

Denote that $A^{\frac{1}{2}} = Q\Lambda^{\frac{1}{2}}Q^T$, $U \in O(n)$and $\tilde{A} = A^{\frac{1}{2}}U$, then we have that

$$\tilde{A}\tilde{A}^T = A^{\frac{1}{2}}UU^TA^{\frac{1}{2}} = A. \tag{2}$$

For any $n \times n$ nonsingular matrix M, MM^T is positive definite. Thus we can consider the following projection from $GL(n)$ to $SPD(n)$:

$$\begin{aligned} \pi : GL(n) &\rightarrow SPD(n) \\ \tilde{A} &\mapsto \tilde{A}\tilde{A} := A. \end{aligned} \tag{3}$$

For any $A \in SPD(n)$, we have

$$\pi^{-1}(A) = \{\tilde{A}|\tilde{A} = A^{\frac{1}{2}}U, U \in O(n)\}. \tag{4}$$

978-1-7281-2345-5/19/$31.00 ©2019 IEEE

Authorized licensed use limited to: BEIJING INSTITUTE OF TECHNOLOGY. Downloaded on January 02,2021 at 07:38:50 UTC from IEEE Xplore. Restrictions apply.

In the following discussion, we always say that $\tilde{A}$ is the expression of $A \in SPD(n)$ in $GL(n)$, and $\pi^{-1}(A)$ is the fiber of A.

Proposition 1. [9] *$GL(n)$ is an open submanifold of $M(n)$ and $T_A GL(n) = M(n)$, for any $A \in SPD(n)$.*

Before constructing Riemannian submersion, we recall Euclidean inner product in $GL(n)$:

$$\langle X, Y\rangle_A = \mathrm{tr}(X^T Y), A \in GL(n), X, Y \in T_A GL(n),$$

and the Frobenius distance:

$$\begin{aligned} d_F(A,B) &= \left(\sum_{i,j} |a_{ij}-b_{ij}|^2\right)^{\frac{1}{2}} \\ &= \sqrt{\mathrm{tr}((A-B)(A-B)^T)}, A, B \in GL(n). \end{aligned} \tag{5}$$

Definition 1. *Given $\tilde{A} \in GL(n)$, define the direct sum decomposition of $T_{\tilde{A}} GL(n)$ as*

$$T_{\tilde{A}} GL(n) = H(\tilde{A}) \oplus V(\tilde{A}), \tag{6}$$

where

$$V(\tilde{A}) = \{\tilde{X} \in T_{\tilde{A}} \mid d\pi_{\tilde{A}}(X) = 0\} \tag{7}$$

is called the vertical space and

$$H(\tilde{A}) = \{\tilde{Y} \in T_{\tilde{A}} GL(n) \mid \langle \tilde{X}, \tilde{Y}\rangle_{\tilde{A}} = 0, \forall \tilde{X} \in V(\tilde{A})\} \tag{8}$$

is called the horizontal space, where $d\pi$ denotes the differential of π.

Proposition 2.

$$H(\tilde{A}) = \{K\tilde{A} \mid K^T = K\}, V(\tilde{A}) = \{\tilde{A}S \mid S^T = -S\}. \tag{9}$$

Proof. An element in $V(\tilde{A})$ corresponds a tangent vector of a curve in $\pi^{-1}(\pi(\tilde{A}))$ passing through $\tilde{A}$, where

$$\pi^{-1}(\pi(\tilde{A})) = \{\tilde{A}Q \mid Q \in O(n\}. \tag{10}$$

A curve in $\pi^{-1}(\pi(\tilde{A}))$ passing through $\tilde{A}$ also corresponds a curve in $O(n)$ passed through the unit element I, and the tangent vectors on I is the set of all skew-symmetric matrices.

Thus we obtain

$$V(\tilde{A}) = \{\tilde{A}S \mid S^T = -S\}. \tag{11}$$

Given any skew-symmetric matrix S, since

$$(\exp(tS))^T = \exp(tS^T) = \exp(-tS) = (\exp(tS))^{-1}, \tag{12}$$

$\exp(tS) \in O(n)$, and $\gamma(t) = \tilde{A}\exp(tS)$ is a curve in $\pi^{-1}(\pi(\tilde{A}))$ passing through $\tilde{A}$.

Choose any curve $c : (-\delta, \delta) \to GL(n)$ passed through $\tilde{A}$ in $GL(n)$, and $\tilde{X} = \dot{c}(0)$.

If $\tilde{X} \in H(\tilde{A})$, then for any skew-symmetric S we have

$$0 = \langle \tilde{X}, \tilde{A}S\rangle_{\tilde{A}} = \mathrm{tr}\left(\tilde{X}^T \tilde{A} S\right), \tag{13}$$

which means

$$\tilde{X}^T \tilde{A} = (\tilde{X}^T \tilde{A})^T = \tilde{A}^T \tilde{X}, \tag{14}$$

thus $\tilde{X} = (\tilde{A}^T)^{-1}\tilde{X}^T\tilde{A}$.

Denote $K = (\tilde{A}^T)^{-1}\tilde{X}^T$, then $\tilde{X} = K\tilde{A}$, and K is symmetric since

$$K^T = \tilde{X}(\tilde{A})^{-1} = K\tilde{A}(\tilde{A})^{-1} = K. \tag{15}$$

Conversely, choose any symmetric matrix K, and $\tilde{X} = K\tilde{A}$, then

$$(\tilde{X})^T\tilde{A} = (\tilde{A})^T K^T \tilde{A} = (\tilde{A})^T K \tilde{A} = (\tilde{A})^T \tilde{X}, \tag{16}$$

which means that $(\tilde{X})^T\tilde{A}$ is symmetric [13], and $\tilde{X} \in H(\tilde{A})$.

In a word, we obtain an one-to-one correspondence between the set of all symmetric matrices and $H(\tilde{A})$, thus we have

$$H(\tilde{A}) = \{K\tilde{A} \mid K^T = K\}. \tag{17}$$

□

Proposition 3. *For any $A \in SPD(n)$, $\tilde{A}$ is an expression of A in $GL(n)$, then for any $X \in T_A SPD(n)$, there exists $\tilde{X} \in H(\tilde{A})$ such that $X = \tilde{A}\tilde{X} + \tilde{X}\tilde{A}$.*

Proof. Since $d\pi_{\tilde{A}}$ is subjective, we only need to prove that the tangent vector $\tilde{X}$ such that $X = d\pi_{\tilde{A}}(\tilde{X})$ satisfies the given equation.

Choose a smooth curve $\tilde{c} : (-\delta, \delta) \to GL(n)$ such that $\tilde{c}(0) = \tilde{A}, \dot{\tilde{c}}(0) = \tilde{X}$, then we have

$$\begin{aligned} X &= d\pi_{\tilde{A}}(\tilde{X}) \\ &= \tfrac{d}{dt}|_{t=0}\pi(\tilde{c}(t)) \\ &= \tfrac{d}{dt}|_{t=0}\tilde{c}(t)\tilde{c}(t)^T \\ &= \tilde{c}(0)\dot{\tilde{c}}(0)^T + \dot{\tilde{c}}(0)\tilde{c}(0)^T \\ &= \tilde{A}\tilde{X}^T + \tilde{X}\tilde{A}^T. \end{aligned} \tag{18}$$

□

Now we are going to consider what properties this new metric should satisfy such that $T_A SPD(n)$ is isomorphic to $H(\tilde{A})$.

For $A \in SPD(n), \tilde{A} \in GL(n), X, Y \in T_A SPD(n)$, by the previous proposition, we know that there exists $\tilde{X}, \tilde{Y} \in H(\tilde{A})$ such that

$$X = \tilde{A}\tilde{X}^T + \tilde{X}\tilde{A}^T, Y = \tilde{A}\tilde{Y}^T + \tilde{Y}\tilde{A}^T. \tag{19}$$

Firstly, calculate $\langle \tilde{X}, \tilde{Y}\rangle_{\tilde{A}}$. It is obvious that there exists a symmetric matrix K such that $\tilde{Y} = K\tilde{A}$, by the Euclidean inner product of $GL(n)$ we have that

$$\begin{aligned} \langle \tilde{X}, \tilde{Y}\rangle_{\tilde{A}} &= \mathrm{tr}\left(\tilde{X}^T\tilde{Y}\right) \\ &= \mathrm{tr}\left(\tilde{X}^T K \tilde{A}\right) \\ &= \tfrac{1}{2}\left(\mathrm{tr}\left(\tilde{A}\tilde{X}^T K\right) + \mathrm{tr}\left(\tilde{X}\tilde{A}^T K\right)\right) \\ &= \tfrac{1}{2}\mathrm{tr}(AK), \end{aligned} \tag{20}$$

in which we need the following equation:

$$\mathrm{tr}\left(\tilde{X}^T K \tilde{A}\right) = \mathrm{tr}\left(\tilde{A}^T K^T \tilde{X}\right) = \mathrm{tr}\left(\tilde{X}\tilde{A}^T K\right). \tag{21}$$

Meanwhile, we have

$$Y = \tilde{A}\tilde{Y}^T + \tilde{Y}\tilde{A}^T = \tilde{A}\tilde{A}^T K + K\tilde{A}\tilde{A}^T = AK + KA. \tag{22}$$

Authorized licensed use limited to: BEIJING INSTITUTE OF TECHNOLOGY. Downloaded on January 02,2021 at 07:38:50 UTC from IEEE Xplore. Restrictions apply.

The following theorem shows that, if we define

$$g_A(X,Y) = \frac{1}{2}\mathrm{tr}(XK), \tag{23}$$

where $A \in SPD(n), X, Y \in T_A SPD(n)$, K is symmetric and $AK + KA = Y$, then $g_A(\cdot,\cdot)$ is an Riemannian metric. With this metric, $H(\tilde{A})$ is isomorphic to $T_A SPD(n)$ naturally, thus $\pi : GL(n) \to SPD(n)$ is an Riemannian submersion [9].

Theorem 1. *$g_A(\cdot,\cdot)$ is an Riemannian metric.*

Proof. Firstly, we prove that if $A \in SPD(n), B \in S(n)$, then the solution of this matrix equation $AK + KA = B$ exists, and the solution is unique and symmetric.

Since A and $-A$ does not share a common eigenvalue, which implies the uniqueness of solution. Thus we only need to find a symmetric solution.

Similarly, consider the orthogonal decomposition of A:

$$A = Q\Lambda Q^T, \tag{24}$$

where $Q \in O(n)$ is an orthogonal matrix, and $\Lambda = diag(\lambda_1, ..., \lambda_n)$, where $\lambda_i > 0$ are eigenvalues of A.

Denote $C = (c_{ij})$, with $C = Q^T BQ$, then define $D = (d_{ij})$ with

$$d_{ij} = \frac{1}{\lambda_i + \lambda_j} c_{ij}, \forall 1 \le i, j \le n. \tag{25}$$

We can clarify that $K = QDQ^T$ is the unique solution of $AK + KA = B$.

Precisely, we have that

$$\begin{aligned} AK + KA &= Q\Lambda Q^T QDQ^T + QDQ^T Q\Lambda Q \\ &= Q(\Lambda D + D\Lambda)Q^T \\ : &= QEQ^T, \end{aligned} \tag{26}$$

where $E = \Lambda D + D\Lambda$.

Denote $E = (e_{ij})$, since

$$e_{ij} = \lambda_i \cdot \frac{1}{\lambda_i + \lambda_j} c_{ij} + \frac{1}{\lambda_i + \lambda_j} c_{ij} \cdot \lambda_j = c_{ij}, \tag{27}$$

thus $E = C$, which means that

$$QEQ^T = QCQ^T = B. \tag{28}$$

From the definition of K, we can know that K depends on A smoothly. After clarifying symmetry, positive-definiteness and bi-linearity, we can say that $g_A(\cdot,\cdot)$ is indeed an Riemannian metric.

Symmetry:

$$\begin{aligned} g_A(X,Y) &= \tfrac{1}{2}\mathrm{tr}(XK) = \tfrac{1}{2}\mathrm{tr}(Q^T XQK), \\ g_A(Y,X) &= \tfrac{1}{2}\mathrm{tr}(YL) = \tfrac{1}{2}\mathrm{tr}(Q^T YQL), \end{aligned} \tag{29}$$

where $AK + KA = Y, AL + LA = X$.

Denote $Q^T XQ = (x_{ij}), Q^T YQ = (y_{ij})$, and $D = (d_{ij}), E = (e_{ij})$ such that

$$d_{ij} = \frac{1}{\lambda_i + \lambda_j} y_{ij}, e_{ij} = \frac{1}{\lambda_i + \lambda_j} x_{ij}, \tag{30}$$

then we have $K = QDQ^T, L = QEQ^T$.

Thus we obtain

$$\begin{aligned} g_A(X,Y) &= \tfrac{1}{2}\mathrm{tr}(Q^T XQQ^T DQ) \\ &= \tfrac{1}{2}\mathrm{tr}(Q^T XQD) \\ &= \tfrac{1}{2}\sum_{i,j}\tfrac{x_{ij}y_{ij}}{\lambda_i+\lambda_j} \\ &= g_A(Y,X). \end{aligned} \tag{31}$$

Positive-definiteness:

$$\begin{aligned} g_A(X,X) &= \tfrac{1}{2}\mathrm{tr}(XK) \\ &= \tfrac{1}{2}\mathrm{tr}((AK + KA)K) \\ &= \mathrm{tr}(AK^2) \\ &= \mathrm{tr}((A^{\frac{1}{2}}K)(A^{\frac{1}{2}}K)^T) \ge 0. \end{aligned} \tag{32}$$

Moreover,

$$g_A(X,X) = 0 \Longleftrightarrow A^{\frac{1}{2}}K = 0 \Longleftrightarrow K = 0 \Longleftrightarrow X = 0. \tag{33}$$

Bi-linearity: It is clearly clarified since tr and the solution of a linear equation are both linear. □

B. Riemannian distance

To get the distance between two matrices, we translate the minimal distance between two points in $SPD(n)$ to the minimal distance between two fibers by horizontal lifting. Then we can get the closed-form expression of the minimal distance.

Theorem 2. *For any $A, B \in SPD(n)$, define*

$$d(A,B) = \sqrt{\mathrm{tr}A + \mathrm{tr}B - 2\mathrm{tr}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}}, \tag{34}$$

then $d(\cdot,\cdot)$ is a distance function.

Proof. Given any $\tilde{A}, \tilde{B} \in GL(n)$, although the straight line segment joints $\tilde{A}$ and $\tilde{B}$ in $M(n)$ (the set of $n \times n$ real matrices) does not belong to $GL(n)$ totally, we can use piecewise smooth curve to approximate this segment. Since $GL(n)$ is dense in $M(n)$, the lower bound of length of the curve joints $\tilde{A}$ and $\tilde{B}$ in $GL(n)$ is the length of this line segment, which is

$$\sqrt{\mathrm{tr}\left(\left(\tilde{A} - \tilde{B}\right)^T\left(\tilde{A} - \tilde{B}\right)\right)}. \tag{35}$$

By horizontal lifting, we can define the distance of $A, B \in SPD(n)$ by the minimal distance of two points from $\pi^{-1}(A)$ and $\pi^{-1}(B)$ respectively, written by

$$d(A,B) = \min_{\tilde{A},\tilde{B}}\sqrt{\mathrm{tr}\left(\left(\tilde{A} - \tilde{B}\right)^T\left(\tilde{A} - \tilde{B}\right)\right)}. \tag{36}$$

Since $\tilde{A} = A^{\frac{1}{2}}Q_1, \tilde{B} = B^{\frac{1}{2}}Q_2$, where $Q_1, Q_2 \in O(n)$, thus we can rewrite $d^2(A,B)$ by

$$\begin{aligned} &\min_{Q_1,Q_2} \mathrm{tr}\left(\left(A^{\frac{1}{2}}Q_1 - B^{\frac{1}{2}}Q_2\right)^T\left(A^{\frac{1}{2}}Q_1 - B^{\frac{1}{2}}Q_2\right)\right) \\ =& \min_{Q_1,Q_2}\left(\mathrm{tr}(A) + \mathrm{tr}(B) - 2\mathrm{tr}\left(Q_2Q_1^TA^{\frac{1}{2}}B^{\frac{1}{2}}\right)\right) \\ =& \mathrm{tr}(A) + \mathrm{tr}(B) - 2\max_{Q_1,Q_2}\mathrm{tr}\left(Q_2Q_1^TA^{\frac{1}{2}}B^{\frac{1}{2}}\right). \end{aligned} \tag{37}$$

Authorized licensed use limited to: BEIJING INSTITUTE OF TECHNOLOGY. Downloaded on January 02,2021 at 07:38:50 UTC from IEEE Xplore. Restrictions apply.

By the knowledge of matrix analysis [13], we obtain

$$\max_{Q_1,Q_2} \operatorname{tr}\left(Q_2 Q_1^T A^{\frac{1}{2}} B^{\frac{1}{2}}\right) \geq \operatorname{tr}\left(A^{\frac{1}{2}} B A^{\frac{1}{2}}\right)^{\frac{1}{2}} := \operatorname{tr}(AB)^{\frac{1}{2}}, \tag{38}$$

and the equality holds if and only if $Q_1^T A^{\frac{1}{2}} B^{\frac{1}{2}} Q_2$ is diagonal.

Since $A^{-\frac{1}{2}} ABA^{\frac{1}{2}} = A^{\frac{1}{2}} BA^{\frac{1}{2}}$, $A^{\frac{1}{2}} BA^{\frac{1}{2}}$ is similar to AB, thus

$$(AB)^{\frac{1}{2}} = (A^{\frac{1}{2}} BA^{\frac{1}{2}})^{\frac{1}{2}} \tag{39}$$

is well-define.

Finally we have

$$d(A,B) = \sqrt{\operatorname{tr}(A) + \operatorname{tr}(B) - 2\operatorname{tr}(AB)^{\frac{1}{2}}}. \tag{40}$$

By definition it is obvious that $d(\cdot,\cdot)$ is a distance, which means that it satisfies symmetry, positive-definiteness and the triangle inequality. □

In fact, Wong et al. used the similar calculation to get distance function, here we introduce piece-wise smooth curve to endow geometric interpretation of this distance and simplify the proof of the distance axioms, which avoid cumbersome and complicated verification. Moreover, it is amazing that this distance is coincident to the Wasserstein distance of 0-mean Gaussian distribution, where the Wasserstein distance between two Gaussian distribution $X \sim N(x|\mu_1, \Sigma_1), Y \sim N(y|\mu_2, \Sigma_2)$ is defined by [7] [8]

$$\begin{aligned} &W(X,Y) \\ = &\sqrt{\|\mu_1 - \mu_2\|^2 + \operatorname{tr}(\Sigma_1) + \operatorname{tr}(\Sigma_2) - 2\operatorname{tr}\left(\Sigma_1^{\frac{1}{2}} \Sigma_2 \Sigma_1^{\frac{1}{2}}\right)^{\frac{1}{2}}}. \end{aligned} \tag{41}$$

C. Natural Gradient

As an extension of the general Euclidean gradient to manifold, natural gradient plays an important role in the application of manifolds [10] [16]. In this part we are going to calculate the closed-form of natural gradient of $SPD(n)$ with this metric.

Definition 2. *M is a Riemannian manifold, $\langle \cdot,\cdot \rangle_p$ is the Riemannian metric at $p \in M$. Given a smooth map $f : M \to \mathbb{R}$, for any $p \in M$, and a curve γ such that $p = \gamma(0), X = \dot{\gamma}(0)$, the natural gradient of f at p is a tangent vector denoted by $\nabla f(p)$ such that*

$$\langle \nabla f, X \rangle_p = \left.\frac{d}{dt}\right|_{t=0} f(\gamma(t)). \tag{42}$$

Theorem 3. *Given a smooth map $f : SPD(n) \to \mathbb{R}$, the natural gradient of f at $A \in SPD(n)$ is*

$$\nabla f(A) = 2(A\frac{\partial f(A)}{\partial A} + \frac{\partial f(A)}{\partial A}A). \tag{43}$$

Proof. For any $A \in SPD(n), X \in T_A SPD(n)$, we have

$$\begin{aligned} & g_A\left(2(A\tfrac{\partial f(A)}{\partial A} + \tfrac{\partial f(A)}{\partial A}A), X\right) \\ = & \operatorname{tr}\left(A\tfrac{\partial f(A)}{\partial A}K + \tfrac{\partial f(A)}{\partial A}AK\right) \\ = & \operatorname{tr}\left(\tfrac{\partial f(A)}{\partial A}(AK + KA)\right) \\ = & \operatorname{tr}\left(\tfrac{\partial f(A)}{\partial A}X\right) = \tfrac{d}{dt}\big|_{t=0} f(\gamma(t)), \end{aligned} \tag{44}$$

where γ is a curve satisfies $\gamma(0) = A, \dot{\gamma}(0) = X$.

Thus we get ∇f as claimed. □

D. Computation Speed

We have discussed the geometric interpretation of this distance in theory. However, from the form of this distance function, it is difficult to see any advantages in practice application. In this part, we will give experience result to show the computation advantage of this distance, compared with the other classical Riemannian distances in $SPD(n)$.

Before experience, we give the closed-form of two classical Riemannian distances.

For any $A \in SPD(n), X, Y \in T_A SPD(n)$, the affine-invariant metric on $SPD(n)$ defined by

$$g_A(X,Y) = \operatorname{tr}(A^{-1}XA^{-1}Y), \tag{45}$$

the geodesic distance between $A, B \in SPD(n)$ is

$$d(A,B) = \sqrt{\operatorname{tr}\left(\log^2\left(A^{-\frac{1}{2}} BA^{-\frac{1}{2}}\right)\right)}. \tag{46}$$

For any $A \in SPD(n), X, Y \in T_A SPD(n)$, the Log-Euclidean metric on $SPD(n)$ defined by

$$g_A(X,Y) = g_I(d\log_A(X), d\log_A(Y)), \tag{47}$$

the geodesic distance between $A, B \in SPD(n)$ is

$$d(A,B) = \|\log(A) - \log(B)\|. \tag{48}$$

From the closed-form of these Riemannian distances, we can compare the time complexity of calculating the Riemannian distances between two matrices from $SPD(n)$. The affine-invariant metric involves in two time-cost steps, including getting $A^{-\frac{1}{2}}$ and computing the eigenvalues of $A^{-\frac{1}{2}} BA^{-\frac{1}{2}}$. Compared with the affine-invariant metric, the major time spent in the Log-Euclidean metric is computing the matrix logarithm $\log(\cdot)$, while in the new metric is computing the eigenvalue of AB. Here we use the conclusion that AB is similar to $A^{\frac{1}{2}} BA^{\frac{1}{2}}$, then we only need to calculate the eigenvalues of AB.

In experience, we denote the affine-invariant distance by 'RD1', the Log-Euclidean distance by 'RD2', the new distance by 'RD3'. By choosing different dimension n, we plot the time costs with respect to different dimension of these three distance functions.

The experience result shows that the new distance function and the Log-Euclidean distance has a similar and high computation speed, which is useful and efficient in practice. Compared with that, the high computation cost limits the application of affine-invariant metric.

III. Discussion

Actually, $SPD(n)$ is a useful object of signal process. In signal view, the PSD matrix of a real signal translates a real signal to a point in $SPD(n)$, which can be used to do signal classification. In a statistical view, the covariant matrix of a real signal, where we can consider the sample obeys a

Authorized licensed use limited to: BEIJING INSTITUTE OF TECHNOLOGY. Downloaded on January 02,2021 at 07:38:50 UTC from IEEE Xplore. Restrictions apply.

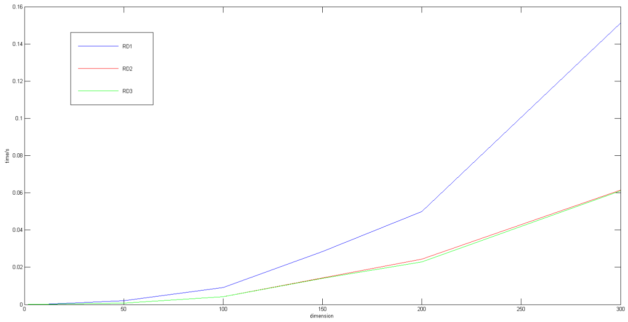

Fig. 1. Plot of the cost time for different dimension.

Gaussian distribution, also gives a connection from signal to $SPD(n)$. To some extent, the practical application demand promotes the development of different structures of $SPD(n)$, and the rich structures of $SPD(n)$ have a widely use in application conversely.

Wong et al. constructed a connection between $SPD(n)$ and $GL(n)$ and then got a Riemannian distance by horizontal lift. Applications of this new distance have got some highly encouraging results. However, their studies have either been short of the closed-form expression of this new metric or have lacked a proper explanation of this distance. In this study we endow a real geometry structure on $SPD(n)$ by using Riemannian submersion, in this case $SPD(n)$ is a trivial fiber bundle. Furthermore, we confirmed the existence and uniqueness of this Riemannian metric after translating its limitation into a form of Sylvester matrix equation and constructed the closed-formed expression of this metric. Introduction of piecewise curves simplifies the calculation of distance function without extra complicated verification. Moreover, the computation on $SPD(n)$ with this metric is fast and efficient.

In theory, this is the first study to our knowledge to apply Riemannian submersion and horizontal lift to define a manifold structure. Our results provide theoretical basis to this new structure. Such studies have shown a new and effective way to figure out complicated or unknown manifolds. In application, this structure admits a higher computation speed on $SPD(n)$ with accuracy since it possesses geometric interpretation and simple expression.

References

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

[2] C. R. Givens, R. M. Shortt. A class of Wasserstein metrics for probability distributions. Michigan Mathematical Journal, 31(1984), 231-240.

[3] H. Sun, L. Peng, Z. Zhang. Information Geometry and Its Applications. Advances in Mathematics (China), 40(2011), 257-269.

[4] J. M. Lee. Introduction to smooth manifolds. Springer Berlin, 2003.

[5] J. Stillwell. Naive Lie Theory. Springer Berlin, 2008.

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

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

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

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

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

[11] R. A. Horn, C. R. Johnson. Matrix Analysis. Cambridge University Press, 1985.

[12] R. Abraham, J. E. Marsden, T. Ratiu. Manifolds, Tensor Analysis, and Applications. Addison-Wesley, 1983.

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

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

[15] R. Kulikowsi. Signal Theory. IRE Transactions on Circuit Theory, 4(1958), 340-340.

[16] S. Amari. Natural Gradient Works Efficiently in Learning. Neural computation. MIT Press, 1998.

[17] S. Amari, H. Nagaoka. Methods of Information Geometry. Methods of Information Geometry. American Mathematical Society, 2007.

[18] V. Arsigny, P. Fillard, X. Pennec, N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine, 56(2006), 411-421.

[19] W. Klingenberg. Riemannian Geometry. The American Mathematical Monthly, 92(1985), 1345-1501.

[20] W. M. Boothby. An introduction to differentiable manifolds and Riemannian geometry(Second edition). Academic Press, 1986.

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

[22] Y. Li and K. M. Wong. Riemannian distances for signal classification by power spectral density. Journal of Selected Topics in Signal Processing, 7(2013), 655-669.

Authorized licensed use limited to: BEIJING INSTITUTE OF TECHNOLOGY. Downloaded on January 02,2021 at 07:38:50 UTC from IEEE Xplore. Restrictions apply.