# A Gyrovector Space Approach for Symmetric Positive Semi-definite Matrix Learning

Xuan Son Nguyen[0000−0002−2776−2254]

ETIS, UMR 8051, CY Cergy Paris Université, ENSEA, CNRS, Cergy, France
xuan-son.nguyen@ensea.fr

This supplementary material provides the proofs for the Theorems and Lemmas in our paper entitled "A Gyrovector Space Approach for Symmetric Positive Semi-definite Matrix Learning". Please see the paper for references.

The following notations will be used in our proofs. Denote by $\mathrm{Sym}_n$ the space of $n \times n$ symmetric matrices, $\mathrm{L}_n^+$ the space of $n \times n$ lower triangular matrices with positive diagonal entries, $\mathrm{L}_n$ the space of $n \times n$ lower triangular matrices, $\mathrm{Gr}_{n,p}$ the $p$-dimensional subspaces of $\mathbb{R}^n$.

## 1 Riemannian Geometry of SPD Manifolds

The space of SPD matrices is part of the vector space of square matrices. However, as mentioned in [2,36], employing the Euclidean metric for computations in this space can be problematic from both practical and theoretical points of view, i.e., the boundary problem or the tensor swelling effect. To address these issues, many Riemannian metrics on SPD manifolds have been proposed [2,24,36]. Below we briefly review the Affine-invariant, Log-Euclidean, and Log-Cholesky metrics studied in our work.

### 1.1 Affine-Invariant Metric

Based on the general principle of designing Riemannian metrics [57], Pennec et al. [36] proposed the Affine-invariant metric that is invariant under the action of affine transformations of the underlying space, i.e.,

$$< \mathbf{A}_1 | \mathbf{A}_2 >_{\mathbf{P}} = < \mathbf{Q} \star \mathbf{A}_1 | \mathbf{Q} \star \mathbf{A}_2 >_{\mathbf{Q} \star \mathbf{P}}$$

where $\mathbf{P} \in \mathrm{Sym}_n^+$, $\mathbf{A}_1, \mathbf{A}_2 \in T_{\mathbf{P}} \mathrm{Sym}_n^+$, $\mathbf{Q} \star \mathbf{P} = \mathbf{Q}\mathbf{P}\mathbf{Q}^T$ is the action of the linear group on $\mathrm{Sym}_n^+$, and $\mathbf{Q} \star \mathbf{A}_1 = \mathbf{Q}\mathbf{A}_1\mathbf{Q}^T$ is the action of the linear group on $\mathrm{Sym}_n$[1]. The dot product at the identity is defined as $< \mathbf{A}_1 | \mathbf{A}_2 >= \mathrm{Trace}(\mathbf{A}_1\mathbf{A}_2) + \beta \, \mathrm{Trace}(\mathbf{A}_1) \, \mathrm{Trace}(\mathbf{A}_2)$ with $\beta > -\frac{1}{n}$.

The Riemannian exponential map at a point can be obtained [36] as

$$\mathrm{Exp}_{\mathbf{P}}(\mathbf{A}) = \mathbf{P}^{\frac{1}{2}} \exp\Big(\mathbf{P}^{-\frac{1}{2}} \mathbf{A} \mathbf{P}^{-\frac{1}{2}}\Big) \mathbf{P}^{\frac{1}{2}}, \tag{14}$$

[1] Indeed, the action of the linear group on $\mathrm{Sym}_n^+$ is naturally extended to tangent vectors [36].

where $\mathbf{P} \in \mathrm{Sym}_n^+$, $\mathbf{A} \in T_{\mathbf{P}} \mathrm{Sym}_n^+$. By inverting the Riemannian exponential map, one obtains the Riemannian logarithmic map

$$\mathrm{Log}_{\mathbf{P}}(\mathbf{Q}) = \mathbf{P}^{\frac{1}{2}} \log\Big(\mathbf{P}^{-\frac{1}{2}}\mathbf{Q}\mathbf{P}^{-\frac{1}{2}}\Big)\mathbf{P}^{\frac{1}{2}}, \tag{15}$$

where $\mathbf{P}, \mathbf{Q} \in \mathrm{Sym}_n^+$. The parallel transport of a tangent vector $\mathbf{A} \in T_{\mathbf{P}} \mathrm{Sym}_n^+$ from $\mathbf{P}$ to $\mathbf{Q}$ along geodesics joining $\mathbf{P}$ and $\mathbf{Q}$ is given [60] by

$$\mathcal{T}_{\mathbf{P}\to\mathbf{Q}}(\mathbf{A}) = (\mathbf{Q}\mathbf{P}^{-1})^{\frac{1}{2}}\mathbf{A}\big((\mathbf{Q}\mathbf{P}^{-1})^{\frac{1}{2}}\big)^T. \tag{16}$$

### 1.2 Log-Euclidean Metric

Arsigny et al. [2] shown that the space of SPD matrices can be given a commutative Lie group structure by endowing it with the Log-Euclidean metric described as

$$< \mathbf{A}_1 | \mathbf{A}_2 >_{\mathbf{P}} = < D_{\mathbf{P}} \log(\mathbf{A}_1) | D_{\mathbf{P}} \log(\mathbf{A}_2) >_{\mathbf{I}},$$

where $\mathbf{P} \in \mathrm{Sym}_n^+$, $\mathbf{A}_1, \mathbf{A}_2 \in T_{\mathbf{P}} \mathrm{Sym}_n^+$, $D_{\mathbf{P}} \log(\mathbf{A}_1)$ and $D_{\mathbf{P}} \log(\mathbf{A}_2)$ are respectively the differentials of the matrix logarithm at $\mathbf{P}$ along tangent vectors $\mathbf{A}_1$ and $\mathbf{A}_2$, and $< .|. >_{\mathbf{I}}$ is any metric at the tangent space at $\mathbf{I}$.

One can derive [2] the Riemannian exponential and logarithmic maps at any point as

$$\mathrm{Exp}_{\mathbf{P}}(\mathbf{A}) = \exp(\log(\mathbf{P}) + D_{\mathbf{P}} \log(\mathbf{A})), \tag{17}$$

$$\mathrm{Log}_{\mathbf{P}}(\mathbf{Q}) = D_{\log(\mathbf{P})} \exp(\log(\mathbf{Q}) - \log(\mathbf{P})), \tag{18}$$

where $\mathbf{P}, \mathbf{Q} \in \mathrm{Sym}_n^+$, $\mathbf{A} \in T_{\mathbf{P}} \mathrm{Sym}_n^+$.

While the Log-Euclidean metric does not yield full affine-invariance, it shares very similar properties with the Affine-invariant metric. It allows to turn Riemannian computations into Euclidean computations in the logarithmic domain. This enables direct generalizations of traditional machine learning algorithms to the SPD manifold setting [2].

### 1.3 Log-Cholesky Metric

The Log-Cholesky metric has recently been proposed by Lin [24]. Under this framework, the space of SPD matrices can also be given a Lie group structure by endowing it with the following metric:

$$< \mathbf{A}_1 | \mathbf{A}_2 >_{\mathbf{P}} = g_{\mathbf{L}}\Big(\mathbf{L}(\mathbf{L}^{-1}\mathbf{A}_1\mathbf{L}^{-T})_{\frac{1}{2}}, \mathbf{L}(\mathbf{L}^{-1}\mathbf{A}_2\mathbf{L}^{-T})_{\frac{1}{2}}\Big),$$

where $\mathbf{P} \in \mathrm{Sym}_n^+$, $\mathbf{A}_1, \mathbf{A}_2 \in T_{\mathbf{P}} \mathrm{Sym}_n^+$, $\mathbf{L} = \mathscr{L}(\mathbf{P})$, $\mathbf{L}^{-T} = (\mathbf{L}^{-1})^T$, $\mathbf{S}_{\frac{1}{2}} = \lfloor \mathbf{S} \rfloor + \mathbb{D}(\mathbf{S})/2$ for any square matrix $\mathbf{S}$, and $g_{\mathbf{L}}(.,.)$ is defined as

$$g_{\mathbf{L}}(\mathbf{X}, \mathbf{Y}) = < \lfloor \mathbf{X} \rfloor, \lfloor \mathbf{Y} \rfloor >_F + < \mathbb{D}(\mathbf{L})^{-1}\mathbb{D}(\mathbf{X}), \mathbb{D}(\mathbf{L})^{-1}\mathbb{D}(\mathbf{Y}) >_F,$$

where $< .,. >_F$ denotes the Frobenius inner product.

The Riemannian exponential map is given by

$$\mathrm{Exp}_{\mathbf{P}}(\mathbf{W}) = \widetilde{\mathrm{Exp}}_{\mathscr{L}(\mathbf{P})}(D_{\mathbf{P}}\mathscr{L}(\mathbf{W}))\big(\widetilde{\mathrm{Exp}}_{\mathscr{L}(\mathbf{P})}(D_{\mathbf{P}}\mathscr{L}(\mathbf{W}))\big)^T, \tag{19}$$

where $D_{\mathbf{P}}\mathscr{L}(\mathbf{W})$ is the differential of $\mathscr{L}$ at $\mathbf{P}$ along direction $\mathbf{W}$, i.e.,

$$D_{\mathbf{P}}\mathscr{L}(\mathbf{W}) = \mathscr{L}(\mathbf{P})\big(\mathscr{L}(\mathbf{P})^{-1}\mathbf{W}\mathscr{L}(\mathbf{P})^{-T}\big)_{\frac{1}{2}}, \tag{20}$$

and $\widetilde{\mathrm{Exp}}(.)$ is defined as

$$\widetilde{\mathrm{Exp}}_{\mathbf{L}}(\mathbf{X}) = \lfloor \mathbf{L} \rfloor + \lfloor \mathbf{X} \rfloor + \mathbb{D}(\mathbf{L})\exp(\mathbb{D}(\mathbf{X})\mathbb{D}(\mathbf{L})^{-1}). \tag{21}$$

The Riemannian logarithmic map is given by

$$\mathrm{Log}_{\mathbf{P}}(\mathbf{Q}) = (D_{\mathscr{L}(\mathbf{P})}\mathscr{S})\big(\widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P})}\mathscr{L}(\mathbf{Q})\big), \tag{22}$$

where $\mathscr{S}(\mathbf{L}) = \mathbf{L}\mathbf{L}^T$, $D_{\mathbf{L}}\mathscr{S}(\mathbf{X})$ is the differential of $\mathscr{S}$ at $\mathbf{L}$ along direction $\mathbf{X}$, i.e.,

$$D_{\mathbf{L}}\mathscr{S}(\mathbf{X}) = \mathbf{L}\mathbf{X}^T + \mathbf{X}\mathbf{L}^T, \tag{23}$$

and $\widetilde{\mathrm{Log}}(.)$ is defined as

$$\widetilde{\mathrm{Log}}_{\mathbf{L}}(\mathbf{K}) = \lfloor \mathbf{K} \rfloor - \lfloor \mathbf{L} \rfloor + \mathbb{D}(\mathbf{L})\log(\mathbb{D}(\mathbf{L})^{-1}\mathbb{D}(\mathbf{K})). \tag{24}$$

The parallel transport of a tangent vector $\mathbf{A} \in T_{\mathbf{P}}\,\mathrm{Sym}_n^+$ from $\mathbf{P} \in \mathrm{Sym}_n^+$ to $\mathbf{Q} \in \mathrm{Sym}_n^+$ along geodesics joining $\mathbf{P}$ and $\mathbf{Q}$ is given by

$$\mathcal{T}_{\mathbf{P}\to\mathbf{Q}}(\mathbf{A}) = \mathbf{K}\big(\lfloor \mathbf{X} \rfloor + \mathbb{D}(\mathbf{K})\mathbb{D}(\mathbf{L})^{-1}\mathbb{D}(\mathbf{X})\big)^T + \big(\lfloor \mathbf{X} \rfloor + \mathbb{D}(\mathbf{K})\mathbb{D}(\mathbf{L}^{-1})\mathbb{D}(\mathbf{X})\big)\mathbf{K}^T, \tag{25}$$

where $\mathbf{L} = \mathscr{L}(\mathbf{P})$, $\mathbf{K} = \mathscr{L}(\mathbf{Q})$, and $\mathbf{X} = \mathbf{L}\big(\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-T}\big)_{\frac{1}{2}}$.

The Log-Cholesky metric enjoys a nice property of the Log-Euclidean metric that the Fréchet mean admits a closed-form expression. Furthermore, the computation of the Log-Cholesky parallel transport is much faster than that of the Affine-invariant and Log-Euclidean parallel transports. These properties make the Log-Cholesky metric a good choice for high-dimensional problems [24].

## 2 Proof of Lemma 1

*Proof.* Using Eqs. (14), (15), and (16) leads to the conclusion of the Lemma.

## 3 Proof of Lemma 2

*Proof.* Using Eqs. (14) and (15), it is straightforward to see that

$$t \otimes_{ai} \mathbf{P} = \mathrm{Exp}^{ai}_{\mathbf{I}_n}(t\,\mathrm{Log}^{ai}_{\mathbf{I}_n}(\mathbf{P})) = \exp(\log(\mathbf{P}))^t = \mathbf{P}^t.$$

# 4 Proof of Theorem 1

*Proof.* First, note that the binary operation $\oplus_{ai}$ verifies the Left Cancellation Law [44,45,46], i.e.,

$$\ominus_{ai}\mathbf{P} \oplus_{ai} (\mathbf{P} \oplus_{ai} \mathbf{Q}) = \mathbf{Q},$$

for any $\mathbf{P}, \mathbf{Q} \in \mathrm{Sym}_n^+$.

The gyroautomorphism can be determined from the binary operation as in [44,45,46]. By axiom (G3) and the Left Cancellation Law,

$$\mathrm{gyr}_{ai}[\mathbf{P}, \mathbf{Q}]\mathbf{R} = \big(\ominus_{ai} (\mathbf{P} \oplus_{ai} \mathbf{Q})\big) \oplus_{ai} \big(\mathbf{P} \oplus_{ai} (\mathbf{Q} \oplus_{ai} \mathbf{R})\big). \tag{26}$$

Using the expression of the binary operation $\oplus_{ai}$ given in Lemma 1, we deduce that

$$\mathrm{gyr}_{ai}[\mathbf{P}, \mathbf{Q}]\mathbf{R} = (\mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}})^{-\frac{1}{2}}\mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}\mathbf{R}\mathbf{Q}^{\frac{1}{2}}\mathbf{P}^{\frac{1}{2}}(\mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}})^{-\frac{1}{2}}.$$

Let $F_{ai}(\mathbf{P}, \mathbf{Q}) = (\mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}})^{-\frac{1}{2}}\mathbf{P}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}$. Then

$$F_{ai}(\mathbf{P}, \mathbf{Q})\mathbf{Q}^{\frac{1}{2}}\mathbf{P}^{\frac{1}{2}}(\mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}})^{-\frac{1}{2}} = (\mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}})^{-\frac{1}{2}}\mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}}(\mathbf{P}^{\frac{1}{2}}\mathbf{Q}\mathbf{P}^{\frac{1}{2}})^{-\frac{1}{2}} = \mathbf{I}_n.$$

Therefore

$$\mathrm{gyr}_{ai}[\mathbf{P}, \mathbf{Q}]\mathbf{R} = F_{ai}(\mathbf{P}, \mathbf{Q})\mathbf{R}(F_{ai}(\mathbf{P}, \mathbf{Q}))^{-1}.$$

It is then easy to verify axioms G1,G2,G4,V1,V2,V3,V4,V5 for Affine-invariant gyrovector spaces.

# 5 Proof of Lemma 3

*Proof.* Let $L$ be the left translation defined as

$$L_{\mathbf{P}}(\mathbf{Q}) = \exp(\log(\mathbf{P}) + \log(\mathbf{Q})).$$

Since the Log-Euclidean metric is a bi-invariant metric, the Levi-Civita connection coincides with the Cartan connection and the parallel transport of a tangent vector $\mathbf{V} \in T_{\mathbf{P}} \mathrm{Sym}_n^+$ is induced by the left translation [52,56,58], i.e.,

$$\mathcal{T}^{le}_{\mathbf{P}\to\mathbf{Q}}(\mathbf{V}) = D_{\mathbf{P}} L_{\mathbf{Q}\mathbf{P}^{-1}}(\mathbf{V}).$$

Thus, when $\mathbf{Q} = \mathbf{I}_n$,

$$\mathcal{T}^{le}_{\mathbf{P}\to\mathbf{I}_n}(\mathbf{V}) = D_{\mathbf{P}} L_{\mathbf{P}^{-1}}(\mathbf{V}). \tag{27}$$

Note that

$$(\log \circ L_{\mathbf{P}^{-1}})(\mathbf{R}) = \log(\mathbf{P}^{-1}) + \log(\mathbf{R})$$

Hence

$$D_{\exp(\log(\mathbf{P}^{-1})+\log(\mathbf{R}))} \log \circ D_{\mathbf{R}} L_{\mathbf{P}^{-1}} = D_{\mathbf{R}} \log.$$

When $\mathbf{R} = \mathbf{P}$, we get

$$D_{\mathbf{P}} L_{\mathbf{P}^{-1}} = D_{\mathbf{P}} \log . \tag{28}$$

Combining Eqs. (27) and (28) leads to

$$\mathcal{T}^{le}_{\mathbf{P} \to \mathbf{I}_n}(\mathbf{V}) = D_{\mathbf{P}} \log(\mathbf{V}). \tag{29}$$

Let $\mathcal{T}^{le}_{\mathbf{I}_n \to \mathbf{P}}(\mathrm{Log}^{le}_{\mathbf{I}_n}(\mathbf{Q})) = \mathbf{V}$. Then $\mathcal{T}^{le}_{\mathbf{P} \to \mathbf{I}_n}(\mathbf{V}) = \mathrm{Log}^{le}_{\mathbf{I}_n}(\mathbf{Q}) = \log(\mathbf{Q})$. From Eq. (29), we get

$$D_{\mathbf{P}} \log(\mathbf{V}) = \log(\mathbf{Q}).$$

Then, from Eq. (17),

$$\mathrm{Exp}^{le}_{\mathbf{P}}(\mathbf{V}) = \exp(\log(\mathbf{P}) + D_{\mathbf{P}} \log(\mathbf{V})),$$

which results in

$$\mathrm{Exp}^{le}_{\mathbf{P}}(\mathbf{V}) = \exp(\log(\mathbf{P}) + \log(\mathbf{Q})).$$

We thus have

$$\mathrm{Exp}^{le}_{\mathbf{P}}(\mathcal{T}^{le}_{\mathbf{I}_n \to \mathbf{P}}(\mathrm{Log}^{le}_{\mathbf{I}_n}(\mathbf{Q}))) = \exp(\log(\mathbf{P}) + \log(\mathbf{Q})).$$

Therefore

$$\mathbf{P} \oplus_{le} \mathbf{Q} = \exp(\log(\mathbf{P}) + \log(\mathbf{Q})).$$

## 6 Proof of Lemma 4

*Proof.* The conclusion of the Lemma is straightforward since the Log-Euclidean exponential and logarithmic maps coincide respectively with the Affine-invariant exponential and logarithmic maps at the identity.

## 7 Proof of Theorem 2

*Proof.* First, note that the binary operation $\oplus_{le}$ verifies the Left Cancellation Law. From Eq. (26),

$$\begin{aligned} \mathrm{gyr}_{le}[\mathbf{P}, \mathbf{Q}]\mathbf{R} &= \big( \ominus_{le} (\mathbf{P} \oplus_{le} \mathbf{Q})\big) \oplus_{le} \big(\mathbf{P} \oplus_{le} (\mathbf{Q} \oplus_{le} \mathbf{R})\big) \\ &\overset{(1)}{=} \big( \ominus_{le} (\mathbf{P} \oplus_{le} \mathbf{Q})\big) \oplus_{le} \big((\mathbf{P} \oplus_{le} \mathbf{Q}) \oplus_{le} \mathbf{R}\big) \\ &\overset{(2)}{=} \mathbf{R}, \end{aligned} \tag{30}$$

The derivation of Eq. (30) follows.

(1) follows from the associativity of the binary operation $\oplus_{le}$.

(2) follows from the Left Cancellation Law.

It is then easy to verify axioms G1,G2,G4,V1,V2,V3,V4,V5 for Log-Euclidean gyrovector spaces.

## 8 Proof of Lemma 5

*Proof.* Let $\mathbf{W} = \mathcal{T}_{\mathbf{I}_n \to \mathbf{P}}(\mathrm{Log}^{ch}_{\mathbf{I}_n}(\mathbf{Q}))$. Then

$$\begin{aligned}\mathbf{P} \oplus_{lc} \mathbf{Q} &= \mathrm{Exp}^{ch}_{\mathbf{P}}(\mathbf{W}) \\ &\stackrel{(1)}{=} \widetilde{\mathrm{Exp}}_{\mathscr{L}(\mathbf{P})}(D_{\mathbf{P}}\mathscr{L}(\mathbf{W}))\big(\widetilde{\mathrm{Exp}}_{\mathscr{L}(\mathbf{P})}(D_{\mathbf{P}}\mathscr{L}(\mathbf{W}))\big)^T \\ &= \widetilde{\mathrm{Exp}}_{\mathscr{L}(\mathbf{P})}(\mathbf{W}')\big(\widetilde{\mathrm{Exp}}_{\mathscr{L}(\mathbf{P})}(\mathbf{W}')\big)^T,\end{aligned} \tag{31}$$

where (1) follows from Eq. (19), and $\mathbf{W}' = D_{\mathbf{P}}\mathscr{L}(\mathbf{W})$.

From Eq. (21),

$$\widetilde{\mathrm{Exp}}_{\mathscr{L}(\mathbf{P})}(\mathbf{W}') = \lfloor \mathscr{L}(\mathbf{P}) \rfloor + \lfloor \mathbf{W}' \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P})) \exp\big(\mathbb{D}(\mathbf{W}')\mathbb{D}(\mathscr{L}(\mathbf{P}))^{-1}\big). \tag{32}$$

From Eq. (25),

$$\mathbf{W} = \mathscr{L}(\mathbf{P})(\lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X))^T + (\lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X))\mathscr{L}(\mathbf{P})^T.$$

Hence

$$\begin{aligned}\mathscr{L}(\mathbf{P})^{-1}\mathbf{W}\mathscr{L}(\mathbf{P})^{-T} &= (\lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X))^T\mathscr{L}(\mathbf{P})^{-T} \\ &\quad + \mathscr{L}(\mathbf{P})^{-1}(\lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X)).\end{aligned}$$

Note that for any $\mathbf{L} \in \mathrm{L}_n$, $(\mathbf{L}+\mathbf{L}^T)_{\frac{1}{2}} = \mathbf{L}$. Since $\mathscr{L}(\mathbf{P})^{-1}(\lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X)) \in \mathrm{L}_n$, we get

$$\big(\mathscr{L}(\mathbf{P})^{-1}\mathbf{W}\mathscr{L}(\mathbf{P})^{-T}\big)_{\frac{1}{2}} = \mathscr{L}(\mathbf{P})^{-1}(\lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X)).$$

We thus have

$$\begin{aligned}\mathscr{L}(\mathbf{P})\big(\mathscr{L}(\mathbf{P})^{-1}\mathbf{W}\mathscr{L}(\mathbf{P})^{-T}\big)_{\frac{1}{2}} &= \mathscr{L}(\mathbf{P})\mathscr{L}(\mathbf{P})^{-1}(\lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X)) \\ &= \lfloor X \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(X) \\ &= \lfloor \mathscr{L}(\mathbf{Q}) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\log(\mathbb{D}(\mathscr{L}(\mathbf{Q}))).\end{aligned}$$

Therefore

$$\begin{aligned}\mathbf{W}' &= (D_{\mathbf{P}}\mathscr{L})(\mathbf{W}) \\ &\stackrel{(1)}{=} \mathscr{L}(\mathbf{P})\big(\mathscr{L}(\mathbf{P})^{-1}\mathbf{W}\mathscr{L}(\mathbf{P})^{-T}\big)_{\frac{1}{2}} \\ &= \lfloor \mathscr{L}(\mathbf{Q}) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\log(\mathbb{D}(\mathscr{L}(\mathbf{Q}))),\end{aligned} \tag{33}$$

where (1) follows from Eq. (20).

This leads to

$$\mathbb{D}(\mathbf{W}') = \mathbb{D}(\mathscr{L}(\mathbf{P}))\log(\mathbb{D}(\mathscr{L}(\mathbf{Q}))),$$

which means that

$$\exp\big(\mathbb{D}(\mathbf{W}')\mathbb{D}(\mathscr{L}(\mathbf{P}))^{-1}\big) = \mathbb{D}(\mathscr{L}(\mathbf{Q})). \tag{34}$$

From Eq. (33), we also get

$$\lfloor \mathbf{W}' \rfloor = \lfloor \mathscr{L}(\mathbf{Q}) \rfloor. \tag{35}$$

Combining Eqs. (31), (32), (34), and (35) leads to

$$\begin{aligned}\mathbf{P} \oplus_{lc} \mathbf{Q} =& \big(\lfloor \mathscr{L}(\mathbf{P}) \rfloor + \lfloor \mathscr{L}(\mathbf{Q}) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(\mathscr{L}(\mathbf{Q}))\big).\\ &\big(\lfloor \mathscr{L}(\mathbf{P}) \rfloor + \lfloor \mathscr{L}(\mathbf{Q}) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))\mathbb{D}(\mathscr{L}(\mathbf{Q}))\big)^T.\end{aligned}$$

## 9 Proof of Lemma 6

*Proof.* Let $\mathbf{W} = t \operatorname{Log}^{ch}_{\mathbf{I}_n}(\mathbf{P})$. Then

$$\begin{aligned} t \otimes_{lc} \mathbf{P} &= \operatorname{Exp}^{ch}_{\mathbf{I}_n}(\mathbf{W}) \\ &\overset{(1)}{=} \widetilde{\operatorname{Exp}}_{\mathbf{I}_n}(D_{\mathbf{I}_n}\mathscr{L}(\mathbf{W}))\big(\widetilde{\operatorname{Exp}}_{\mathbf{I}_n}(D_{\mathbf{I}_n}\mathscr{L}(\mathbf{W}))\big)^T, \end{aligned} \tag{36}$$

where (1) follows from Eq. (19).

Note that

$$\begin{aligned} \mathbf{W} &= t \operatorname{Log}^{ch}_{\mathbf{I}_n}(\mathbf{P}) \\ &\overset{(1)}{=} t D_{\mathbf{I}_n}\mathscr{S}\big(\widetilde{\operatorname{Log}}_{\mathbf{I}_n}(\mathscr{L}(\mathbf{P}))\big) \\ &\overset{(2)}{=} t\Big(\big(\widetilde{\operatorname{Log}}_{\mathbf{I}_n}(\mathscr{L}(\mathbf{P}))\big)^T + \widetilde{\operatorname{Log}}_{\mathbf{I}_n}(\mathscr{L}(\mathbf{P}))\Big), \end{aligned} \tag{37}$$

where (1) and (2) follow respectively from Eqs. (22) and (23).

Let $\mathbf{W}' = D_{\mathbf{I}_n}\mathscr{L}(\mathbf{W})$. Then

$$\begin{aligned} \mathbf{W}' &= D_{\mathbf{I}_n}\mathscr{L}(\mathbf{W}) \\ &\overset{(1)}{=} (\mathbf{W})_{\frac{1}{2}} \\ &\overset{(2)}{=} t\widetilde{\operatorname{Log}}_{\mathbf{I}_n}(\mathscr{L}(\mathbf{P})) \\ &\overset{(3)}{=} t\big(\lfloor \mathscr{L}(\mathbf{P}) \rfloor + \log(\mathbb{D}(\mathscr{L}(\mathbf{P})))\big). \end{aligned} \tag{38}$$

The derivation of Eq. (38) follows.
(1) follows from Eq. (20).
(2) follows from Eq. (37) and the fact that $t\widetilde{\operatorname{Log}}_{\mathbf{I}_n}(\mathscr{L}(\mathbf{P})) \in \mathrm{L}_n$.
(3) follows from Eq. (24).
Therefore

$$\begin{aligned} \widetilde{\operatorname{Exp}}_{\mathbf{I}_n}(D_{\mathbf{I}_n}\mathscr{L}(\mathbf{W})) &= \widetilde{\operatorname{Exp}}_{\mathbf{I}_n}(\mathbf{W}') \\ &\overset{(1)}{=} \lfloor \mathbf{W}' \rfloor + \exp(\mathbb{D}(\mathbf{W}')) \\ &\overset{(2)}{=} t\lfloor \mathscr{L}(\mathbf{P}) \rfloor + \exp(t \log(\mathbb{D}(\mathscr{L}(\mathbf{P})))) \\ &= t\lfloor \mathscr{L}(\mathbf{P}) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))^t, \end{aligned} \tag{39}$$

where (1) and (2) follow respectively from Eqs. (21) and (38).

Combining Eqs. (36) and (39) leads to

$$t \otimes_{lc} \mathbf{P} = \big(t\lfloor \mathscr{L}(\mathbf{P}) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))^t\big)\big(t\lfloor \mathscr{L}(\mathbf{P}) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}))^t\big)^T.$$

## 10 Proof of Theorem 3

*Proof.* Note that the binary operation $\oplus_{lc}$ verifies the Left Cancellation Law. From Eq. (26),

$$\begin{aligned}
\mathrm{gyr}_{lc}[\mathbf{P},\mathbf{Q}]\mathbf{R} &= \big(\ominus_{lc}(\mathbf{P} \oplus_{lc} \mathbf{Q})\big) \oplus_{lc} \big(\mathbf{P} \oplus_{lc} (\mathbf{Q} \oplus_{lc} \mathbf{R})\big) \\
&\overset{(1)}{=} \big(\ominus_{lc}(\mathbf{P} \oplus_{lc} \mathbf{Q})\big) \oplus_{lc} \big((\mathbf{P} \oplus_{lc} \mathbf{Q}) \oplus_{lc} \mathbf{R}\big) \\
&\overset{(2)}{=} \mathbf{R},
\end{aligned} \tag{40}$$

The derivation of Eq. (40) follows.

(1) follows from the associativity of the binary operation $\oplus_{lc}$.

(2) follows from the Left Cancellation Law.

It is then easy to verify axioms G1,G2,G4,V1,V2,V3,V4,V5 for Log-Cholesky gyrovector spaces.

## 11 Proof of Lemma 7

### Affine-invariant gyrovector spaces.

*Proof.* By assumption that $\mathrm{Log}^{ai}_{\mathbf{P}_1} \mathbf{Q}_1$ is the Affine-invariant parallel transport of $\mathrm{Log}^{ai}_{\mathbf{P}_0} \mathbf{Q}_0$ from $\mathbf{P}_0$ to $\mathbf{P}_1$ along geodesics connecting $\mathbf{P}_0$ and $\mathbf{P}_1$, then from Eq. (16),

$$\mathrm{Log}^{ai}_{\mathbf{P}_1}(\mathbf{Q}_1) = \mathbf{R}\,\mathrm{Log}^{ai}_{\mathbf{P}_0}(\mathbf{Q}_0)\mathbf{R}^T,$$

where $\mathbf{R} = (\mathbf{P}_1\mathbf{P}_0^{-1})^{\frac{1}{2}}$.

Thus

$$\begin{aligned}
\mathbf{Q}_1 &= \mathrm{Exp}^{ai}_{\mathbf{P}_1}(\mathbf{R}\,\mathrm{Log}^{ai}_{\mathbf{P}_0}(\mathbf{Q}_0)\mathbf{R}^T) \\
&= \mathbf{P}_1^{\frac{1}{2}} \exp(\mathbf{P}_1^{-\frac{1}{2}}\mathbf{R}\,\mathrm{Log}^{ai}_{\mathbf{P}_0}(\mathbf{Q}_0)\mathbf{R}^T\mathbf{P}_1^{-\frac{1}{2}})\mathbf{P}_1^{\frac{1}{2}}.
\end{aligned} \tag{41}$$

Note that

$$\mathrm{Log}^{ai}_{\mathbf{P}_0}(\mathbf{Q}_0) = \mathbf{P}_0^{\frac{1}{2}} \log(\mathbf{P}_0^{-\frac{1}{2}}\mathbf{Q}_0\mathbf{P}_0^{-\frac{1}{2}})\mathbf{P}_0^{\frac{1}{2}}.$$

Hence from Eq. (41),

$$\begin{aligned}
\mathbf{Q}_1 &= \mathbf{P}_1^{\frac{1}{2}} \exp(\mathbf{U}\log(\mathbf{P}_0^{-\frac{1}{2}}\mathbf{Q}_0\mathbf{P}_0^{-\frac{1}{2}})\mathbf{U}^T)\mathbf{P}_1^{\frac{1}{2}} \\
&\overset{(1)}{=} \mathbf{P}_1^{\frac{1}{2}}\mathbf{U}\exp(\log(\mathbf{P}_0^{-\frac{1}{2}}\mathbf{Q}_0\mathbf{P}_0^{-\frac{1}{2}}))\mathbf{U}^T\mathbf{P}_1^{\frac{1}{2}} \\
&= \mathbf{P}_1^{\frac{1}{2}}\mathbf{U}\mathbf{P}_0^{-\frac{1}{2}}\mathbf{Q}_0\mathbf{P}_0^{-\frac{1}{2}}\mathbf{U}^T\mathbf{P}_1^{\frac{1}{2}} \\
&= \mathbf{R}\mathbf{Q}_0\mathbf{R}^T = (\mathbf{P}_1\mathbf{P}_0^{-1})^{\frac{1}{2}}\mathbf{Q}_0((\mathbf{P}_1\mathbf{P}_0^{-1})^{\frac{1}{2}})^T,
\end{aligned} \tag{42}$$

where $\mathbf{U} = \mathbf{P}_1^{-\frac{1}{2}} \mathbf{R} \mathbf{P}_0^{\frac{1}{2}}$, and (1) follows from the fact that $\mathbf{U}\mathbf{U}^T = \mathbf{I}_n$.

From Eq. (5),

$$\mathrm{gyr}_{ai}[\mathbf{P}_1, \ominus_{ai}\mathbf{P}_0](\ominus_{ai}\mathbf{P}_0 \oplus_{ai} \mathbf{Q}_0) = (\mathbf{P}_1^{\frac{1}{2}}\mathbf{P}_0^{-1}\mathbf{P}_1^{\frac{1}{2}})^{-\frac{1}{2}}\mathbf{P}_1^{\frac{1}{2}}\mathbf{P}_0^{-1}\mathbf{Q}_0\mathbf{P}_1^{-\frac{1}{2}}(\mathbf{P}_1^{\frac{1}{2}}\mathbf{P}_0^{-1}\mathbf{P}_1^{\frac{1}{2}})^{\frac{1}{2}}.$$

Let $\mathbf{B} = (\mathbf{P}_1^{\frac{1}{2}}\mathbf{P}_0^{-1}\mathbf{P}_1^{\frac{1}{2}})^{-\frac{1}{2}}\mathbf{P}_1^{\frac{1}{2}}\mathbf{P}_0^{-1}$, $\mathbf{C} = \mathbf{P}_1^{-\frac{1}{2}}(\mathbf{P}_1^{\frac{1}{2}}\mathbf{P}_0^{-1}\mathbf{P}_1^{\frac{1}{2}})^{\frac{1}{2}}$. Then

$$\mathrm{gyr}_{ai}[\mathbf{P}_1, \ominus_{ai}\mathbf{P}_0](\ominus_{ai}\mathbf{P}_0 \oplus_{ai} \mathbf{Q}_0) = \mathbf{B}\mathbf{Q}_0\mathbf{C}.$$

Hence

$$\mathbf{P}_1^{\frac{1}{2}}\,\mathrm{gyr}_{ai}[\mathbf{P}_1, \ominus_{ai}\mathbf{P}_0](\ominus_{ai}\mathbf{P}_0 \oplus_{ai} \mathbf{Q}_0)\mathbf{P}_1^{\frac{1}{2}} = \mathbf{P}_1^{\frac{1}{2}}\mathbf{B}\mathbf{Q}_0\mathbf{C}\mathbf{P}_1^{\frac{1}{2}}. \tag{43}$$

We remark that

$$(\mathbf{P}_1^{\frac{1}{2}}\mathbf{B})^2 = \mathbf{P}_1^{\frac{1}{2}}\mathbf{B}\mathbf{P}_1^{\frac{1}{2}}\mathbf{B} = \mathbf{P}_1\mathbf{P}_0^{-1},$$

$$(\mathbf{C}\mathbf{P}_1^{\frac{1}{2}})^2 = \mathbf{C}\mathbf{P}_1^{\frac{1}{2}}\mathbf{C}\mathbf{P}_1^{\frac{1}{2}} = (\mathbf{P}_1\mathbf{P}_0^{-1})^T.$$

Therefore

$$\mathbf{P}_1^{\frac{1}{2}}\mathbf{B}\mathbf{Q}_0\mathbf{C}\mathbf{P}_1^{\frac{1}{2}} = (\mathbf{P}_1\mathbf{P}_0^{-1})^{\frac{1}{2}}\mathbf{Q}_0((\mathbf{P}_1\mathbf{P}_0^{-1})^{\frac{1}{2}})^T. \tag{44}$$

Combining Eqs. (42), (43), and (44), we obtain

$$\mathbf{Q}_1 = \mathbf{P}_1^{\frac{1}{2}}\,\mathrm{gyr}_{ai}[\mathbf{P}_1, \ominus_{ai}\mathbf{P}_0](\ominus_{ai}\mathbf{P}_0 \oplus_{ai} \mathbf{Q}_0)\mathbf{P}_1^{\frac{1}{2}},$$

which leads to

$$\ominus_{ai}\mathbf{P}_1 \oplus_{ai} \mathbf{Q}_1 = \mathrm{gyr}_{ai}[\mathbf{P}_1, \ominus_{ai}\mathbf{P}_0](\ominus_{ai}\mathbf{P}_0 \oplus_{ai} \mathbf{Q}_0).$$

**Log-Euclidean gyrovector spaces.**

*Proof.* Notice that

$$\begin{aligned}\mathbf{Q}_0 &= \mathrm{Exp}^{le}_{\mathbf{P}_0}(\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0)) \\ &\overset{(1)}{=} \exp\big(\log(\mathbf{P}_0) + D_{\mathbf{P}_0}\log\big(\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0)\big)\big) \\ &\overset{(2)}{=} \exp\big(\log(\mathbf{P}_0) + \mathcal{T}_{\mathbf{P}_0\to\mathbf{I}_n}(\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0))\big)\end{aligned}$$

where (1) and (2) follow respectively from Eqs. (17) and (29).

It is known [53] that a SPD matrix has a unique symmetric logarithm, and since $\log(\mathbf{P}_0) + \mathcal{T}_{\mathbf{P}_0\to\mathbf{I}_n}(\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0))$ is symmetric, we have

$$\log(\mathbf{Q}_0) = \log(\mathbf{P}_0) + \mathcal{T}_{\mathbf{P}_0\to\mathbf{I}_n}(\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0)).$$

We thus get

$$\log(\mathbf{Q}_0) - \log(\mathbf{P}_0) = \mathcal{T}_{\mathbf{P}_0\to\mathbf{I}_n}(\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0)). \tag{45}$$

Similarly, we have

$$\log(\mathbf{Q}_1) - \log(\mathbf{P}_1) = \mathcal{T}_{\mathbf{P}_1 \to \mathbf{I}_n}(\mathrm{Log}^{le}_{\mathbf{P}_1}(\mathbf{Q}_1)). \tag{46}$$

Since $\mathrm{Syn}^+_n$ are complete, simply-connected and flat manifolds [2], the parallel transport is path independent. By the assumption that $\mathrm{Log}^{le}_{\mathbf{P}_1}(\mathbf{Q}_1)$ is the Log-Euclidean parallel transport of $\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0)$ from $\mathbf{P}_0$ to $\mathbf{P}_1$ along geodesics connecting $\mathbf{P}_0$ and $\mathbf{P}_1$, we deduce that

$$\mathcal{T}_{\mathbf{P}_0 \to \mathbf{I}_n}(\mathrm{Log}^{le}_{\mathbf{P}_0}(\mathbf{Q}_0)) = \mathcal{T}_{\mathbf{P}_1 \to \mathbf{I}_n}(\mathrm{Log}^{le}_{\mathbf{P}_1}(\mathbf{Q}_1)). \tag{47}$$

Combining Eqs. (45), (46), and (47) results in

$$\log(\mathbf{Q}_0) - \log(\mathbf{P}_0) = \log(\mathbf{Q}_1) - \log(\mathbf{P}_1),$$

which leads to

$$\ominus_{le}\mathbf{P}_1 \oplus_{le} \mathbf{Q}_1 = \ominus_{le}\mathbf{P}_0 \oplus_{le} \mathbf{Q}_0.$$

Therefore

$$\ominus_{le}\mathbf{P}_1 \oplus_{le} \mathbf{Q}_1 = \mathrm{gyr}_{le}[\mathbf{P}_1, \ominus_{le}\mathbf{P}_0](\ominus_{le}\mathbf{P}_0 \oplus_{le} \mathbf{Q}_0).$$

**Log-Cholesky gyrovector spaces.**

*Proof.* We have

$$\begin{aligned}\mathcal{T}_{\mathbf{P}_0 \to \mathbf{P}_1}(\mathrm{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0)) =& \mathscr{L}(\mathbf{P}_1)\big(\lfloor \mathbf{X} \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathbf{X})\big)^T + \\ &+ \big(\lfloor \mathbf{X} \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathbf{X})\big)\mathscr{L}(\mathbf{P}_1)^T,\end{aligned} \tag{48}$$

where $\mathbf{X} = \mathscr{L}(\mathbf{P}_0)\big(\mathscr{L}(\mathbf{P}_0)^{-1}\,\mathrm{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0)\mathscr{L}(\mathbf{P}_0)^{-T}\big)_{\frac{1}{2}}$.

Note that

$$\begin{aligned}\mathrm{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0) &\overset{(1)}{=} D_{\mathscr{L}(\mathbf{P}_0)}\mathscr{S}\big(\widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0))\big) \\ &\overset{(2)}{=} \mathscr{L}(\mathbf{P}_0)\big(\widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0))\big)^T + \widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0))\mathscr{L}(\mathbf{P}_0)^T,\end{aligned}$$

where (1) and (2) follow respectively from Eqs. (22) and (23).

Thus

$$\begin{aligned}\mathscr{L}(\mathbf{P}_0)^{-1}\,\mathrm{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0)\mathscr{L}(\mathbf{P}_0)^{-T} =& \big(\widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0))\big)^T \mathscr{L}(\mathbf{P}_0)^{-T} + \\ &+ \mathscr{L}(\mathbf{P}_0)^{-1}\widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0)).\end{aligned}$$

Since $\mathscr{L}(\mathbf{P}_0)^{-1}\widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0)) \in \mathrm{L}_n$, we get

$$\big(\mathscr{L}(\mathbf{P}_0)^{-1}\,\mathrm{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0)\mathscr{L}(\mathbf{P}_0)^{-T}\big)_{\frac{1}{2}} = \mathscr{L}(\mathbf{P}_0)^{-1}\widetilde{\mathrm{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0)).$$

Therefore

$$\begin{aligned}\mathbf{X} &= \mathscr{L}(\mathbf{P}_0)\big(\mathscr{L}(\mathbf{P}_0)^{-1}\operatorname{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0)\mathscr{L}(\mathbf{P}_0)^{-T}\big)_{\frac{1}{2}}\\&= \mathscr{L}(\mathbf{P}_0)\mathscr{L}(\mathbf{P}_0)^{-1}\widetilde{\operatorname{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0))\\&= \widetilde{\operatorname{Log}}_{\mathscr{L}(\mathbf{P}_0)}(\mathscr{L}(\mathbf{Q}_0))\\&\overset{(1)}{=} \lfloor\mathscr{L}(\mathbf{Q}_0)\rfloor - \lfloor\mathscr{L}(\mathbf{P}_0)\rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_0))\log\big(\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathscr{L}(\mathbf{Q}_0))\big),\end{aligned} \tag{49}$$

where (1) follows from Eq. (24).

Let $\mathbf{W} = \mathcal{T}_{\mathbf{P}_0\to\mathbf{P}_1}(\operatorname{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0))$. Replace $\mathbf{X}$ in Eq. (48) with its expression in Eq. (49), we get

$$\begin{aligned}\mathbf{W} =&\mathscr{L}(\mathbf{P}_1)\Big(\lfloor\mathscr{L}(\mathbf{Q}_0)\rfloor - \lfloor\mathscr{L}(\mathbf{P}_0)\rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\log\big(\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathscr{L}(\mathbf{Q}_0))\big)\Big)^T\\&+\Big(\lfloor\mathscr{L}(\mathbf{Q}_0)\rfloor - \lfloor\mathscr{L}(\mathbf{P}_0)\rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\log\big(\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathscr{L}(\mathbf{Q}_0))\big)\Big)\mathscr{L}(\mathbf{P}_1)^T.\end{aligned}$$

Let $\mathbf{U} = \lfloor\mathscr{L}(\mathbf{Q}_0)\rfloor - \lfloor\mathscr{L}(\mathbf{P}_0)\rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\log\big(\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathscr{L}(\mathbf{Q}_0))\big)$. Then

$$\mathscr{L}(\mathbf{P}_1)^{-1}\mathbf{W}\mathscr{L}(\mathbf{P}_1)^{-T} =\mathbf{U}^T\mathscr{L}(\mathbf{P}_1)^{-T} + \mathscr{L}(\mathbf{P}_1)^{-1}\mathbf{U}.$$

Therefore

$$\begin{aligned}D_{\mathbf{P}_1}\mathscr{L}(\mathbf{W}) &\overset{(1)}{=} \mathscr{L}(\mathbf{P}_1)\big(\mathscr{L}(\mathbf{P}_1)^{-1}\mathbf{W}\mathscr{L}(\mathbf{P}_1)^{-T}\big)_{\frac{1}{2}}\\&= \mathscr{L}(\mathbf{P}_1)\big(\mathbf{U}^T\mathscr{L}(\mathbf{P}_1)^{-T} + \mathscr{L}(\mathbf{P}_1)^{-1}\mathbf{U}\big)_{\frac{1}{2}}\\&\overset{(2)}{=} \mathscr{L}(\mathbf{P}_1)\mathscr{L}(\mathbf{P}_1)^{-1}\mathbf{U}\\&= \mathbf{U},\end{aligned} \tag{50}$$

where (1) follows from Eq. (20), and (2) follows from the fact that $\mathscr{L}(\mathbf{P}_1)^{-1}\mathbf{U} \in \mathrm{L}_n$.

By assumption

$$\begin{aligned}\mathbf{Q}_1 &= \operatorname{Exp}^{ch}_{\mathbf{P}_1}(\mathcal{T}_{\mathbf{P}_0\to\mathbf{P}_1}(\operatorname{Log}^{ch}_{\mathbf{P}_0}(\mathbf{Q}_0)))\\&\overset{(1)}{=} \widetilde{\operatorname{Exp}}_{\mathscr{L}(\mathbf{P}_1)}\big(D_{\mathbf{P}_1}\mathscr{L}(\mathbf{W})\big)\Big(\widetilde{\operatorname{Exp}}_{\mathscr{L}(\mathbf{P}_1)}\big(D_{\mathbf{P}_1}\mathscr{L}(\mathbf{W})\big)\Big)^T,\end{aligned}$$

where (1) follows from Eq. (19). We thus have

$$\begin{aligned}\mathscr{L}(\mathbf{Q}_1) &= \widetilde{\operatorname{Exp}}_{\mathscr{L}(\mathbf{P}_1)}\big(D_{\mathbf{P}_1}\mathscr{L}(\mathbf{W})\big)\\&\overset{(1)}{=} \lfloor D_{\mathbf{P}_1}\mathscr{L}(\mathbf{W})\rfloor + \lfloor\mathscr{L}(\mathbf{P}_1)\rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\exp\big(\mathbb{D}(D_{\mathbf{P}_1}\mathscr{L}(\mathbf{W}))\mathbb{D}(\mathscr{L}(\mathbf{P}_1))^{-1}\big)\\&\overset{(2)}{=} \lfloor\mathbf{U}\rfloor + \lfloor\mathscr{L}(\mathbf{P}_1)\rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\exp\big(\mathbb{D}(\mathbf{U})\mathbb{D}(\mathscr{L}(\mathbf{P}_1))^{-1}\big)\\&= \lfloor\mathscr{L}(\mathbf{Q}_0)\rfloor - \lfloor\mathscr{L}(\mathbf{P}_0)\rfloor + \lfloor\mathscr{L}(\mathbf{P}_1)\rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathscr{L}(\mathbf{Q}_0)),\end{aligned}$$

where (1) and (2) follow respectively from Eqs. (21) and (50). We deduce that

$$\lfloor \mathscr{L}(\mathbf{Q}_1) \rfloor = \lfloor \mathscr{L}(\mathbf{Q}_0) \rfloor - \lfloor \mathscr{L}(\mathbf{P}_0) \rfloor + \lfloor \mathscr{L}(\mathbf{P}_1) \rfloor, \tag{51}$$

and

$$\mathbb{D}(\mathscr{L}(\mathbf{Q}_1)) = \mathbb{D}(\mathscr{L}(\mathbf{P}_1))\mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathscr{L}(\mathbf{Q}_0)). \tag{52}$$

Note that

$$\mathscr{L}(\ominus_{lc}\mathbf{P}_0 \oplus_{lc} \mathbf{Q}_0) = \lfloor \mathscr{L}(\mathbf{Q}_0) \rfloor - \lfloor \mathscr{L}(\mathbf{P}_0) \rfloor + \mathbb{D}(\mathscr{L}(\mathbf{P}_0))^{-1}\mathbb{D}(\mathscr{L}(\mathbf{Q}_0)). \tag{53}$$

Combining Eqs. (51), (52), and (53) leads to

$$\mathscr{L}(\ominus_{lc}\mathbf{P}_1 \oplus_{lc} \mathbf{Q}_1) = \mathscr{L}(\ominus_{lc}\mathbf{P}_0 \oplus_{lc} \mathbf{Q}_0),$$

or equivalently,

$$\ominus_{lc}\mathbf{P}_1 \oplus_{lc} \mathbf{Q}_1 = \ominus_{lc}\mathbf{P}_0 \oplus_{lc} \mathbf{Q}_0.$$

Therefore

$$\ominus_{lc}\mathbf{P}_1 \oplus_{lc} \mathbf{Q}_1 = \mathrm{gyr}_{lc}[\mathbf{P}_1, \ominus_{lc}\mathbf{P}_0](\ominus_{lc}\mathbf{P}_0 \oplus_{lc} \mathbf{Q}_0).$$

## 12 Proof of Theorem 4

*Proof.* We first prove that the basic operations on $\mathrm{S}^+_{n,p}$ verify the Left Gyroassociative Law. Let $\mathbf{P}, \mathbf{Q}, \mathbf{R} \in \mathrm{S}^+_{n,p}$, $\mathbf{P} = \mathbf{U}_P \overline{\mathbf{P}} \mathbf{U}_P^T$, $\mathbf{Q} = \mathbf{U}_Q \overline{\mathbf{Q}} \mathbf{U}_Q^T$, and $\mathbf{R} = \mathbf{U}_R \overline{\mathbf{R}} \mathbf{U}_R^T$. According to Definition [9],

$$\mathbf{Q} \oplus_{spsd} \mathbf{R} = \mathbf{U}_e(\mathbf{O}^T_{\mathbf{U}_Q \to \mathbf{U}_e} \overline{\mathbf{Q}} \mathbf{O}_{\mathbf{U}_Q \to \mathbf{U}_e} \oplus \mathbf{O}^T_{\mathbf{U}_R \to \mathbf{U}_e} \overline{\mathbf{R}} \mathbf{O}_{\mathbf{U}_R \to \mathbf{U}_e}) \mathbf{U}_e^T.$$

Hence

$$\mathbf{P} \oplus_{spsd} (\mathbf{Q} \oplus_{spsd} \mathbf{R}) = \mathbf{U}_e \big(\mathbf{P}' \oplus (\mathbf{Q}' \oplus \mathbf{R}')\big) \mathbf{U}_e^T, \tag{54}$$

where $\mathbf{P}' = \mathbf{O}^T_{\mathbf{U}_P \to \mathbf{U}_e} \overline{\mathbf{P}} \mathbf{O}_{\mathbf{U}_P \to \mathbf{U}_e}$, $\mathbf{Q}' = \mathbf{O}^T_{\mathbf{U}_Q \to \mathbf{U}_e} \overline{\mathbf{Q}} \mathbf{O}_{\mathbf{U}_Q \to \mathbf{U}_e}$,
and $\mathbf{R}' = \mathbf{O}^T_{\mathbf{U}_R \to \mathbf{U}_e} \overline{\mathbf{R}} \mathbf{O}_{\mathbf{U}_R \to \mathbf{U}_e}$.

From the definition of the gyroautomorphism in $\mathrm{S}^+_{n,p}$,

$$(\mathbf{P} \oplus_{spsd} \mathbf{Q}) \oplus_{spsd} \mathrm{gyr}_{spsd}[\mathbf{P}, \mathbf{Q}]\mathbf{R} = \mathbf{U}_e \big((\mathbf{P}' \oplus \mathbf{Q}') \oplus \mathrm{gyr}[\mathbf{P}', \mathbf{Q}']\mathbf{R}'\big) \mathbf{U}_e^T. \tag{55}$$

Combining Eqs. (54), (55), and the Left Gyroassociative Law in gyrovector spaces of SPD matrices, we get

$$\mathbf{P} \oplus_{spsd} (\mathbf{Q} \oplus_{spsd} \mathbf{R}) = (\mathbf{P} \oplus_{spsd} \mathbf{Q}) \oplus_{spsd} \mathrm{gyr}_{spsd}[\mathbf{P}, \mathbf{Q}]\mathbf{R}.$$

Similarly, one can prove that the basic operations on $\mathrm{S}^+_{n,p}$ verify the Left Reduction Property, Gyrocommutative Law, and axioms (V2), (V3), and (V4).

## 13 Derivation of Our SPD Neural Networks

We first introduce a definition of gyroderivative in Gyrovector Spaces of SPD Matrices.

**Definition 12 (Gyroderivative in Gyrovector Spaces of SPD Matrices).** *Let* $(\mathrm{Sym}_n^+, \oplus, \otimes)$ *be a gyrovector space, and* $h : \mathbb{R} \rightarrow \mathrm{Sym}_n^+$ *be a map. If the limit*

$$\frac{dh}{dt}(t) = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \otimes (\ominus h(t) \oplus h(t + \delta t))$$

*exists for any* $t \in \mathbb{R}$*, then the map* $h$ *is said to be differentiable on* $\mathbb{R}$*, and the gyroderivative of* $h(t)$ *is* $\frac{dh}{dt}(t)$*.*

Note that the gyroderivative considered here is different from the derivative used for computing tangent vectors to a curve on manifolds [51], which is a map from a set of smooth real-valued functions to $\mathbb{R}$. From Definition 12, we can derive the chain rule in gyrovector spaces of SPD matrices similar to the Gyro-chain-rule [12] in hyperbolic spaces.

**Lemma 8 (Gyro-chain-rule in Gyrovector Spaces of SPD Matrices).** *Let* $g : \mathbb{R} \rightarrow \mathbb{R}$ *be a differentiable map, and* $h : \mathbb{R} \rightarrow \mathrm{Sym}_n^+$ *be a map with a well-defined gyroderivative in a gyrovector space* $(\mathrm{Sym}_n^+, \oplus, \otimes)$*. If* $f := h \circ g$*, then*

$$\frac{df}{dt}(t) = \frac{dg}{dt}(t) \otimes \frac{dh}{dt}(g(t)),$$

*where* $\frac{dg}{dt}(t)$ *is the ordinary derivative.*

*Proof.* The Lemma can be proved by applying the techniques in [12].

$$\begin{aligned}\frac{df}{dt}(t) &= \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \otimes (\ominus f(t) \oplus f(t + \delta t)) \\ &= \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \otimes (\ominus h(g(t)) \oplus h(g(t) + \delta t(g'(t) + \mathcal{O}(\delta t)))).\end{aligned}$$

Let $l_1 = \frac{g'(t)}{\delta t(g'(t) + \mathcal{O}(\delta t))}$, $l_2 = \frac{\mathcal{O}(\delta t)}{\delta t(g'(t) + \mathcal{O}(\delta t))}$. Then $\frac{1}{\delta t} = l_1 + l_2$ and we have

$$\frac{df}{dt}(t) = \lim_{\delta t \rightarrow 0} (l_1 + l_2) \otimes (\ominus h(g(t)) \oplus h(g(t) + \delta t(g'(t) + \mathcal{O}(\delta t)))).$$

Let $L_1 = \lim_{\delta t \rightarrow 0} l_1 \otimes (\ominus h(g(t)) \oplus h(g(t) + \delta t(g'(t) + \mathcal{O}(\delta t))))$, $L_2 = \lim_{\delta t \rightarrow 0} l_2 \otimes (\ominus h(g(t)) \oplus h(g(t) + \delta t(g'(t) + \mathcal{O}(\delta t))))$. Then by axiom (V2),

$$\frac{df}{dt}(t) = L_1 \oplus L_2.$$

Note that

$$L_2 = 0 \otimes (\ominus h(g(t)) \oplus h(g(t) + \delta t(g'(t) + \mathcal{O}(\delta t)))) = \mathbf{I}$$

by axiom (V1). Hence

$$\frac{df}{dt}(t) = L_1 \oplus \mathbf{I} = L_1.$$

Let $u = \delta t(g'(t) + \mathcal{O}(\delta t))$. Then

$$\frac{df}{dt}(t) = \lim_{u \to 0} \frac{g'(t)}{u} \otimes (\ominus h(g(t)) \oplus h(g(t) + u)).$$

Note that $\frac{g'(t)}{u} = g'(t) \frac{1}{g(t)+u-g(t)}$. Then by axiom (V3), we get

$$\frac{df}{dt}(t) = \frac{dg}{dt}(t) \otimes \frac{dh}{dt}(g(t)).$$

We now derive the update equations for our models. We consider a class of models that are invariant to time rescaling. Following [12,43], we first study time transformations in the continuous-time setting and then translate continuous-time models back to the discrete-time setting. In the following, we use indices $h_t$ for discrete time and brackets $h(t)$ for continuous time. From the definition of gyroderivative in gyrovector spaces of SPD matrices, using axiom (V3) and the Left Cancellation Law, we have

$$h(t + \delta t) \approx h(t) \oplus \delta t \otimes \frac{dh}{dt}(t) \tag{56}$$

for small $\delta t$. Let $T$ be a time variable and $H(T) = h(\alpha T)$, $X(T) = x(\alpha T)$. Using the chain rule in gyrovector spaces of SPD matrices, we obtain

$$\frac{dH}{dT}(T) = \alpha \otimes \frac{dh}{dT}(\alpha T). \tag{57}$$

Let $h(t+1) = \psi(h(t), x(t))$[2]. Note that Eq. (56) is equivalent to

$$\ominus h(t) \oplus h(t + \delta t) \approx \delta t \otimes \frac{dh}{dt}(t).$$

With discretization step $\delta t = 1$, we have

$$\begin{aligned} \frac{dh}{dT}(\alpha T) &= \ominus H(T) \oplus h(\alpha T + 1) \\ &= \ominus H(T) \oplus \psi(H(T), X(T)). \end{aligned}$$

Eq. (57) now becomes

$$\frac{dH}{dT}(T) = \alpha \otimes \big( \ominus H(T) \oplus \psi(H(T), X(T)) \big).$$

By renaming $H$ to $h$, $X$ to $x$, and $T$ to $t$, we obtain

$$\frac{dh}{dt}(t) = \alpha \otimes \big( \ominus h(t) \oplus \psi(h(t), x(t)) \big),$$

[2] We drop model parameters to simplify notations.

which results in

$$h(t) \oplus \frac{dh}{dt}(t) = h(t) \oplus \alpha \otimes \big( \ominus h(t) \oplus \psi(h(t), x(t))\big).$$

According to Eq. (56), we have $h(t+1) = h(t) \oplus \frac{dh}{dt}(t)$. Then

$$h(t+1) = h(t) \oplus \alpha \otimes \big( \ominus h(t) \oplus \psi(h(t), x(t))\big). \tag{58}$$

Now, setting $\psi(h(t), x(t)) = \varphi^{\otimes_a}(\mathbf{W}_h \otimes_v h(t) + \mathbf{W}_x \otimes_v \phi(x(t)))$ and translating Eq. (58) back to discrete-time models, we obtain the following recurrent equations

$$\mathbf{P}_t = \varphi^{\otimes_a}(\mathbf{W}_h \otimes_v \mathbf{H}_{t-1} + \mathbf{W}_x \otimes_v \phi(\mathbf{X}_t)),$$

$$\mathbf{H}_t = \mathbf{H}_{t-1} \oplus \alpha \otimes ((\ominus \mathbf{H}_{t-1}) \oplus \mathbf{P}_t),$$

where $\mathbf{X}_t, \mathbf{P}_t, \mathbf{H}_{t-1}, \mathbf{H}_t \in Sym_n^+$, $\mathbf{W}_h, \mathbf{W}_x \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$ are learnable parameters.

## 14 More Details on Our Implementation and Training

The procedure SpdRotate(.) proceeds as follows. We first compute two bases of $\text{span}(\mathbf{X}^u)$ and $\text{span}(\mathbf{W})$, denoted respectively by $\overline{\mathbf{X}}$ and $\overline{\mathbf{W}}$, such that

$$d_{\mathrm{V}(n,p)}(\overline{\mathbf{X}}, \overline{\mathbf{W}}) = d_{\mathrm{Gr}(n,p)}(\text{span}(\mathbf{X}^u), \text{span}(\mathbf{W})),$$

where $d_{\mathrm{V}(n,p)}(.,.)$ and $d_{\mathrm{Gr}(n,p)}(.,.)$ are the distances between two points in $\mathrm{V}(n,p)$ and $\mathrm{Gr}(n,p)$, respectively. From [3], $\overline{\mathbf{X}} = \mathbf{X}^u \mathbf{Y}$, $\overline{\mathbf{W}} = \mathbf{W}\mathbf{V}$, where $\mathbf{Y}$ and $\mathbf{V}$ are obtained from the SVD of $(\mathbf{X}^u)^T \mathbf{W}$, i.e.,

$$(\mathbf{X}^u)^T \mathbf{W} = \mathbf{Y}(\cos \boldsymbol{\Sigma})\mathbf{V}^T.$$

The output of SpdRotate($\cdot$) is then computed as

$$\mathbf{X}^{spd} = \mathbf{V}\overline{\mathbf{X}}^T \mathbf{X}_t \overline{\mathbf{X}} \mathbf{V}^T.$$

For the first experiment in the ablation study (**Full-rank vs. low-rank learning**), since we only consider learning the parameter $\mathbf{W}$ in the full-rank case, we transform the problem of learning the constrained parameter (rotation matrix) $\mathbf{W}$ into the one of learning an unconstrained parameter using the Cayley parameterization [54,55]. This method is based on the transform

$$\mathbf{W} = (\mathbf{I}_n + \mathbf{A})(\mathbf{I}_n - \mathbf{A})^{-1},$$

which constructs a rotation matrix $\mathbf{W}$ from a skew-symmetric matrix $\mathbf{A}$. We parametrize the rotation matrix $\mathbf{W}$ through the upper or lower triangular part of the matrix $\mathbf{A}$, which results in $\frac{n(n-1)}{2}$ trainable parameters. Other advanced methods [54,55] can also be used for improving performance.

| Dataset | HDM05 | FPHA | NTU60 (X-Sub) | NTU60 (X-View) |
|---|---|---|---|---|
| $p = 14$ | 74.19 | 90.94 | 89.78 | 90.15 |
| $p = 18$ | 76.47 | 92.35 | 91.14 | 92.56 |
| $p = 22$ | 77.05 | 95.47 | 93.54 | 94.19 |

**Table 6.** Accuracies of SPSD-AI for different values of $p$ in the low-rank learning setting ($n = 28$, $\mathbf{W} = \mathbf{I}_{n,p}[:, :p]$).

We use a temporal pyramid representation for each sequence. At temporal pyramid $M$, a sequence is partitioned into $M$ subsequences of equal size. Each subsequence is fed to a model with its own parameter set. The outputs from all the models are concatenated to create a final representation of the sequence. In our experiments, the number of temporal pyramids $M$ is set to 3.

For all the datasets, we use interpolation to create sequences of the same length (100 frames in our experiments). We use a batch size of 32 for HDM05 and FPHA datasets, and a batch size of 256 for NTU60 dataset. For our networks and the methods whose codes are used in our experiments, we run each model three times and report the best accuracy from these three runs [12].

## 15 More Results

Tab. 6 reports the accuracies of SPSD-AI for different values of $p$ in the low-rank learning setting when the parameter $\mathbf{W}$ is fixed ($\mathbf{W} = \mathbf{I}_{n,p}[:, :p]$). The accuracies of SPSD-AI increase as the rank $p$ increases. These results show the important impact of the rank $p$ on the performance of SPSD-AI.

Tab. 7 reports the accuracies and computation times of SPSD-AI in the full-rank learning setting with and without the parameter $\mathbf{W}$. When the parameter $\mathbf{W}$ is not used, our models are described by the following update equations:

$$\mathbf{P}_t = \varphi^{\otimes_a}(\mathbf{W}_h \otimes_v \mathbf{H}_{t-1} + \mathbf{W}_x \otimes_v \mathbf{X}_t),$$

$$\mathbf{H}_t = \mathbf{H}_{t-1} \oplus \alpha \otimes ((\ominus \mathbf{H}_{t-1}) \oplus \mathbf{P}_t).$$

In all cases, the introduction of the parameter $\mathbf{W}$ yields better results. However, this also results in higher computational costs.

Tab. 8 shows the accuracies and computation times of SPSD-AI in the low-rank learning setting when the parameter $\mathbf{W}$ is learned or fixed ($\mathbf{W} = \mathbf{I}_{n,p}[:, :p]$). When the parameter $\mathbf{W}$ is learned, we use the TensorFlow RiemOpt library [59] for optimization on Stiefel manifolds. The value of $p$ is set to 14. The results indicate that learning the parameter $\mathbf{W}$ leads to better accuracies than setting it to $\mathbf{I}_{n,p}[:, :p]$. This suggests an alternative method that first learns the parameter $\mathbf{W}$ and then fixes it during the training of the other parameters.

Finally, Tab. 9 reports the training times of our networks on HDM05 dataset. While the Log-Cholesky metric performs worst in terms of accuracy in our frame-

| Dataset | HDM05 | | FPHA | | NTU60 (X-Sub) | | NTU60 (X-View) | |
|---|---|---|---|---|---|---|---|---|
| | W | NW | W | NW | W | NW | W | NW |
| Accuracy (%) | 81.32 | 78.14 | 96.58 | 96.00 | 95.86 | 94.72 | 97.44 | 95.11 |
| Training time (min) | 1.09 | 0.94 | 0.64 | 0.54 | 9.19 | 7.81 | 8.56 | 7.30 |

**Table 7.** Accuracies and training times (minutes) per epoch of SPSD-AI in the full-rank learning setting with (W) and without (NW) the parameter $\mathbf{W}$.

| Dataset | HDM05 | | FPHA | | NTU60 (X-Sub) | | NTU60 (X-View) | |
|---|---|---|---|---|---|---|---|---|
| | WL | WF | WL | WF | WL | WF | WL | WF |
| Accuracy (%) | 75.53 | 74.19 | 92.06 | 90.94 | 91.05 | 89.78 | 92.28 | 90.15 |
| Training time (min) | 0.70 | 0.66 | 0.42 | 0.39 | 5.96 | 5.61 | 5.59 | 5.26 |

**Table 8.** Accuracies and training times (minutes) per epoch of SPSD-AI in the low-rank learning setting ($p = 14$) when the parameter $\mathbf{W}$ is learned (WL) or fixed (WF) beforehand.

work, it is more advantageous than the Affine-invariant and Log-Euclidean metrics in terms of computation time.

## References


51. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press (2007) 13
52. Gallier, J., Quaintance, J.: Differential Geometry and Lie Groups: A Computational Perspective. Springer International Publishing (2020) 4
53. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press (1991) 9
54. Li, J., Li, F., Todorovic, S.: Efficient Riemannian Optimization on the Stiefel Manifold via the Cayley Transform. In: ICLR (2020) 15
55. Liu, W., Lin, R., Liu, Z., Rehg, J.M., Paull, L., Xiong, L., Song, L., Weller, A.: Orthogonal Over-Parameterized Training. In: CVPR. pp. 7251–7260 (2021) 15
56. Lorenzi, M., Pennec, X.: Geodesics, Parallel Transport and One-Parameter Subgroups for Diffeomorphic Image Registration. IJCV **105**, 111–127 (2013) 4
57. Pennec, X., Ayache, N.: Uniform distribution, distance and expectation problems for geometric features processing. Journal of Mathematical Imaging and Vision **9**(1), 49–67 (1998) 1
58. Postnikov, M.M.: Geometry VI. Springer (2001) 4
59. Smirnov, O.: TensorFlow RiemOpt: A Library for Optimization on Riemannian Manifolds (2021) 16
60. Sra, S., Hosseini, R.: Conic Geometric Optimization on the Manifold of Positive Definite Matrices. SIAM Journal on Optimization **25**(1), 713–739 (2015) 2

| SPSD-AI | SPSD-LE | SPSD-LC |
| --- | --- | --- |
| 1.09 | 0.98 | 0.74 |

**Table 9.** Training times (minutes) per epoch of our networks on HDM05 dataset (full-rank learning setting with the parameter $\mathbf{W}$).