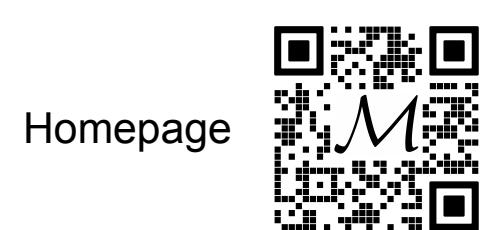
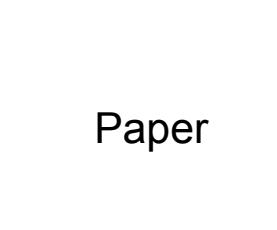


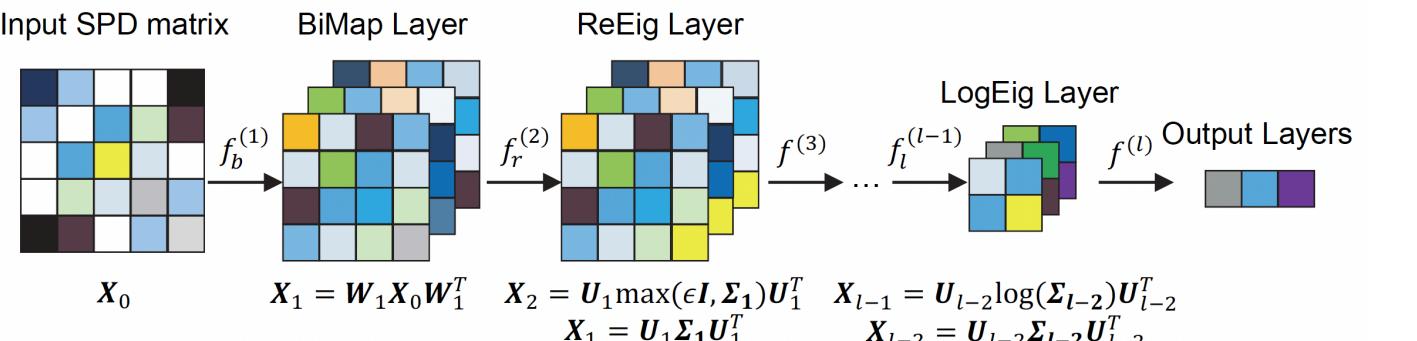
Riemannian Multinomial Logistics Regression for SPD Neural Networks

Ziheng Chen, Yue Song, Gaowen Liu, Ramana Rao Kompella, Xiaojun Wu, Nicu Sebe

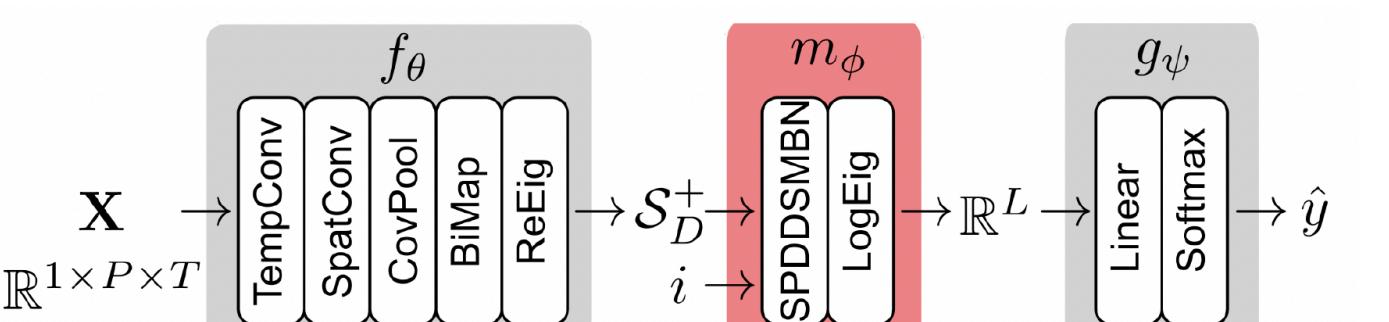


Motivation

Classification layers in typical SPD neural networks are non-intrinsic



Tangent Space [1-2]



Theoretical Contributions

- a general framework for building SPD MLRs under PEMs
- specific SPD MLRs under parameterized LCM and LEM
- An intrinsic theoretical explanation of the most popular LogEig classifier

Preliminaries

Pullback Euclidean Metrics (PEMs) [4]

Definition 2.1 (Pullback Metrics). Suppose \mathcal{M}, \mathcal{N} are smooth manifolds, g is a Riemannian metric on \mathcal{N} , and $f : \mathcal{M} \rightarrow \mathcal{N}$ is a diffeomorphism. Then f can induce a Riemannian metric on \mathcal{M} defined as

$$(f^*g)_p(V_1, V_2) = g_{f(p)}(f_{*,p}(V_1), f_{*,p}(V_2)), \quad (1)$$

where $p \in \mathcal{M}$, $f_{*,p}(\cdot)$ is the differential map of f at p , $V_i \in T_p \mathcal{M}$, and f^*g is the pullback metric by f from \mathcal{N} .

Theorem 2.2 (Pullback Euclidean Metrics (PEMs)). Let $S, S_1, S_2 \in \mathcal{S}_{++}^n$ and $V_1, V_2 \in T_S \mathcal{S}_{++}^n$, $\phi : \mathcal{S}_{++}^n \rightarrow \mathcal{S}^n$ is a diffeomorphism. We define the following operations,

$$S_1 \odot_\phi S_2 = \phi^{-1}(\phi(S_1) + \phi(S_2)), \quad (3)$$

$$g_S^{\phi}(V_1, V_2) = \langle \phi_{*,S}(V_1), \phi_{*,S}(V_2) \rangle, \quad (4)$$

where $\phi_{*,S} : T_S \mathcal{S}_{++}^n \rightarrow T_{\phi(S)} \mathcal{S}^n$ is the differential map of ϕ at S , and $\langle \cdot, \cdot \rangle$ is the standard Frobenius inner product. Then, we have the following conclusions: $\{\mathcal{S}_{++}^n, \odot_\phi\}$ is an Abelian Lie group, $\{\mathcal{S}_{++}^n, g^\phi\}$ is a Riemannian manifold, and g^ϕ is a bi-invariant metric, called Pullback Euclidean Metric (PEM). The associated geodesic distance is

$$d^\phi(S_1, S_2) = \|\phi(S_1) - \phi(S_2)\|_F, \quad (5)$$

where $\|\cdot\|_F$ is the norm induced by $\langle \cdot, \cdot \rangle$. The Riemannian operators are as follows

$$\text{Exp}_{S_1} V = \phi^{-1}(\phi(S_1) + \phi_{*,S_1} V), \quad (6)$$

$$\text{Log}_{S_1} S_2 = \phi_{*,\phi(S_1)}^{-1}(\phi(S_2) - \phi(S_1)), \quad (7)$$

$$\Gamma_{S_1 \rightarrow S_2}(V) = \phi_{*,\phi(S_2)}^{-1} \circ \phi_{*,S_1}(V), \quad (8)$$

where $V \in T_{S_1} \mathcal{S}_{++}^n$ is a tangent vector, Exp_{S_1} is the Riemannian exponential at S_1 , and ϕ_*^{-1} are the differential maps ϕ^{-1} .

SPD Geometries

Name	$g_P(V, W)$	$\text{Log}_P Q$	$\Gamma_{P \rightarrow Q}(V)$
(α, β) -LEM	$\langle \text{mlog}_{*,P}(V), \text{mlog}_{*,P}(W) \rangle^{(\alpha, \beta)}$	$(\text{mlog}_{*,P})^{-1}[\text{mlog}(Q) - \text{mlog}(P)]$	$(\text{mlog}_{*,Q})^{-1} \circ \text{mlog}_{*,P}(V)$
LCM	$\sum_{i>j} V_{ij} \tilde{W}_{ij} + \sum_{j=1}^n \tilde{V}_{jj} W_{jj} L_{jj}^{-2}$	$(\text{Chol}^{-1})_{*,L} [[K] - [L] + \mathbb{D}(L) \text{Diag}(\mathbb{D}(L)^{-1} \mathbb{D}(K))]$	$(\text{Chol}^{-1})_{*,K} [\tilde{V}] + \mathbb{D}(K) \mathbb{D}(L)^{-1} \mathbb{D}(\tilde{V})$

Table 2. Riemannian operators of (α, β) -LEM and LCM on SPD manifolds.

Methods

Exsiting Multinomial Logistics Regression (MLR)

$$\forall k \in \{1, \dots, C\}, p(y = k | x) \propto \exp((\langle a_k, x \rangle - b_k))$$

Euclidean MLR:

$$\begin{aligned} & \text{Reformulation} \\ & p(y = k | x) \propto \exp(\text{sign}(\langle a_k, x - p_k \rangle) \|a_k\| d(x, H_{a_k, p_k})) \\ & H_{a_k, p_k} = \{x \in \mathbb{R}^n : \langle a_k, x - p_k \rangle = 0\} \end{aligned}$$

Euclidean MLR has been generalized into hyperbolic [6] and SPD manifolds [5]. However, existing gyro SPD MLRs require:

- Gyro vector structures
- Solving formulation case by case

SPD MLR under PEMs

Definition 3.1 (SPD hyperplanes). Given $P \in \mathcal{S}_{++}^n$, $A \in T_P \mathcal{S}_{++}^n \setminus \{0\}$, we define the SPD hyperplane as

$$\tilde{H}_{A,P} = \{S \in \mathcal{S}_{++}^n : g_P(\text{Log}_P S, A) = \langle \text{Log}_P S, A \rangle_P = 0\}, \quad (12)$$

where P and A are referred to as shift and normal matrices, respectively.

Definition 3.2 (SPD MLR). SPD MLR is defined as

$$p(y = k | S) \propto \exp(\text{sign}(\langle A_k, \text{Log}_{P_k}(S) \rangle_{P_k}) \|A_k\|_{P_k} d(S, \tilde{H}_{A_k, P_k})), \quad (13)$$

where $P_k \in \mathcal{S}_{++}^n$, $A_k \in T_{P_k} \mathcal{S}_{++}^n \setminus \{0\}$, $\langle \cdot, \cdot \rangle_{P_k} = g_{P_k}$, and $\|\cdot\|_{P_k}$ is the norm on $T_{P_k} \mathcal{S}_{++}^n$ induced by g at P_k , and \tilde{H}_{A_k, P_k} is a margin hyperplane in \mathcal{S}_{++}^n as defined in Eq. (12). $d(S, \tilde{H}_{A_k, P_k})$ denotes the margin distance between S and SPD hyperplane \tilde{H}_{A_k, P_k} , which is formulated as:

$$d(S, \tilde{H}_{A_k, P_k}) = \inf_{Q \in \tilde{H}_{A_k, P_k}} d(S, Q), \quad (14)$$

where $d(S, Q)$ is the geodesic distance induced by g .

SPD hyperplanes are submanifolds, which generalize hyperplane in Euclidean spaces.

Proposition 3.3 (Submanifolds). The SPD hyperplane (as defined in Eq. (12)) under any geometrically complete Riemannian metric g is a regular submanifold of SPD manifolds.

Lemma 3.5. Given a PEM g , the margin distance defined in Eq. (14) has a closed-form solution:

$$d(S, \tilde{H}_{A_k, P_k}) = d(\phi(S), H_{\phi_{*,P_k}(A_k), \phi(P_k)}), \quad (15)$$

$$= \frac{|\langle \phi(S) - \phi(P_k), \phi_{*,P_k}(A_k) \rangle|}{\|A_k\|_{P_k}}, \quad (16)$$

where $|\cdot|$ is the absolute value.

SPD MLR: $p(y = k | S) \propto \exp(\langle A_k, \text{Log}_{P_k}(S) \rangle_{P_k}) = \exp(\langle \phi(S) - \phi(P_k), \phi_{*,P_k}(A_k) \rangle)$

Optmizing $A_k \in T_{P_k} \mathcal{S}_{++}^n \setminus \{0\}$

Lemma 3.6. Given a PEM, any parallel transportation is equivalent to the differential map of a left translation and vice versa.

Lemma 3.7. Given two fixed SPD matrices $Q_1, Q_2 \in \mathcal{S}_{++}^n$, we have the following equivalence for parallel transports under a PEM,

$$\begin{aligned} & \forall \tilde{A}_{1,k} \in T_{Q_1} \mathcal{S}_{++}^n, \exists! \tilde{A}_{2,k} \in T_{Q_2} \mathcal{S}_{++}^n, \\ & \text{s.t. } \Gamma_{Q_1 \rightarrow P_k}(\tilde{A}_{1,k}) = \Gamma_{Q_2 \rightarrow P_k}(\tilde{A}_{2,k}). \end{aligned} \quad (18)$$

Lems 3.6-3.7 indicates that we can either use parallel transportation or left translation.

General Formulation

Theorem 3.8 (SPD MLR under a PEM). Under any PEM, SPD MLR and SPD hyperplane is

$$p(y = k | S) \propto \exp(\langle \phi(S) - \phi(P_k), \phi_{*,P_k}(A_k) \rangle), \quad (19)$$

$$\tilde{H}_{A_k, P_k} = \{S \in \mathcal{S}_{++}^n : \langle \phi(S) - \phi(P_k), \phi_{*,P_k}(A_k) \rangle = 0\}, \quad (20)$$

where $\tilde{A}_k \in T_{P_k} \mathcal{S}_{++}^n \setminus \{0\} \cong \mathcal{S}^n \setminus \{0\}$ is a symmetric matrix, and $P_k \in \mathcal{S}_{++}^n$ is an SPD matrix.

SPD MLR under deformed LEM and LCM

Power deformation can interpolates between different metrics [7, 8]

$$\text{Power deformation: } \tilde{g} = \frac{1}{\theta^2} \text{Pow}_\theta^* g,$$

Corollary 4.1 (SPD MLRs under the deformed LEM and LCM). The SPD MLRs under (α, β) -LEM is

$$p(y = k | S) \propto \exp \left[(\text{mlog}(S) - \text{mlog}(P_k), \tilde{A}_k)^{(\alpha, \beta)} \right], \quad (21)$$

where $\tilde{A}_k \in T_{P_k} \mathcal{S}_{++}^n \cong \mathcal{S}^n$ and $P_k \in \mathcal{S}_{++}^n$. The SPD MLRs under (θ) -LCM is

$$p(y = k | S) \propto \exp \left[\frac{1}{\theta} ([\tilde{K}] - [\tilde{L}_k] + [\text{Dlog}(\mathbb{D}(\tilde{K})) - \text{Dlog}(\mathbb{D}(\tilde{L}_k))] , [\tilde{A}_k] + \frac{1}{2} \mathbb{D}(\tilde{A}_k)) \right], \quad (22)$$

where $\tilde{K} = \text{Chol}(S^\theta)$, $\tilde{L}_k = \text{Chol}(P_k^\theta)$, and $\mathbb{D}(\tilde{A}_k)$ denotes a diagonal matrix with diagonal elements of \tilde{A}_k .

SPD Hyperplanes

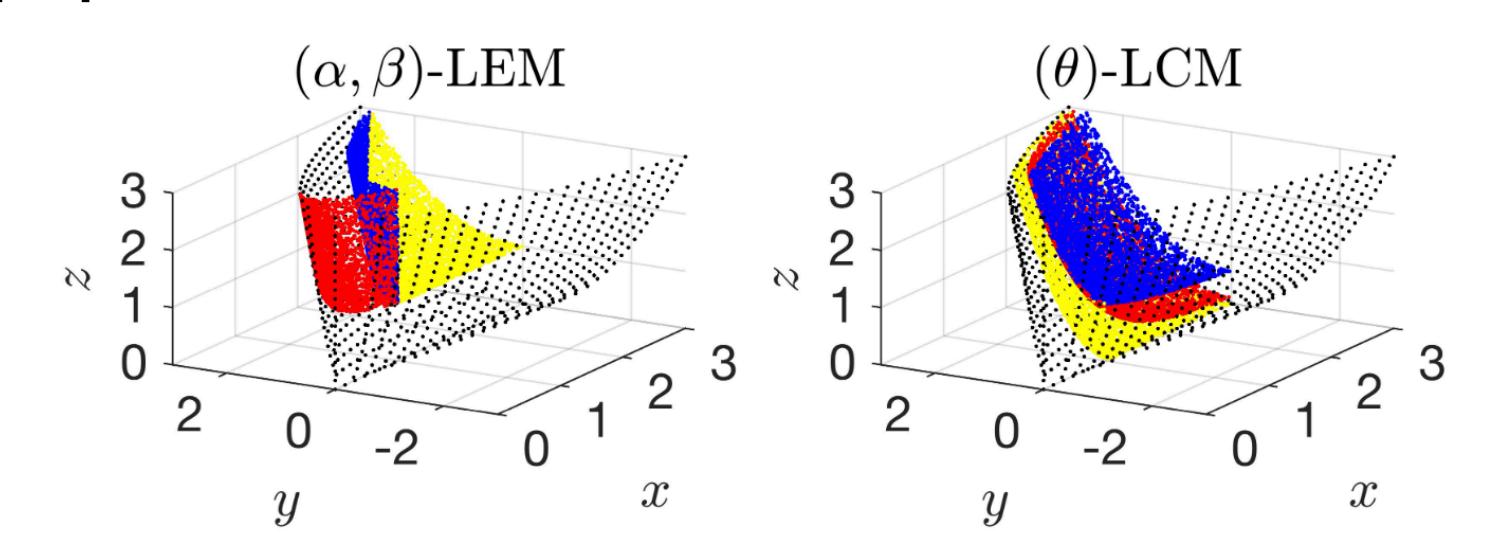


Figure 1. Conceptual illustration of SPD hyperplanes induced by (α, β) -LEM and (θ) -LCM. In each subfigure, the black dots are symmetric positive semi-definite (SPSD) matrices, denoting the boundary of \mathcal{S}_{++}^n , while the blue, red, and yellow dots denote three SPD hyperplanes.

An Intrinsic theoretical explanation for LogEig MLR

Proposition 5.1. Endowing SPD manifolds with the standard LEM, optimizing SPD parameter P_k in Eq. (21) by LEM-based RSGD and Euclidean parameter A_k by Euclidean SGD, the LEM-based SPD MLR is equivalent to a LogEig MLR with parameters in FC layer optimized by Euclidean SGD.

Experiments

SPDNet

Backbone	Classifier	[20,16,8]	[20,16,14,12,10,8]
SPDNet	LogEig MLR	92.88±1.05	93.47±0.45
	Gyro-AIM	94.53±0.95	94.32±0.94
	(1,0)-LEM	93.55±1.21	94.60±0.70
	(1,1)-LEM	95.64±0.83	95.87±0.58
(1,0)-LCM	93.49±1.25	93.93±0.98	
	(1,1)-LCM	62.04±1.05	62.11±2.11
	(0,5)-LCM	94.59±0.82	95.16±0.67
	(0,1)-LCM	94.56±0.73	65.79±0.63