Riemannian Multinomial Logistics Regression for SPD Neural Networks

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

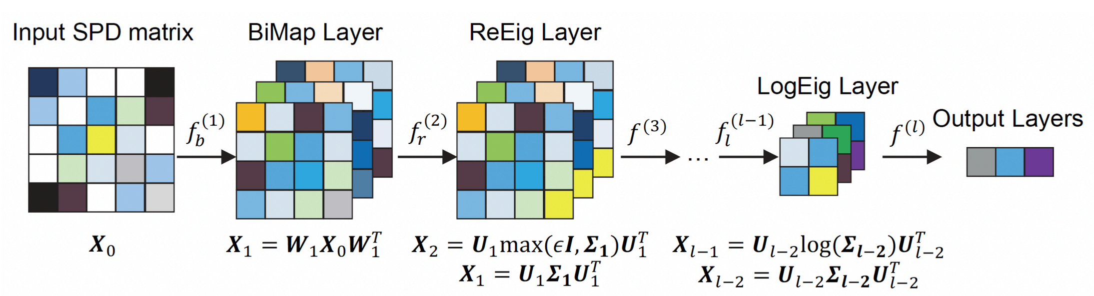
UNIVERSITÀ DI TRENTO · JIANGNAN UNIVERSITY · CISCO

CVPR JUNE 17-21, 2024 SEATTLE, WA

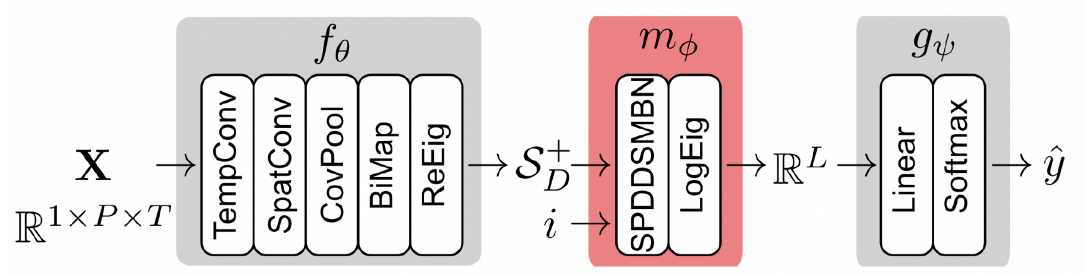
Code Paper Homepage

Motivation

Classification layers in typical SPD neural networks are non-intrinsic



Tangent Space [1-2]



Parameterization [3]

$$Y_t = \mathrm{FM}\left(\left\{M_{t-1}^{(\alpha)}\right\}, \left\{w^{(y,\alpha)}\right\}\right), \quad R_t = \mathrm{T}\left(Y_t, g^{(r)}\right)$$

$$T_t = \mathrm{FM}\left(\left\{R_t, X_t\right\}, w^{(t)}\right), \quad \Phi_t = \mathrm{T}\left(T_t, g^{(p)}\right)$$

$$\forall \alpha \in J, \quad M_t^{(\alpha)} = \mathrm{FM}\left(\left\{M_{t-1}^{(\alpha)}, \Phi_t\right\}, \alpha\right)$$

$$S_t = \mathrm{FM}\left(\left\{M_t^{(\alpha)}\right\}, \left\{w^{(s,\alpha)}\right\}\right), \quad O_t = \mathrm{Chol}\left(\mathrm{ReLU}\left(\mathrm{Chol}\left(\mathrm{T}\left(S_t, g^{(y)}\right)\right)\right)\right)$$

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

Prelimianries

Pullback Euclideam 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 $f : \mathcal{M} \to \mathcal{N}$ is a diffeomorphism. Then f can induce a Riemannian metric on $\mathcal{M}$ defined as

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

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 $V_1, V_2 \in T_S\mathcal{S}^n_{++}$, $\phi : \mathcal{S}^n_{++} \to \mathcal{S}^n$ is a diffeomorphism. We define the following operations,*

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

$$g^\phi_S(V_1, V_2) = \langle \phi_{*,S}(V_1), \phi_{*,S}(V_2)\rangle, \tag{4}$$

*where $\phi_{*,S} : T_S\mathcal{S}^n_{++} \to T_{\phi(S)}\mathcal{S}^n$ 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*

$$d^\phi(S_1, S_2) = \|\phi(S_1) - \phi(S_2)\|_{\mathrm{F}}, \tag{5}$$

where $\|\cdot\|_{\mathrm{F}}$ is the norm induced by $\langle\cdot,\cdot\rangle$. The Riemannian operators are as follows

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

$$\mathrm{Log}_{S_1} S_2 = \phi^{-1}_{*,\phi(S_1)}(\phi(S_2) - \phi(S_1)), \tag{7}$$

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

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	$g_P(V,W)$	$\mathrm{Log}_P Q$	$\Gamma_{P\to Q}(V)$
(α,β)-LEM	$\langle \mathrm{mlog}_{*,P}(V), \mathrm{mlog}_{*,P}(W)\rangle^{(\alpha,\beta)}$	$(\mathrm{mlog}_{*,P})^{-1}[\mathrm{mlog}(Q) - \mathrm{mlog}(P)]$	$(\mathrm{mlog}_{*,Q})^{-1} \circ \mathrm{mlog}_{*,P}(V)$
LCM	$\sum_{i>j} \tilde{V}_{ij}\tilde{W}_{ij} + \sum_{j=1}^n \tilde{V}_{jj}\tilde{W}_{jj}L_{jj}^{-2}$	$(\mathrm{Chol}^{-1})_{*,L}\left[\lfloor K\rfloor - \lfloor L\rfloor + \mathbb{D}(L)\,\mathrm{Dlog}(\mathbb{D}(L)^{-1}\mathbb{D}(K))\right]$	$(\mathrm{Chol}^{-1})_{*,K}\left[\lfloor\tilde{V}\rfloor + \mathbb{D}(K)\mathbb{D}(L)^{-1}\mathbb{D}(\tilde{V})\right]$

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

Methods

Exsiting Multinomial Logistics Regression (MLR)

$$\forall k \in \{1,\ldots,C\}, p(y=k \mid x) \propto \exp\left((\langle a_k, x\rangle - b_k)\right)$$

Euclidean MLR: Reformulation

$$p(y=k \mid x) \propto \exp(\mathrm{sign}(\langle a_k, x - p_k\rangle)\|a_k\| d(x, H_{a_k,p_k}))$$

$$H_{a_k,p_k} = \{x \in \mathbb{R}^n : \langle a_k, x - p_k\rangle = 0\}$$

Eucldiean 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

General Formulation

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

$$p(y=k \mid S) \propto \exp(\langle \phi(S) - \phi(P_k), \phi_{*,I}(\tilde{A}_k)\rangle), \tag{19}$$

$$\tilde{H}_{\tilde{A}_k,P_k} = \{S \in \mathcal{S}^n_{++} : \langle \phi(S) - \phi(P_k), \phi_{*,I}(\tilde{A}_k)\rangle = 0\}, \tag{20}$$

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

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

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

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

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

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

where $P_k \in \mathcal{S}^n_{++}$, $A_k \in T_{P_k}\mathcal{S}^n_{++}\backslash\{\mathbf{0}\}$, $\langle\cdot,\cdot\rangle_{P_k} = g_{P_k}$, 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:

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

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:*

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

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

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

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

Optmizing $A_k \in T_{P_k}\mathcal{S}^n_{++}\backslash\{\mathbf{0}\}$

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 transportations under a PEM,*

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

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

SPD MLR under deformed LEM and LCM

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

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

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

$$p(y=k \mid S) \propto \exp\left[\langle \mathrm{mlog}(S) - \mathrm{mlog}(P_k), \tilde{A}_k\rangle^{(\alpha,\beta)}\right], \tag{21}$$

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

$$p(y=k \mid S) \propto \exp\left[\frac{1}{\theta}\langle \lfloor\tilde{K}\rfloor - \lfloor\tilde{L}_k\rfloor + \left[\mathrm{Dlog}(\mathbb{D}(\tilde{K})) - \mathrm{Dlog}(\mathbb{D}(\tilde{L}_k))\right], \lfloor\tilde{A}_k\rfloor + \frac{1}{2}\mathbb{D}(\tilde{A}_k)\rangle\right], \tag{22}$$

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

SPD Hyperplanes

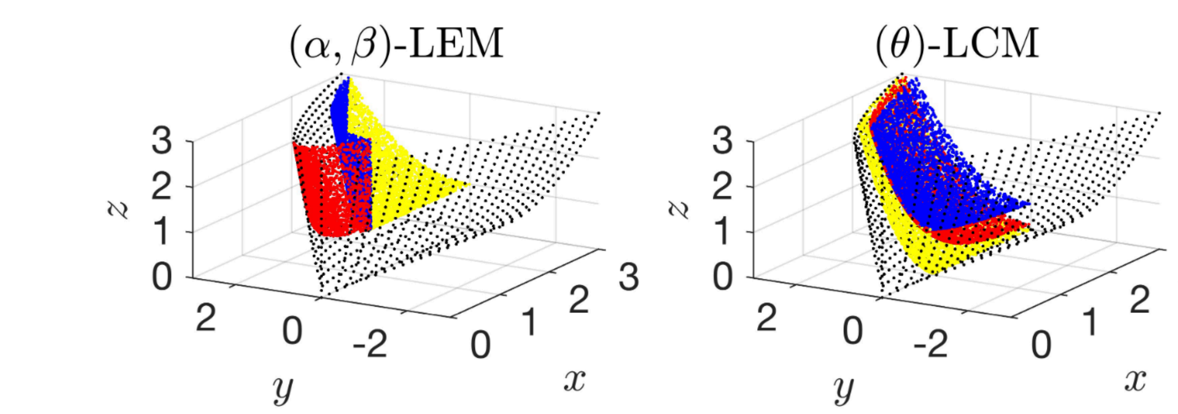


Figure 1. Conceptual illustration of SPD hyperplanes induced by (α,β)-LEM and (θ)-LCM. In each subfigure, the black dots are symmetric positive semi-definite (SPSD) matrices, denoting the boundary of $\mathcal{S}^2_{++}$, while the blue, red, and yellow dots denote three SPD hyperplanes.

An Intrinsic theoretical explanantion 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)-LCM	93.49±1.25	93.93±0.98
	(0.5)-LCM	**94.59±0.82**	**95.16±0.67**

Table 3. Results of SPDNet with different classifiers on the Radar dataset.

Backbone	Classifier	[93,30]	[93,70,30]	[93,70,50,30]
SPDNet	LogEig MLR	57.42±1.31	60.69±0.66	60.76±0.80
	Gyro-AIM	58.07±0.64	60.72±0.62	61.14±0.94
	(1,0)-LEM	57.02±0.75	61.34±0.62	60.78±0.86
	(1)-LCM	62.04±1.05	62.11±2.11	62.89±2.09
	(0.5)-LCM	**65.66±0.73**	**65.79±0.63**	**65.71±0.75**

Table 4. Results of SPDNet with different classifiers on the HDM05 dataset.

TSMNet

Backbone	Classifier	Inter-session	Inter-subject
SPDDSMBN	LogEig MLR	53.83±9.77	49.68±7.88
	Gyro-AIM	53.36±9.92	50.65±8.13
	(1,0)-LEM	53.16±9.73	51.41±7.98
	(1)-LCM	55.71±8.57	51.60±8.43
	(1.5)-LCM	**56.43±8.79**	**51.65±5.90**

Table 5. Results of SPDDSMBN with different classifiers on the Hinss2021 dataset under inter-subject and inter-session scenarios. The presented results are the ones of balanced accuracy under the leaving 5% out cross-validation scenario.

Training Efficiency

Methods	Radar	HDM05	Hinss2021 inter-session	Hinss2021 inter-subject
Baseline	1.36	1.95	0.18	8.31
MLR-Gyro-AIM	1.75	31.64	0.38	13.3
MLR-LEM	1.5	4.7	0.24	10.13
MLR-LCM	**1.35**	**3.29**	**0.18**	**8.35**

Table 6. Comparison of training efficiency (s/epoch) of SPDNet (SPDDSMBN) under different classifiers. The most efficient MLR is highlighted in **bold**.

References

[1] Huang, Zhiwu, and Luc Van Gool. "A riemannian network for spd matrix learning." AAAI, 2017.
[2] Kobler, Reinmar, et al. "SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG." Neurips, 2022.
[3] Chakraborty, Rudrasis, et al. "A statistical recurrent model on the manifold of symmetric positive definite matrices." Neurips, 2018.
[4] Chen, Ziheng, et al. "Adaptive Riemannian metrics on SPD manifolds." arXiv, 2023.
[5] Nguyen, Xuan Son, and Shuo Yang. "Building neural networks on matrix manifolds: A Gyrovector space approach." ICML, 2023.
[6] Ganea, Octavian, Gary Bécigneul, and Thomas Hofmann. "Hyperbolic neural networks." Neurips, 2018.
[7] Chen, Ziheng, et al. "A Lie Group Approach to Riemannian Batch Normalization", ICLR, 2024.
[8] Thanwerdas, Yann, and Xavier Pennec. "The geometry of mixed-Euclidean metrics on symmetric positive definite matrices." Differential Geometry and its Applications, 2022.