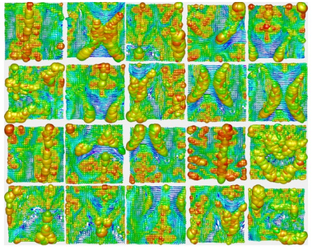
RMLR: Extending Multinomial Logistic Regression into General Geometries

Ziheng Chen[1], Yue Song[2], Rui Wang[3], Xiaojun Wu[3], Nicu Sebe[1]

1 University of Trento, Italy
2 Caltech, USA
3 Jiangnan University, China

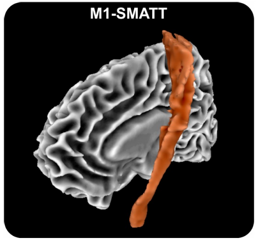
Applications of Riemannian Manifolds

Image generation



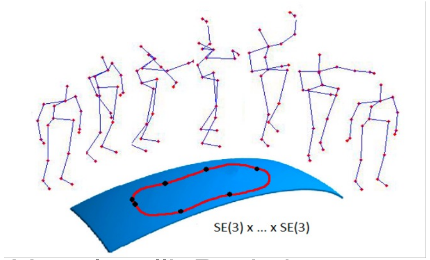
Huang et al., 2019

Medical Imaging



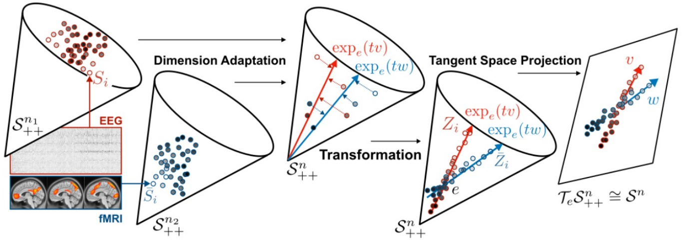
Chakraborty et al., 2020

Action Recognition



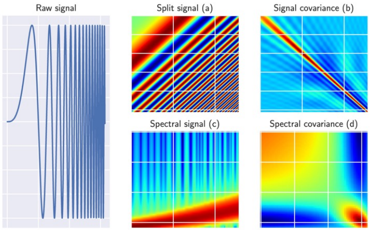
Vemulapalli, Raviteja 2014

Brain-Computer Interfaces



Ju et al., 2024

Radar Classification



Brooks et al., 2020

Huang, Zhiwu, Jiqing Wu, and Luc Van Gool. "Manifold-valued image generation with wasserstein generative adversarial nets." AAAI, 2019.
Chakraborty, Rudrasis, et al. "Manifoldnet: A deep neural network for manifold-valued data with applications." IEEE-TPAMI, 2020.
Vemulapalli, Raviteja, et al. Human action recognition by representing 3D skeletons as points in a lie group. CVPR. 2014.
Ju, Ce, et al. "Deep geodesic canonical correlation analysis for covariance-based neuroimaging data." ICLR, 2024.
Brooks, Daniel, et al. "Deep learning and information geometry for drone micro-Doppler radar classification." RadarConf, 2020.

Motivation

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

$$p(y = k \mid x) \propto \exp\left(\operatorname{sign}(\langle a_k, x - p_k \rangle) \|a_k\| d(x, H_{a_k, p_k})\right),$$
$$H_{a_k, p_k} = \{x \in \mathbb{R}^n : \langle a_k, x - p_k \rangle = 0\},$$

General geometries

Eqs. (2) and (3) can be naturally extended into manifolds $\mathcal{M}$ by Riemannian operators:

$$p(y = k \mid S) \propto \exp\left(\operatorname{sign}(\langle \tilde{A}_k, \operatorname{Log}_{P_k}(S) \rangle_{P_k}) \|\tilde{A}_k\|_{P_k} \tilde{d}(S, \tilde{H}_{\tilde{A}_k, P_k})\right), \tag{4}$$

$$\tilde{H}_{\tilde{A}_k, P_k} = \{S \in \mathcal{M} : g_{P_k}(\operatorname{Log}_{P_k} S, \tilde{A}_k) = 0\}, \tag{5}$$

where $P_k \in \mathcal{M}$, $\tilde{A}_k \in T_{P_k}\mathcal{M} \backslash \{\mathbf{0}\}$, g_{P_k} is the Riemannian metric at P_k, and Log_{P_k} is the Riemannian logarithm at P_k. The margin distance is defined as an infimum:

$$\tilde{d}(S, \tilde{H}_{\tilde{A}_k, P_k})) = \inf_{Q \in \tilde{H}_{\tilde{A}_k, P_k}} d(S, Q). \tag{6}$$

- The key is to solve the margin distance, which could be non-convex on general geometries

Reformulation by Riemannian Trigonometry

$$p(y=k \mid x) \propto \exp\left(\operatorname{sign}(\langle a_k, x-p_k\rangle)\|a_k\| d(x, H_{a_k,p_k})\right),$$
$$H_{a_k,p_k} = \{x \in \mathbb{R}^n : \langle a_k, x-p_k\rangle = 0\},$$

Reformulation

$$d(x, H_{a,p})) = \sin(\angle xpy^*) d(x,p), \quad \text{with } y^* = \underset{y \in H_{a,p}\setminus\{p\}}{\arg\max} (\cos \angle xpy). \tag{7}$$

Riemannian trigonometry

Definition 3.1 (Riemannian Margin Distance)**.** Let $\tilde{H}_{\tilde{A},P}$ be a Riemannian hyperplane defined in Eq. (5), and $S \in \mathcal{M}$. The Riemannian margin distance from S to $\tilde{H}_{\tilde{A},P}$ is defined as

$$d(S, \tilde{H}_{\tilde{A},P}) = \sin(\angle SPY^*) d(S,P), \tag{8}$$

where $d(S,P)$ is the geodesic distance, and $Y^* = \operatorname{argmax}(\cos \angle SPY)$ with $Y \in \tilde{H}_{\tilde{A},P}\setminus\{P\}$. The initial velocities of geodesics define $\cos \angle SPY$:

$$\cos \angle SPY = \frac{\langle \operatorname{Log}_P Y, \operatorname{Log}_P S\rangle_P}{\|\operatorname{Log}_P Y\|_P, \|\operatorname{Log}_P S\|_P}, \tag{9}$$

where $\langle \cdot, \cdot \rangle_P$ is the Riemannian metric at P, and $\|\cdot\|_P$ is the associated norm.

General Results

Riemannian Margin distance

Theorem 3.2. [↓] *The Riemannian margin distance defined in Def. 3.1 is given as*

$$d(S, \tilde{H}_{\tilde{A},P}) = \frac{|\langle \operatorname{Log}_P S, \tilde{A} \rangle_P|}{\|\tilde{A}\|_P}. \tag{10}$$

Putting the Eq. (10) into Eq. (4), we can a closed-form expression for Riemannian MLR.

Riemannian MLR

Theorem 3.3 (RMLR). [↓] *Given a Riemannian manifold $\{\mathcal{M}, g\}$, the Riemannian MLR induced by g is*

$$p(y = k \mid S \in \mathcal{M}) \propto \exp\left(\langle \operatorname{Log}_{P_k} S, \tilde{A}_k \rangle_{P_k}\right), \tag{11}$$

where $P_k \in \mathcal{M}$, $\tilde{A}_k \in T_{P_k}\mathcal{M} \backslash \{\mathbf{0}\}$, and Log *is the Riemannian logarithm.*

Optimization

$$\tilde{A}_k = \Gamma_{Q \rightarrow P_k} A_k, \tag{12}$$

$$\tilde{A}_k = L_{P_k \odot Q_{\odot}^{-1} *, Q} A_k, \tag{13}$$

where $Q \in \mathcal{M}$ is a fixed point, $A_k \in T_Q\mathcal{M} \backslash \{0\}$, Γ is the parallel transportation along geodesic connecting Q and P_k, and $L_{P_k \odot Q_{\odot}^{-1} *, Q}$ denotes the differential map at Q of left translation $L_{P_k \odot Q_{\odot}^{-1}}$

Generality

Table 1: Several MLRs on different geometries are special cases of our MLR.

MLR	Geometries	Requirements	Incorporated by Our MLR
Euclidean MLR (Eq. (1))	Euclidean geometry	N/A	✓(App. C)
Gyro SPD MLRs [50]	AIM, LEM & LCM on $\mathcal{S}^n_{++}$	Gyro structures	✓(Rem. 4.3)
Gyro SPSD MLRs [51]	SPSD product gyro spaces	Gyro structures	✓(App. D)
Flat SPD MLRs [16]	(α, β)-LEM & (θ)-LCM on $\mathcal{S}^n_{++}$	Pullback metrics from the Euclidean space	✓(Rem. 4.3)
Ours	General Geometries	Riemannian logarithm	N/A

Deformed SPD Geometries

Table 12: The associated Riemannian operators and properties of five basic metrics on SPD manifolds.

Metrics	$g_P(V,W)$	$\operatorname{Log}_P Q$	$\Gamma_{P\to Q}(V)$	Properties
(α,β)-LEM	$\langle \log_{*,P}(V), \log_{*,P}(W)\rangle^{(\alpha,\beta)}$	$(\log_{*,P})^{-1}\left[\log(Q)-\log(P)\right]$	$(\log_{*,Q})^{-1}\circ\log_{*,P}(V)$	$\mathrm{O}(n)$-Invariance, Geodesically Completeness
(α,β)-AIM	$\langle P^{-1}V, WP^{-1}\rangle^{(\alpha,\beta)}$	$P^{1/2}\log\left(P^{-1/2}QP^{-1/2}\right)P^{1/2}$	$(QP^{-1})^{1/2}V(P^{-1}Q)^{1/2}$	Lie Group Left-Invariance, $\mathrm{O}(n)$-Invariance, Geodesically Completeness
(α,β)-EM	$\langle V,W\rangle^{(\alpha,\beta)}$	$Q-P$	V	$\mathrm{O}(n)$-Invariance
LCM	$\sum_{i>j}\tilde{V}_{ij}\tilde{W}_{ij}+\sum_{j=1}^{n}\tilde{V}_{jj}\tilde{W}_{jj}L_{jj}^{-2}$	$(\mathrm{Chol}^{-1})_{*,L}\left[\lfloor K\rfloor-\lfloor L\rfloor+\mathbb{D}(L)\,\mathrm{Dlog}(\mathbb{D}(L)^{-1}\mathbb{D}(K))\right]$	$(\mathrm{Chol}^{-1})_{*,K}\left[\lfloor\tilde{V}\rfloor+\mathbb{D}(K)\mathbb{D}(L)^{-1}\mathbb{D}(\tilde{V})\right]$	Lie Group Bi-Invariance, Geodesically Completeness
BWM	$\frac{1}{2}\langle\mathcal{L}_P[V],W\rangle$	$(PQ)^{1/2}+(QP)^{1/2}-2P$	$U\left[\sqrt{\frac{\delta_i+\delta_j}{\sigma_i+\sigma_j}}\left[U^\top VU\right]_{ij}\right]U^\top$	$\mathrm{O}(n)$-Invariance

$$\tilde{g}_P\left(V,W\right)=\frac{1}{\theta^2}g_{P^\theta}\left((\phi_\theta)_{*,P}(V),(\phi_\theta)_{*,P}(W)\right),\forall P\in\mathcal{S}^n_{++},V,W\in T_P\mathcal{S}^n_{++}, \tag{14}$$

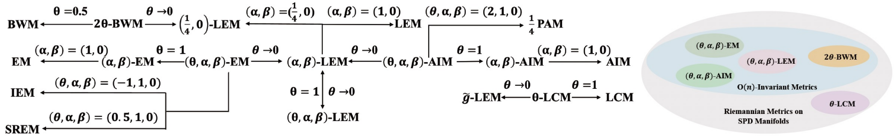


Figure 1: Illustration on the deformation (**left**) and Venn diagram (**right**) of metrics on SPD manifolds, where IEM, SREM, and $\frac{1}{4}$ PAM denotes Inverse Euclidean Metric, Square Root Euclidean Metric, and Polar Affine Metric scaled by $1/4$.

Table 2: Properties of deformed metrics on SPD manifolds ($\theta\neq 0$ and $\min(\alpha,\alpha+n\beta)>0$).

Name	Properties
(θ,α,β)-LEM	Bi-Invariance, $\mathrm{O}(n)$-Invariance, Geodesically Completeness
(θ,α,β)-AIM	Lie Group Left-Invariance, $\mathrm{O}(n)$-Invariance, Geodesically Completeness
(θ,α,β)-EM	$\mathrm{O}(n)$-Invariance
θ-LCM	Lie Group Bi-Invariance, Geodesically Completeness
2θ-BWM	$\mathrm{O}(n)$-Invariance

Manifestation

SPD MLR

Theorem 4.2 (SPD MLRs). [↓] *By abuse of notation, we omit the subscripts k of A_k and P_k. Given SPD feature S, the SPD MLRs, $p(y = k \mid S \in \mathcal{S}^n_{++})$, are proportional to*

$$(\alpha,\beta)\text{-}LEM: \exp\left[\langle \log(S) - \log(P), A\rangle^{(\alpha,\beta)}\right], \tag{16}$$

$$(\theta,\alpha,\beta)\text{-}AIM: \exp\left[\frac{1}{\theta}\langle \log(P^{-\frac{\theta}{2}} S^{\theta} P^{-\frac{\theta}{2}}), A\rangle^{(\alpha,\beta)}\right], \tag{17}$$

$$(\theta,\alpha,\beta)\text{-}EM: \exp\left[\frac{1}{\theta}\langle S^{\theta} - P^{\theta}, A\rangle^{(\alpha,\beta)}\right], \tag{18}$$

$$\theta\text{-}LCM: \exp\left[\frac{1}{\theta}\langle \lfloor \tilde{K} \rfloor - \lfloor \tilde{L} \rfloor + \left[\mathrm{Dlog}(\mathbb{D}(\tilde{K})) - \mathrm{Dlog}(\mathbb{D}(\tilde{L}))\right], \lfloor A \rfloor + \frac{1}{2}\mathbb{D}(A)\rangle\right], \tag{19}$$

$$2\theta\text{-}BWM: \exp\left[\frac{1}{4\theta}\langle (P^{2\theta} S^{2\theta})^{\frac{1}{2}} + (S^{2\theta} P^{2\theta})^{\frac{1}{2}} - 2P^{2\theta}, \mathcal{L}_{P^{2\theta}}(\bar{L} A \bar{L}^{\top})\rangle\right], \tag{20}$$

where $A \in T_I\mathcal{S}^n_{++} \backslash \{0\}$ is a symmetric matrix, $\log(\cdot)$ is the matrix logarithm, $\mathcal{L}_P(V)$ is the solution to the matrix linear system $\mathcal{L}_P[V]P + P\mathcal{L}_P[V] = V$, known as the Lyapunov operator, $\mathrm{Dlog}(\cdot)$ is the diagonal element-wise logarithm, $\lfloor \cdot \rfloor$ is the strictly lower part of a square matrix, and $\mathbb{D}(\cdot)$ is a diagonal matrix with diagonal elements of a square matrix. Besides, $\log_{*,P}$ is the differential maps at P, $\tilde{K} = \mathrm{Chol}(S^{\theta})$, $\tilde{L} = \mathrm{Chol}(P^{\theta})$, and $\bar{L} = \mathrm{Chol}(P^{2\theta})$.

Lie MLR

Theorem 5.2. [↓] *The Lie MLR on $\mathrm{SO}(n)$ is given as*

$$p(y = k \mid R \in \mathrm{SO}(n)) \propto \langle \log(P_k^{\top} S), A_k \rangle, \tag{22}$$

where $P_k \in \mathrm{SO}(n)$ and $A_k \in \mathfrak{so}(n)$.

Visualization

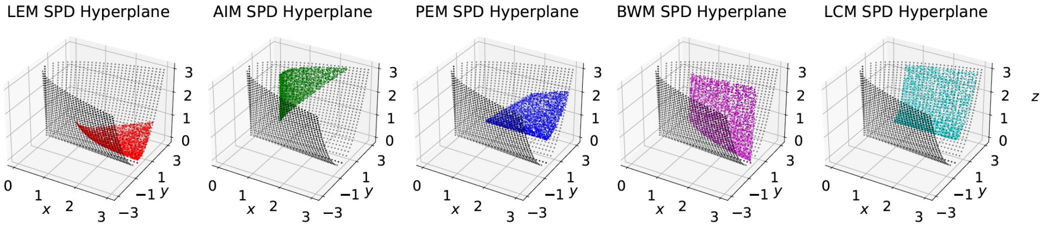


Figure 2: Conceptual illustration of SPD hyperplanes induced by five families of Riemannian metrics. The black dots denote the boundary of $\mathcal{S}^2_{++}$.

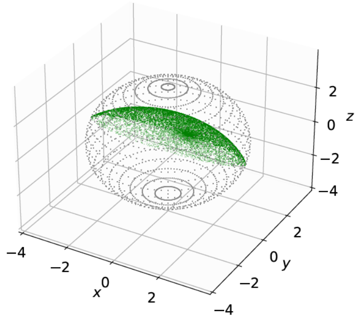


Figure 3: Conceptual illustration of a Lie hyperplane. Each pair of antipodal black dots corresponds to a rotation matrix with an Euler angle of π, while the green dots denote a Lie hyperplane.

Results

Riemannian Feedforward Networks

Table 3: Comparison of SPDNet with LogEig against SPD MLRs on the Radar dataset.

Architectures	LogEig MLR	(θ,α,β)-AIM	(θ,α,β)-EM		(α,β)-LEM		2θ-BWM		θ-LCM	
		(1,1,0)	(1,1,0)	(1,1,$1/8$)	(1,1,0)	(1,1,1)	(0.5)	(0.25)	(1)	(0.5)
2-Block	92.88±1.05	**94.53±0.95**	94.24±0.55	**94.93±0.60**	93.55±1.21	**95.64±0.83**	92.22±0.83	**94.99±0.47**	93.49±1.25	**94.59±0.82**
5-Block	93.47±0.45	**94.32±0.94**	**95.11±0.82**	95.01±0.84	94.60±0.70	**95.87±0.58**	93.69±0.66	**94.84±0.68**	93.93±0.98	**95.16±0.67**

Table 4: Comparison of SPDNet with LogEig against SPD MLRs on the HDM05 dataset.

Architectures	LogEig MLR	(θ,α,β)-AIM	(θ,α,β)-EM		(α,β)-LEM	2θ-BWM	θ-LCM	
		(1,1,0)	(1,1,0)	(0.5,1.0,$1/30$)	(1,1,0)	(0.5)	(1)	(0.5)
1-Block	57.42±1.31	58.07±0.64	66.32±0.63	**71.65±0.88**	56.97±0.61	**70.24±0.92**	63.84±1.31	**65.66±0.73**
2-Block	60.69±0.66	60.72±0.62	66.40±0.87	**70.56±0.39**	60.69±1.02	**70.46±0.71**	62.61±1.46	**65.79±0.63**
3-Block	60.76±0.80	61.14±0.94	66.70±1.26	**70.22±0.81**	60.28±0.91	**70.20±0.91**	62.33±2.15	**65.71±0.75**

Table 5: Inter-session experiments of TSMNet with different MLRs on the Hinss2021 dataset.

Classifiers	LogEig MLR	(θ,α,β)-AIM		(θ,α,β)-EM	(α,β)-LEM	2θ-BWM	θ-LCM	
		(1,1,0)	(0.5,1,0.05)	(1,1,0)	(1,1,0)	(0.5)	(1)	(1.5)
Balanced Acc.	53.83±9.77	53.36±9.92	**55.27±8.68**	**54.48±9.21**	53.51±10.02	**55.54±7.45**	55.71±8.57	**56.43±8.79**

Table 6: Inter-subject experiments of TSMNet with different MLRs on the Hinss2021 dataset.

Classifiers	LogEig MLR	(θ,α,β)-AIM		(θ,α,β)-EM		(α,β)-LEM	2θ-BWM		θ-LCM	
		(1,1,0)	(1.5,1,0)	(1,1,0)	(1.5,1,$1/20$)	(1,1,0)	(0.5)	(0.75)	(1)	(0.5)
Balanced Acc.	49.68±7.88	50.65±8.13	**51.15±7.83**	50.02±5.81	**51.38±5.77**	**51.41±7.98**	50.26±7.23	**51.67±8.73**	52.93±7.76	**54.14±8.36**

Results

Riemannian GCN

Table 8: Comparison of LogEig against SPD MLRs under the SPDGCN architecture.

Classifiers	Disease		Cora		Pubmed	
	Mean±STD	Max	Mean±STD	Max	Mean±STD	Max
LogEig MLR	90.55 ± 4.83	96.85	78.04 ± 1.27	79.6	70.99 ± 5.12	77.6
(θ, α, β)-AIM	94.84 ± 2.27	98.43	79.79 ± 1.44	81.6	77.83 ± 1.08	**80**
(θ, α, β)-EM	90.87 ± 5.14	98.03	79.05 ± 1.23	81	78.16 ± 2.41	79.5
(α, β)-LEM	**96.33 ± 2.19**	**98.82**	**79.89 ± 0.99**	81.8	**78.16 ± 2.41**	79.5
2θ-BWM	91.93 ± 3.64	96.85	73.46 ± 2.18	77.7	73.22 ± 4.06	78.1
θ-LCM	93.01 ± 2.14	98.43	77.59 ± 1.20	80.1	74.46 ± 5.81	78.9

RResNet

Table 7: Comparison of LogEig against SPD MLRs under the RResNet architecture.

Datasets	LogEigMLR	(θ, α, β)-AIM	(θ, α, β)-EM	(α, β)-LEM	2θ-BWM	θ-LCM
HDM05	58.17 ± 2.07	60.23 ± 1.26	**71.89 ± 0.60 (↑ 13.72)**	59.44 ± 0.87	69.85 ± 0.23	65.76 ± 0.96
NTU60	45.22 ± 1.23	48.94 ± 0.68	52.24 ± 1.25	46.99 ± 0.41	50.56 ± 0.59	**53.63 ± 0.95 (↑ 8.41)**

Riemannian VS. Tangent MLR

Table 9: Comparison of LogEig against SPD MLRs for direct classification.

Classifiers	Radar	HDM05	Hinss2021	
			Inster-session	Inster-subject
LogEig MLR	91.93 ± 1.30	48.43 ± 1.25	39.76 ± 7.60	44.66 ± 7.17
(θ, α, β)-AIM	95.21 ± 0.81	49.17 ± 1.08	41.14 ± 7.26	45.89 ± 6.52
(θ, α, β)-EM	92.25 ± 1.20	61.60 ± 0.69	**45.78 ± 8.51 (↑ 6.02)**	45.84 ± 4.75
(α, β)-LEM	95.09 ± 0.57	49.05 ± 0.91	40.88 ± 7.46	**46.02 ± 5.96 (↑ 1.36)**
2θ-BWM	94.89 ± 0.41	**66.77 ± 1.34 (↑ 18.34)**	44.84 ± 8.00	45.21 ± 7.44
θ-LCM	**95.67 ± 0.61 (↑ 3.74)**	58.66 ± 0.51	43.17 ± 6.21	45.10 ± 6.20

Results

Efficiency

Table 16: Training efficiency (s/epoch).

Methods	Radar	HDM05	Hinss2021	
			Inter-session	Inter-subject
Baseline	1.36	1.95	0.18	8.31
AIM-MLR	1.75	31.64	0.38	13.3
EM-MLR	1.34	3.91	0.19	8.23
LEM-MLR	1.5	4.7	0.24	10.13
BWM-MLR	1.75	33.14	0.38	13.84
LCM-MLR	1.35	3.29	0.18	8.35

LieNet

Table 10: Results of LogEig MLR against Lie MLR under the LieNet architecture.

Classifiers	G3D		HDM05	
	Mean±STD	Max	Mean±STD	Max
LogEig MLR	87.91±0.90	89.73	76.92±1.27	79.11
Lie MLR	**89.13±1.7**	**92.12**	**78.24±1.03**	**80.25**

Thanks

Homepage

LinkedIn