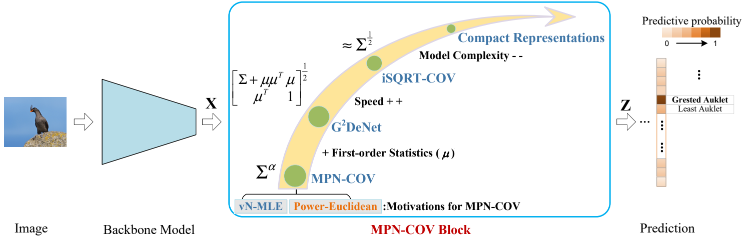
Understanding Matrix Function Normalizations in Covariance Pooling through the Lens of Riemannian Geometry

Ziheng Chen[1], Yue Song[2], Xiao-Jun Wu[3], Gaowen Liu[4] & Nicu Sebe[1]

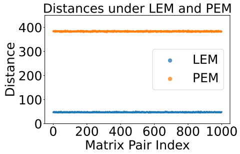
[1] University of Trento, [2] Caltech, [3] Jiangnan University, [4] Cisco Systems

ICLR 2025

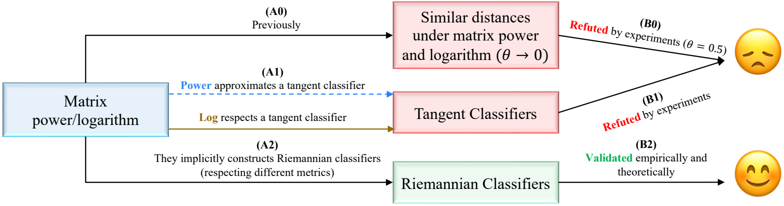
Motivations



Global Covariance Pooling (Wang et al., 2020)



LEM vs. PEM



GCP as Riemannian Multinomial Logistic Regression

Preliminaries

- Matrix logarithm and power:

$$\text{Log-EMLR: } \mathrm{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(\mathrm{mlog}(S)\right); A, b\right)\right), \tag{1}$$

$$\text{Pow-EMLR: } \mathrm{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(S^{\theta}\right); A, b\right)\right), \tag{2}$$

Table 1: Seven power-deformed metrics on SPD the manifold.

Name	Riemannian Metric $g_P(V, W)$	Riemannian Logarithm $\mathrm{Log}_P Q$	Deformation ($\theta \neq 0$)
(α, β)-LEM (Thanwerdas and Pennec, 2023)	$\langle \mathrm{mlog}_{*,P}(V), \mathrm{mlog}_{*,P}(W) \rangle^{(\alpha,\beta)}$	$(\mathrm{mlog}_{*,P})^{-1}\left[\mathrm{mlog}(Q) - \mathrm{mlog}(P)\right]$	$\frac{1}{\theta^2} \mathrm{Pow}_\theta^* g^{(\alpha,\beta)\text{-LE}}$
(α, β)-AIM (Thanwerdas and Pennec, 2023)	$\langle P^{-1}V, WP^{-1} \rangle^{(\alpha,\beta)}$	$P^{1/2}\, \mathrm{mlog}\left(P^{-1/2}QP^{-1/2}\right) P^{1/2}$	$\frac{1}{\theta^2} \mathrm{Pow}_\theta^* g^{(\alpha,\beta)\text{-AI}}$
(α, β)-EM (Thanwerdas and Pennec, 2023)	$\langle V, W \rangle^{(\alpha,\beta)}$	$Q - P$	$\frac{1}{\theta^2} \mathrm{Pow}_\theta^* g^{(\alpha,\beta)\text{-E}}$
(θ_1, θ_2)-EM (Thanwerdas and Pennec, 2022)	$\frac{1}{\theta_1\theta_2} \langle \mathrm{Pow}_{\theta_1 *,P}(V), \mathrm{Pow}_{\theta_2 *,P}(W) \rangle$	$(\mathrm{Pow}_{\theta *,P})^{-1}(Q^\theta - P^\theta)$, with $\theta = (\theta_1 + \theta_2)/2$	N/A
LCM (Lin, 2019)	$\sum_{i>j} \tilde{V}_{ij}\tilde{W}_{ij} + \sum_{j=1}^{n} \tilde{V}_{jj}\tilde{W}_{jj}L_{jj}^{-2}$	$(\mathrm{Chol}^{-1})_{*,L}\left[\lfloor K \rfloor - \lfloor L \rfloor + \mathbb{D}(L)\,\mathrm{Dlog}(\mathbb{D}(L)^{-1}\mathbb{D}(K))\right]$	$\frac{1}{\theta^2} \mathrm{Pow}_\theta^* g^{\mathrm{LC}}$
BWM (Bhatia et al., 2019)	$\frac{1}{2}\langle \mathcal{L}_P[V], W \rangle$	$(PQ)^{1/2} + (QP)^{1/2} - 2P$	$\frac{1}{4\theta^2} \mathrm{Pow}_{2\theta}^* g^{\mathrm{BW}}$
GBWM (Han et al., 2023)	$\frac{1}{2}\langle \mathcal{L}_{P,M}[V], W \rangle$	$M\left(M^{-1}PM^{-1}Q\right)^{1/2} + \left(QM^{-1}PM^{-1}\right)^{1/2} M - 2P$	$\frac{1}{4\theta^2} \mathrm{Pow}_{2\theta}^* g^{M\text{-BW}}$

Tangent Classifiers

- Matrix logarithm as $\mathrm{Log}_I \Rightarrow$ Considering all the logarithms.

Table 2: Log_I under seven families of SPD metrics.

Metric	$\mathrm{Log}_I P$	Metric	$\mathrm{Log}_I P$
(α, β)-LEM (θ, α, β)-AIM	$\mathrm{mlog}(P)$	(θ, α, β)-EM (θ_1, θ_2)-EM 2θ-BWM $(2\theta, P^{2\theta})$-BWM	$\frac{1}{\theta_0}(P^{\theta_0} - I)$
θ-LCM	$\frac{1}{\theta}\left[\lfloor \tilde{L} \rfloor + \lfloor \tilde{L} \rfloor^\top + 2\,\mathrm{Dlog}(\mathbb{D}(\tilde{L}))\right]$		

$$\boxed{\mathrm{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(S^\theta\right); A, b\right)\right)}$$

Table 3: Pow-TMLR vs. Pow-EMLR under the architecture of ResNet-18.

Method	ImageNet-1k		Cars	
	Top-1 Acc (%)	Top-5 Acc (%)	Top-1 Acc (%)	Top-5 Acc (%)
Pow-TMLR	71.62	89.73	51.14	74.29
Pow-EMLR	**73**	**90.91**	**80.43**	**94.15**

😭

From Tangent to Riemannian Classifiers

- Pow in GCP (Pow-EMLR): $\boxed{\mathrm{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(S^{\theta}\right);A,b\right)\right)}$
- SPD MLR (Chen et al., 2024b):

$$\text{LEM-based:}\quad \exp\left[\langle \log(S)-\log(P_k), A_k\rangle\right], \tag{3}$$

$$\text{PEM-based:}\quad \exp\left[\frac{1}{\theta}\langle S^{\theta}-P_k^{\theta}, A_k\rangle\right], \tag{4}$$

- Difference: the SPD parameters

Theories: Matrix Functions as RMLRs

$$\text{Pow-EMLR: } \boxed{\operatorname{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(S^{\theta}\right);A,b\right)\right)} \tag{5}$$

$$\text{ScalePow-EMLR : } \boxed{\operatorname{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(\frac{1}{\theta}S^{\theta}\right);A,b\right)\right)} \tag{6}$$

Theorem 1

Under PEM with $\theta > 0$, optimizing each SPD parameter P_k in Eq. (4) by PEM-based RSGD and Euclidean parameter A_k by Euclidean SGD, the PEM-based SPD MLR is equivalent to a Euclidean MLR in the co-domain of $\phi_\theta(\cdot) : \mathcal{S}^n_{++} \to \mathcal{S}^n_{++}$, defined as

$$\phi_\theta(S) = \frac{1}{\theta}S^{\theta}, \theta > 0, \forall S \in \mathcal{S}^n_{++}. \tag{7}$$

- Similar results w.r.t LEM and matrix logarithm can be found in (Chen et al., 2024a, Prop. 5.1.).

Experiments

Table 4: Results of iSQRT-COV on four datasets with different covariance matrix classifiers. The backbone network on ImageNet is ResNet-18, while the one on the other three FGVC datasets is ResNet-50. Power is set to be $1/2$ for Pow-TMLR, ScalePow-EMLR and Pow-EMLR.

Classifier	ImageNet-1k		Aircrafts		Birds		Cars	
	Top-1 Acc (%)	Top-5 Acc (%)	Top-1 Acc (%)	Top-5 Acc (%)	Top-1 Acc (%)	Top-5 Acc (%)	Top-1 Acc (%)	Top-5 Acc (%)
Cho-TMLR	N/A	N/A	78.97	91.81	48.07	72.59	51.06	74.33
Pow-TMLR	71.62	89.73	69.58	88.68	52.97	77.80	51.14	74.29
ScalePow-EMLR	72.43	90.44	71.05	89.86	63.48	84.69	80.31	94.07
Pow-EMLR	73	90.91	72.07	89.83	63.29	84.66	80.43	94.15

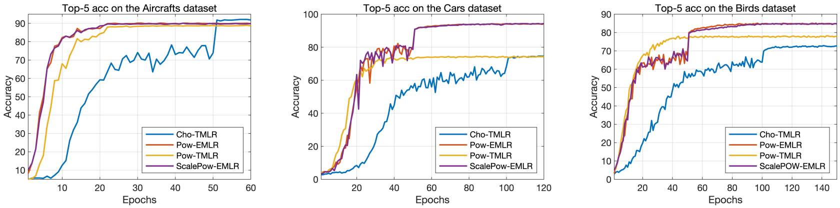


Figure 1: The validation top-5 accuracy on the three FGVC datasets for iSQRT-COV with different classifiers using the ResNet-50 backbone.

Bhatia, R., Jain, T., and Lim, Y. (2019). On the Bures-Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191.

Chen, Z., Song, Y., Liu, G., Kompella, R. R., Wu, X., and Sebe, N. (2024a). Riemannian multiclass logistics regression for SPD neural networks. In *CVPR*.

Chen, Z., Song, Y., Xiaojun, and Sebe, N. (2024b). RMLR: Extending multinomial logistic regression into general geometries. In *NeurIPS*.

Han, A., Mishra, B., Jawanpuria, P., and Gao, J. (2023). Learning with symmetric positive definite matrices via generalized Bures-Wasserstein geometry. In *International Conference on Geometric Science of Information*, pages 405–415. Springer.

Lin, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIMAX*.

Thanwerdas, Y. and Pennec, X. (2022). The geometry of mixed-Euclidean metrics on symmetric positive definite matrices. *Differential Geometry and its Applications*, 81:101867.

Thanwerdas, Y. and Pennec, X. (2023). O (n)-invariant Riemannian metrics on SPD matrices. *Linear Algebra and its Applications*, 661:163–201.

Wang, Q., Xie, J., Zuo, W., Zhang, L., and Li, P. (2020). Deep CNNs meet global covariance pooling: better representation and generalization. *IEEE TPAMI*.

Thank You

LinkedIn

Homepage