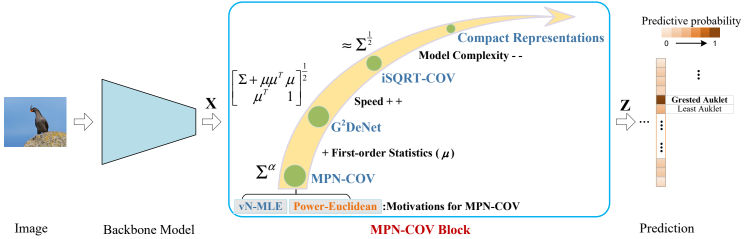
Understanding Matrix Function Normalizations in Covariance Pooling through the Lens of Riemannian Geometry

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

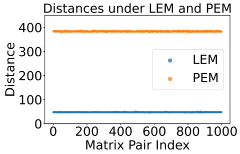
[1] University of Trento, [2] Caltech, [3] Jiangnan University, [4] Cisco Systems

ICLR 2025

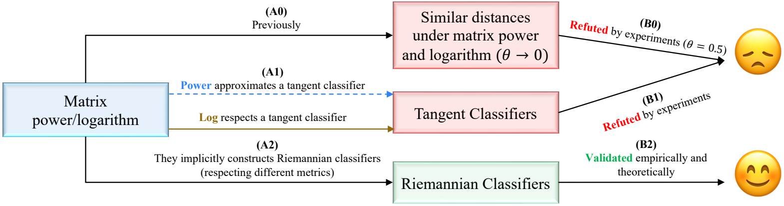
Motivations



Global Covariance Pooling (Wang et al., 2020)



LEM vs. PEM



GCP as Riemannian Multinomial Logistic Regression

Preliminaries

Table 1: Riemannian operators and deformed metrics of seven basic metrics on SPD manifolds.

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

GCP Revisiting

- Global Covariance Pooling (GCP):

$$X \xrightarrow{\mathrm{Cov}} \Sigma \xrightarrow{f_{\mathrm{M}}} \tilde{\Sigma} \xrightarrow{f_{\mathrm{vec}}} x \xrightarrow{f_{\mathrm{FC}}} \tilde{x} \xrightarrow{f_{\mathrm{EC}}} \hat{y}, \tag{1}$$

- Matrix logarithm and power:

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

$$\text{Pow-EMLR: } \mathrm{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(S^{\theta}\right); A, b\right)\right), \tag{3}$$

Riemannian Logarithm

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

Table 2: Log_I under seven families of SPD metrics.

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

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

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

Riemannian MLR

- Euclidean MLR:

$$\begin{aligned} \text{Softmax}(Ax) &\Rightarrow p(y=k \mid x) \propto \exp\left(\langle a_k, x\rangle - b_k\right), \forall k \in \{1, \ldots, C\} \\ &\Rightarrow p(y=k \mid x) \propto \exp(\operatorname{sign}(\langle a_k, x - p_k\rangle)\|a_k\| d(x, H_{a_k,p_k})), \end{aligned} \tag{4}$$

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

- SPD MLR (Chen et al., 2024b):

$$\text{LEM-based:} \quad \exp\left[\langle \log(S) - \log(P_k), A_k\rangle\right], \tag{5}$$

$$\text{PEM-based:} \quad \exp\left[\frac{1}{\theta}\langle S^\theta - P_k^\theta, A_k\rangle\right], \tag{6}$$

- Difference: the SPD parameters

Theories: Matrix Function as GCP

$$\text{Pow-EMLR: } \boxed{\mathrm{softmax}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(S^{\theta}\right); A, b\right)\right)} \tag{7}$$

Theorem

Under PEM with $\theta > 0$, optimizing each SPD parameter P_k in Eq. (6) 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 $\phi_\theta(\cdot) : \mathcal{S}^n_{++} \rightarrow \mathcal{S}^n_{++}$, defined as

$$\phi_\theta(S) = \frac{1}{\theta} S^{\theta}, \theta > 0, \forall S \in \mathcal{S}^n_{++}. \tag{8}$$

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

Summary

Table 4: Intrinsic explanations of some classifiers for GCP. For Cho-TMLR, $\tilde{L} = \mathrm{Chol}(S^{\theta})$. For Pow-TMLR, $\theta_0 = \frac{\theta_1+\theta_2}{2}$ for (θ_1, θ_2)-EM, $\theta_0 = \theta$ for (θ, α, β)-EM, $\theta_0 = 2\theta$ for 2θ-BWM and $(2\theta, \phi_{2\theta}(S))$-BWM. Here, $f_{\mathrm{s}}(\cdot)$ denotes the softmax, $\mathcal{F}(\cdot)$ denotes the FC layer, and $\tilde{V} = \frac{1}{\theta}\left[\lfloor \tilde{L} \rfloor + \lfloor \tilde{L} \rfloor^{\top} + 2\,\mathrm{Dlog}(\mathbb{D}(\tilde{L}))\right]$ with $S^{\theta} = \tilde{L}\tilde{L}^{\top}$ as the Cholesky decomposition.

	Log-EMLR	Pow-EMLR	ScalePow-EMLR	Pow-TMLR	Cho-TMLR
Expression	$f_{\mathrm{s}}(\mathcal{F}(f_{\mathrm{vec}}(\mathrm{mlog}(S))))$	$f_{\mathrm{s}}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(S^{\theta}\right)\right)\right)$ $(\theta > 0)$	$f_{\mathrm{s}}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(\frac{1}{\theta}S^{\theta}\right)\right)\right)$ $(\theta > 0)$	$f_{\mathrm{s}}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(\frac{1}{\theta_0}(S^{\theta_0} - I)\right)\right)\right)$	$f_{\mathrm{s}}\left(\mathcal{F}\left(f_{\mathrm{vec}}\left(\tilde{V}\right)\right)\right)$
Explanation	SPD MLR	SPD MLR	SPD MLR	Tangent Classifier	Tangent Classifier
Metrics	LEM	$(\theta, 1, 0)$-EM	$(\theta, 1, 0)$-EM	(θ, α, β)-EM, (θ_1, θ_2)-EM, 2θ-BWM, $(2\theta, \phi_{2\theta}(S))$-BWM	θ-LCM
Used in GCP	✓(Eq. (2))	✓ ($\theta = 0.5$ in Eq. (3))	✗	✗	✗
Reference	(Chen et al., 2024a, Prop. 5.1)	Thm. 1	Thm. 1	Tab. 2	Tab. 2

Experiments

Table 5: 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

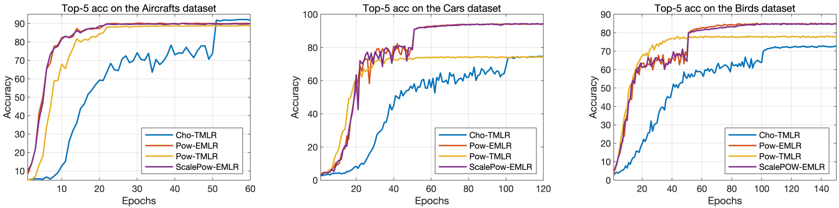


Figure 1: The validation top-5 accuracy on the three FGVC datasets for iSQRT-COV with different classifiers using the ResNet-50 backbone.

Ablations

Table 6: Ablations of Pow-EMLR, ScalePow-EMLR, and Pow-TMLR under different settings.

(a) Ablations of different powers.

Classifier	Aircrafts Top-1 Acc (%)	Aircrafts Top-5 Acc (%)	Cars Top-1 Acc (%)	Cars Top-5 Acc (%)
Pow-TMLR-0.25	65.41	86.71	41.47	66.66
ScalePow-EMLR-0.25	**72.76**	**90.31**	61.78	84.04
Pow-EMLR-0.25	71.47	90.04	**62.88**	84.14
Pow-TMLR-0.5	67.9	88.75	55.01	77.95
ScalePow-EMLR-0.5	74.29	91.12	62.42	84.82
Pow-EMLR-0.5	74.17	91.21	62.83	84.85
Pow-TMLR-0.7	65.92	87.49	50.68	74.12
ScalePow-EMLR-0.7	**74.26**	**91.15**	**64.22**	**83.67**
Pow-EMLR-0.7	74.17	90.49	61.41	82.39

(b) Results under the AlexNet.

Dataset	Result	Pow-TMLR	Pow-EMLR
Aircrafts	Top-1 Acc (%)	38.01	**65.02**
	Top-5 Acc (%)	74.4	**87.79**
Cars	Top-1 Acc (%)	28.57	**59.13**
	Top-5 Acc (%)	59.51	**82.04**

Table 7: Comparison of Pow-EMLR, ScalePow-EMLR and Pow-TMLR under transformer (SoT-7) backbone (Xie et al., 2021) on the ImageNet-1k dataset.

Classifier	Top-1 Acc (%)	Top-5 Acc (%)
Pow-TMLR	75.79	92.91
ScalePow-EMLR	**76.14**	**93.18**
Pow-EMLR	76.11	93.05

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*.

Xie, J., Zeng, R., Wang, Q., Zhou, Z., and Li, P. (2021). Sot: Delving deeper into classification head for transformer. *arXiv preprint arXiv:2104.10935*.

Thank You