Fast and Stable Riemannian Metrics on SPD Manifolds via Cholesky Product Geometry

Ziheng Chen, Yue Song, Xiaojun Wu, Nicu Sebe

UNIVERSITÀ DI TRENTO

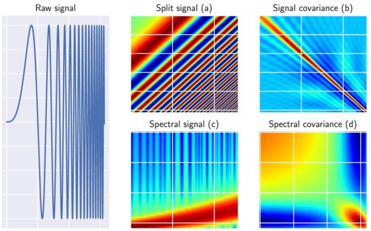
清華大學 Tsinghua University

上海期智研究院 SHANGHAI QI ZHI INSTITUTE

江南大学 JIANGNAN UNIVERSITY

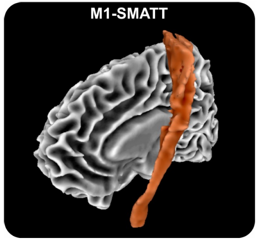
Motivation

Radar Classification



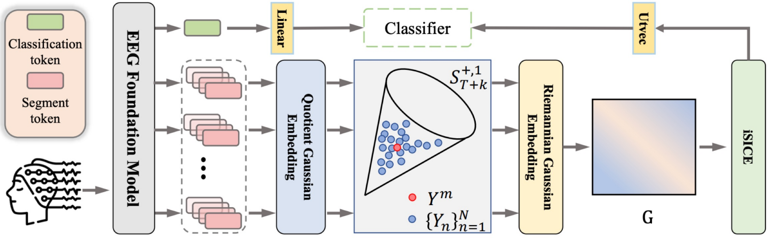
Brooks et al., 2020

Medical



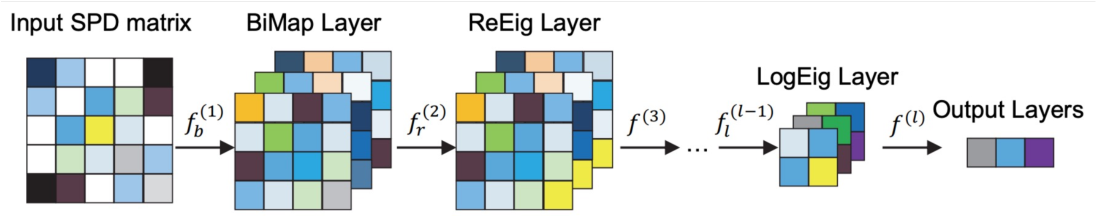
Chakraborty et al., 2020

Brain-Computer Interfaces



Chen et al., 2026

SPD Networks



$$X_0 \qquad X_1 = W_1 X_0 W_1^T \qquad X_2 = U_1 \max(\epsilon I, \Sigma_1) U_1^T,\; X_1 = U_1 \Sigma_1 U_1^T \qquad X_{l-1} = U_{l-2} \log(\Sigma_{l-2}) U_{l-2}^T,\; X_{l-2} = U_{l-2}\Sigma_{l-2}U_{l-2}^T$$

Huang et al., 2017

- Key point: **fast**, **simple**, and **stable** Riemannian metrics

PRODUCT GEOMETRIES

Current Geometries

$$g_P^{\mathrm{AI}}(V,W) = \langle P^{-1}V, WP^{-1}\rangle, \qquad g_P^{\mathrm{LE}}(V,W) = \langle \mathrm{mlog}_{*,P}(V), \mathrm{mlog}_{*,P}(W)\rangle,$$
$$g_P^{\theta\text{-E}}(V,W) = \frac{1}{\theta^2}\langle \mathrm{Pow}_{\theta*,P}\, V, \mathrm{Pow}_{\theta*,P}\, W\rangle, \quad g_P^{\mathrm{LC}}(V,W) = \langle \lfloor X \rfloor, \lfloor Y \rfloor\rangle + \langle \mathbb{L}^{-1}\mathbb{X}, \mathbb{L}^{-1}\mathbb{Y}\rangle, \tag{1}$$

$$g_P^{M\text{-BW}}(V,W) = \frac{1}{2}\langle \mathcal{L}_{P,M}(V), W\rangle. \tag{2}$$

- The one based on the **Cholesky** is fast, simple, and stable.

Cholesky geometries. The Euclidean space of $n \times n$ lower triangular matrices is denoted $\mathcal{L}^n$. Its open subset, whose diagonal elements are all positive, is denoted by $\mathcal{L}^n_{++}$. The Cholesky space $\mathcal{L}^n_{++}$ forms a submanifold of $\mathcal{L}^n$ (Lin, 2019). For a Cholesky matrix $L \in \mathcal{L}^n_{++}$ and tangent vectors $X, Y \in T_L\mathcal{L}^n_{++}$, the Riemannian metric on the Cholesky manifold, referred to as the diagonal log metric, is

$$g_L^{\mathrm{DL}}(X,Y) = \langle \lfloor X \rfloor, \lfloor Y \rfloor\rangle + \langle \mathbb{L}^{-1}\mathbb{X}, \mathbb{L}^{-1}\mathbb{Y}\rangle. \tag{3}$$

LCM is the pullback metric of g^{DL} by the Cholesky decomposition. As shown by Chen et al. (2024d, Thm. III.1.), the diagonal log metric is the pullback metric, by the diagonal log map, of the Euclidean metric over $\mathcal{L}^n$, which rationalizes our nomenclature.

Product Structure

$$\{\mathcal{L}^n_{++}, g^{\mathrm{DL}}\} = \{\mathcal{SL}^n, g^{\mathrm{E}}\} \times \overbrace{\{\mathbb{R}_{++}, g^{\mathbb{R}_{++}}\} \times \cdots \times \{\mathbb{R}_{++}, g^{\mathbb{R}_{++}}\}}^{n}. \tag{5}$$

- **Euclidean** off-diagonals + **Non-Euclidean** diagonals

OVERVIEW

$$\{\mathcal{L}^n_{++}, g^{\mathrm{DL}}\} = \{\mathcal{SL}^n, g^{\mathrm{E}}\} \times \overbrace{\{\mathbb{R}_{++}, g^{\mathbb{R}_{++}}\} \times \cdots \times \{\mathbb{R}_{++}, g^{\mathbb{R}_{++}}\}}^{n}. \tag{5}$$

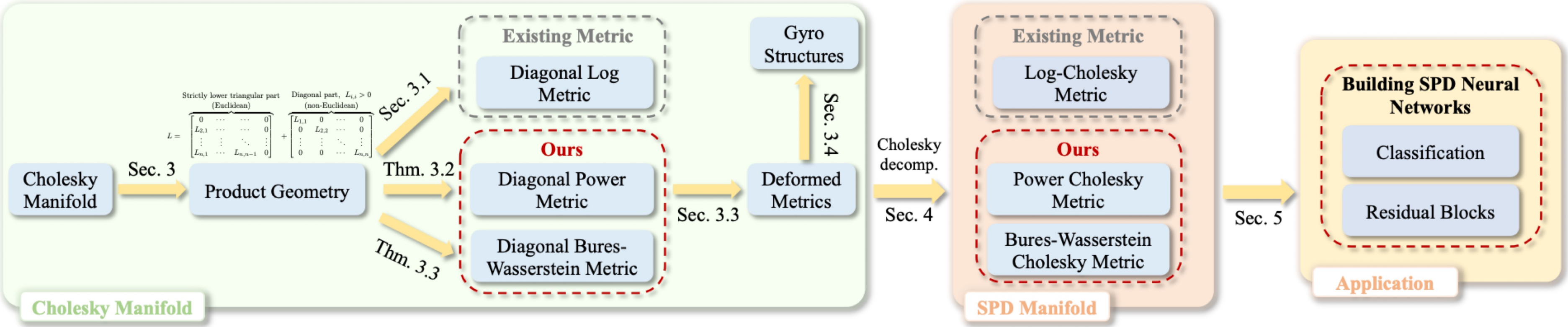


Figure 1: Overview of our theoretical framework. Here, L is a Cholesky matrix.

PCM AND BWCM

Cholesky Geometries

Operators	Diagonal Log Metric	θ-DPM	$(\theta, \mathbb{M})$-DBWM
$g_L(X,Y)$	$\lfloor X \rfloor + \lfloor Y \rfloor + \langle \mathbb{L}^{-1}\mathbb{X}, \mathbb{L}^{-1}\mathbb{Y} \rangle$	$\langle \lfloor X \rfloor, \lfloor Y \rfloor \rangle + \langle \mathbb{L}^{\theta-1}\mathbb{X}, \mathbb{L}^{\theta-1}\mathbb{Y} \rangle$	$\langle \lfloor X \rfloor, \lfloor Y \rfloor \rangle + \frac{1}{4} \langle \mathbb{L}^{\theta-2}\mathbb{X}, \mathbb{M}^{-1}\mathbb{Y} \rangle$
$\gamma_{(L,X)}(t)$	$\lfloor L \rfloor + t\lfloor X \rfloor + \mathbb{L}\exp(t\mathbb{L}^{-1}\mathbb{X})$	$\lfloor L \rfloor + t\lfloor X \rfloor + \mathbb{L}\left(I + t\theta\mathbb{L}^{-1}\mathbb{X}\right)^{\frac{1}{\theta}}$	$\lfloor L \rfloor + t\lfloor X \rfloor + \mathbb{L}\left(I + t\frac{\theta}{2}\mathbb{L}^{-1}\mathbb{X}\right)^{\frac{2}{\theta}}$
$\operatorname{Log}_L(K)$	$\lfloor K \rfloor - \lfloor L \rfloor + \mathbb{L}\log(\mathbb{L}^{-1}\mathbb{K})$	$\lfloor K \rfloor - \lfloor L \rfloor + \frac{1}{\theta}\mathbb{L}\left[\left(\mathbb{L}^{-1}\mathbb{K}\right)^{\theta} - I\right]$	$\lfloor K \rfloor - \lfloor L \rfloor + \frac{2}{\theta}\mathbb{L}\left[\left(\mathbb{L}^{-1}\mathbb{K}\right)^{\frac{\theta}{2}} - I\right]$
$\mathrm{PT}_{L\to K}(X)$	$\lfloor X \rfloor + (\mathbb{L}^{-1}\mathbb{K})\mathbb{X}$	$\lfloor X \rfloor + \left(\mathbb{L}^{-1}\mathbb{K}\right)^{1-\theta}\mathbb{X}$	$\lfloor X \rfloor + \left(\mathbb{L}^{-1}\mathbb{K}\right)^{1-\frac{\theta}{2}}\mathbb{X}$
$\mathrm{d}^2(L,K)$	$\Vert \lfloor K \rfloor - \lfloor L \rfloor \Vert_{\mathrm{F}}^2 + \Vert \log(\mathbb{K}) - \log(\mathbb{L}) \Vert_{\mathrm{F}}^2$	$\Vert \lfloor K \rfloor - \lfloor L \rfloor \Vert_{\mathrm{F}}^2 + \frac{1}{\theta^2}\left\Vert \mathbb{K}^{\theta} - \mathbb{L}^{\theta} \right\Vert_{\mathrm{F}}^2$	$\Vert \lfloor K \rfloor - \lfloor L \rfloor \Vert_{\mathrm{F}}^2 + \frac{1}{\theta^2}\left\Vert \mathbb{M}^{-\frac{1}{2}}\left(\mathbb{K}^{\frac{\theta}{2}} - \mathbb{L}^{\frac{\theta}{2}}\right) \right\Vert_{\mathrm{F}}^2$
$\mathrm{WFM}(\{w_i\},\{L_i\})$	$\sum_i w_i \lfloor L_i \rfloor + \exp\left(\sum_i w_i \log(\mathbb{L}_i)\right)$	$\sum_i w_i \lfloor L_i \rfloor + \left(\sum_i w_i \mathbb{L}_i^{\theta}\right)^{\frac{1}{\theta}}$	$\sum_i w_i \lfloor L_i \rfloor + \left(\sum_i w_i \mathbb{L}_i^{\frac{\theta}{2}}\right)^{\frac{2}{\theta}}$
$L \oplus K$	$\lfloor L \rfloor + \lfloor K \rfloor + \mathbb{L}\mathbb{K}$	$\lfloor L \rfloor + \lfloor K \rfloor + \left(\mathbb{L}^{\theta} + \mathbb{K}^{\theta} - I\right)^{\frac{1}{\theta}}$	$\lfloor L \rfloor + \lfloor K \rfloor + \left(\mathbb{L}^{\frac{\theta}{2}} + \mathbb{K}^{\frac{\theta}{2}} - I\right)^{\frac{2}{\theta}}$
$t \odot L$	$t\lfloor L \rfloor + \mathbb{L}^t$	$t\lfloor L \rfloor + \left(t\mathbb{L}^{\theta} + (1-t)I\right)^{\frac{1}{\theta}}$	$t\lfloor L \rfloor + \left(t\mathbb{L}^{\frac{\theta}{2}} + (1-t)I\right)^{\frac{2}{\theta}}$

SPD Geometries

$$
\begin{aligned}
&\gamma^{\mathcal{S}}_{(P,V)}(t) = \mathrm{Chol}^{-1}\left(\gamma^{\mathcal{C}}_{(L,\widetilde{V})}(t)\right), && \operatorname{Log}^{\mathcal{S}}_P(Q) = \left(\mathrm{Chol}_{*,P}\right)^{-1}\left(\operatorname{Log}^{\mathcal{C}}_L(K)\right), \\
&\operatorname{Exp}^{\mathcal{S}}_P(V) = \mathrm{Chol}^{-1}\left(\operatorname{Exp}^{\mathcal{C}}_L\left(\widetilde{V}\right)\right), && \mathrm{PT}^{\mathcal{S}}_{P\to Q}(V) = \left(\mathrm{Chol}_{*,Q}\right)^{-1}\left(\mathrm{PT}^{\mathcal{C}}_{L\to K}\left(\widetilde{V}\right)\right), \\
&\mathrm{d}^{\mathcal{S}}(P,Q) = \mathrm{d}^{\mathcal{C}}(L,K), && \mathrm{WFM}^{\mathcal{S}}(\{P_i\},\{w_i\}) = \mathrm{Chol}^{-1}\left(\mathrm{WFM}^{\mathcal{C}}\left(\{L_i\},\{w_i\}\right)\right), \\
&P \oplus^{\mathcal{S}} Q = \mathrm{Chol}^{-1}(L \widetilde{\oplus}^{\mathcal{C}} K), && t \odot^{\mathcal{S}} P = \mathrm{Chol}^{-1}(t \widetilde{\odot}^{\mathcal{C}} L),
\end{aligned}
\tag{11}
$$

Applications

Classifier

$$p(y=k \mid X) \propto \exp\left(\operatorname{sign}(\langle A_k, \operatorname{Log}_{P_k}(X)\rangle_{P_k})\|A_k\|_{P_k} \operatorname{d}(X, H_{A_k,P_k})\right), \quad \forall X \in \mathcal{M}$$

$$\forall k \in \{1,\ldots,C\}, \quad p(y=k \mid x) \propto \exp\left(\operatorname{sign}(\langle a_k, x-p_k\rangle)\|a_k\| \operatorname{d}(x, H_{a_k,p_k})\right)$$

Theorem 5.1. [↓] *Given an input SPD matrix $S \in \mathcal{S}^n_{++}$, the C-class SPD MLRs under θ-PCM and $(\theta, \mathbb{M})$-BWCM are*

$$\theta\text{-PCM}: p(y=k \mid S \in \mathcal{S}^n_{++}) \propto \exp\left[\langle \lfloor K \rfloor - \lfloor L_k \rfloor, \lfloor A_k \rfloor \rangle + \frac{1}{2\theta}\langle \mathbb{K}^\theta - \mathbb{L}_k^\theta, \mathbb{A}_k\rangle\right], \tag{14}$$

$$(\theta, \mathbb{M})\text{-BWCM}: p(y=k \mid S \in \mathcal{S}^n_{++}) \propto \exp\left[\langle \lfloor K \rfloor - \lfloor L_k \rfloor, \lfloor A_k \rfloor \rangle + \frac{1}{4\theta}\langle \mathbb{K}^{\frac{\theta}{2}} - \mathbb{L}_k^{\frac{\theta}{2}}, \mathbb{M}^{-1}\mathbb{A}_k\rangle\right], \tag{15}$$

where $S = KK^\top$ and $P_k = L_k L_k^\top$ are Cholesky decompositions. The parameters are $P_k \in \mathcal{S}^n_{++}$ and $A_k \in \mathcal{L}^n$ for each class $k = 1, \cdots, C$.

Residual blocks

$$x^{(i)} = \operatorname{Exp}_{x^{(i-1)}}\left(\ell_i(x^{(i-1)})\right), \text{ where } \ell_i : \mathcal{M} \to T\mathcal{M}$$

$$Y = \operatorname{Exp}_X\left(Q \operatorname{diag}\left(f(\operatorname{spec}(X))\right) Q^T\right),$$

EXPERIMENTS

Effectiveness

Table 2: SPD MLRs under different metrics on the SPDNet backbone. The best two results are highlighted in **red** and **blue**.

(a) Radar

Metric	Acc	Time
AIM	**94.53 ± 0.95**	0.80
LEM	93.55 ± 1.21	0.76
LCM	93.49 ± 1.25	0.72
θ-PCM	**95.79 ± 0.38**	0.72
θ-BWCM	93.93 ± 0.79	0.71

(a) HDM05

Metric	1-Block		2-Block		3-Block	
	Acc	Time	Acc	Time	Acc	Time
AIM	58.07 ± 0.64	17.32	60.72 ± 0.62	18.75	61.14 ± 0.94	19.23
LEM	56.97 ± 0.61	2.21	60.69 ± 1.02	2.92	60.28 ± 0.91	3.50
LCM	60.69 ± 1.89	1.83	62.61 ± 1.46	2.40	62.33 ± 2.15	2.90
θ-PCM	**62.51 ± 1.65**	1.58	**63.66 ± 1.30**	2.29	**65.75 ± 2.86**	2.76
θ-BWCM	**62.71 ± 0.88**	1.64	**64.52 ± 0.56**	2.27	**67.40 ± 0.90**	2.87

(c) FPHA

Metric	Acc	Time
AIM	85.57 ± 0.50	7.14
LEM	85.90 ± 0.47	0.98
LCM	**86.37 ± 0.59**	0.74
θ-PCM	**89.40 ± 0.13**	0.69
θ-BWCM	86.27 ± 0.60	0.70

Table 3: SPD MLRs on the GyroSPD backbone.

Metric	Radar		HDM05		FPHA	
	Acc	Time	Acc	Time	Acc	Time
AIM	**96.80 ± 0.59**	1.23	66.05 ± 1.80	21.65	85.77 ± 0.52	11.48
LEM	96.58 ± 0.27	1.18	66.42 ± 0.47	2.02	85.87 ± 0.79	1.22
LCM	96.29 ± 0.53	1.12	68.37 ± 0.66	1.66	89.83 ± 0.28	0.98
θ-PCM	**97.04 ± 0.64**	1.18	**71.93 ± 1.21**	1.51	**91.17 ± 0.30**	1.00
θ-BWCM	96.21 ± 0.25	1.05	**72.74 ± 0.43**	1.58	**91.00 ± 0.11**	0.96

Table 3: SPD MLRs on the GyroSPD backbone.

Metric	Radar		HDM05		FPHA	
	Acc	Time	Acc	Time	Acc	Time
AIM	**96.80 ± 0.59**	1.23	66.05 ± 1.80	21.65	85.77 ± 0.52	11.48
LEM	96.58 ± 0.27	1.18	66.42 ± 0.47	2.02	85.87 ± 0.79	1.22
LCM	96.29 ± 0.53	1.12	68.37 ± 0.66	1.66	89.83 ± 0.28	0.98
θ-PCM	**97.04 ± 0.64**	1.18	**71.93 ± 1.21**	1.51	**91.17 ± 0.30**	1.00
θ-BWCM	96.21 ± 0.25	1.05	**72.74 ± 0.43**	1.58	**91.00 ± 0.11**	0.96

EXPERIMENTS

Stability

Table 5: Failure probabilities (%) of geodesics under different metrics with small eigenvalues in $L \in \mathcal{L}^n_{++}$. An output matrix containing any INF or NAN is considered a failure. Here, DLM denotes the diagonal log metric, while DPM and DBWM denote θ-DPM and θ-DBWM, respectively.

ϵ	3 × 3 for small matrices							256 × 256 for large matrices						
	DLM	$\theta = 1.5$		$\theta = 0.5$		$\theta = 0.15$		DLM	$\theta = 1.5$		$\theta = 0.5$		$\theta = 0.15$	
		DPM	DBWM	DPM	DBWM	DPM	DBWM		DPM	DBWM	DPM	DBWM	DPM	DBWM
$1e^{-1}$	0.62	0	0	0	0	0	0	14.29	0	0	0	0	0	0
$1e^{-2}$	5.70	0	0	0	0	0	0	18.48	0	0	0	0	0	0
$1e^{-3}$	51.32	0	0	0	0	0	0	58.35	0	0	0	0	0	0
$1e^{-4}$	94.34	0	0	0	0	0	0	95.02	0	0	0	0	0	0
$1e^{-5}$	99.39	0	0	0	0	0	0	99.47	0	0	0	0	0	0
$1e^{-10}$	100	0	0	0	0	0	0	100	0	0	0	0	0	0
$1e^{-15}$	100	0	0	0	0	0	0	100	0	0	0	0	0	0
$1e^{-20}$	100	0	0	0	0	0	0.002	100	0	0	0	0	0	0.02
$1e^{-21}$	100	0	0	0	0	0	0.03	100	0	0	0	0	0	0.01
$1e^{-22}$	100	0	0	0	0	0	0.25	100	0	0	0	0	0	0.23
$1e^{-23}$	100	0	0	0	0	0	2.26	100	0	0	0	0	0	2.42
$1e^{-24}$	100	0	0	0	0	0	22.98	100	0	0	0	0	0	23.13
$1e^{-25}$	100	0	0	0	0	0	86.34	100	0	0	0	0	0	86.58
$1e^{-30}$	100	0	0	0	0	0	100	100	0	0	0	0	0	100

Experiments

Efficiency

Table 12: Number of matrix functions required per sample for a C-class SPD MLR. Spectral matrix functions include matrix logarithm, matrix power, and the Lyapunov operator.

Metric	Num. spectral matrix functions	Num. Cholesky decomposition
AIM	$1+2C$	0
LEM	$1+C$	0
LCM	0	$1+C$
PEM	$1+C$	0
BWM	$1+3C$	C
PCM	0	$1+C$
BWCM	0	$1+C$

Table 13: Asymptotic per-sample complexity of a C-class SPD MLR for an $n \times n$ input SPD matrix.

Metric	Asymptotic complexity
AIM	$O\big(9(1+2C)n^3\big)$
LEM	$O\big(9(1+C)n^3\big)$
LCM	$O\big(\frac{1+C}{3}n^3\big)$
PEM	$O\big(9(1+C)n^3\big)$
BWM	$O\big((9(1+3C)+\frac{C}{3})n^3\big)$
PCM	$O\big(\frac{1+C}{3}n^3\big)$
BWCM	$O\big(\frac{1+C}{3}n^3\big)$

Table 14: Average runtime (in seconds) of one SPD MLR training step under different dimensions.

Dim	AIM	LEM	LCM	PEM	BWM	PCM	BWCM
32	0.2380	0.0077	0.0046	0.0076	0.2377	0.0040	0.0040
64	1.0139	0.0395	0.0303	0.0473	1.1205	0.0251	0.0225
128	3.6256	0.1832	0.1490	0.1844	4.0674	0.1013	0.1019
256	14.5142	0.7793	0.5833	0.7853	16.5918	0.3848	0.4077
512	60.1918	3.2948	2.5030	3.4357	70.8647	1.7553	1.7526

Thanks

Homepage

LinkedIn

UNIVERSITÀ DI TRENTO

清華大學 Tsinghua University

上海期智研究院 SHANGHAI QI ZHI INSTITUTE

江南大學 JIANGNAN UNIVERSITY