A RIEMANNIAN FRAMEWORK FOR NORMALIZED SPD MATRIX LEARNINGZ S

Anonymous authors
Paper under double-blind review

ABSTRACT

Symmetric positive definite (SPD) matrices are often normalized before learning, yet subsequent methods commonly use geometry defined on the full SPD cone. We study this bounded-spectrum domain as normalized SPD (NSPD) and define a family of Deformed-Euclidean Metrics (DEMs) directly on it. Each metric pulls back Euclidean geometry through a scalar spectral diffeomorphism, giving closed-form Riemannian operators. We instantiate five deformations and develop eigendecomposition-free approximations of their forward maps for large matrices. Experiments in visual recognition, Riemannian classification, and fMRI connectivity generation show gains over matched logarithm-based methods.

1 INTRODUCTION

Symmetric positive definite (SPD) matrices represent second-order statistics in visual recognition (Lin et al., 2015; Li et al., 2017; Wang et al., 2021), electroencephalography (EEG) decoding (Barachant et al., 2012; Pan et al., 2022), and neuroimaging (You & Park, 2021; Collas et al., 2025). Learning on these matrices commonly uses affine-invariant or log-Euclidean geometry (Pennec et al., 2006; Arsigny et al., 2006), with corresponding Riemannian networks and normalization layers (Huang & Gool, 2017; Brooks et al., 2019; Chen et al., 2024a).

Many SPD learning pipelines normalize their matrix inputs. In covariance pooling and iterative whitening, trace or Frobenius normalization places the spectrum in the convergence range of fast matrix-function algorithms (Li et al., 2018; Huang et al., 2019; Song et al., 2022b; Rahman et al., 2026). In EEG decoding, common spatial patterns and related methods use trace normalization (Ramoser et al., 2000; Samek et al., 2012; Pan et al., 2022). Both normalizations also remove overall scale, assigning the same representative to A and cA for $c > 0$. For dimension $n \geq 2$, either operation places every eigenvalue in $(0, 1)$. These normalized matrices thus lie in a bounded subdomain of the SPD cone (Figure 1). Yet subsequent processing can still use geometry defined on the full cone, as in MAtt (Pan et al., 2022). This motivates geometry designed for the normalized domain and learning with its geometric operators.

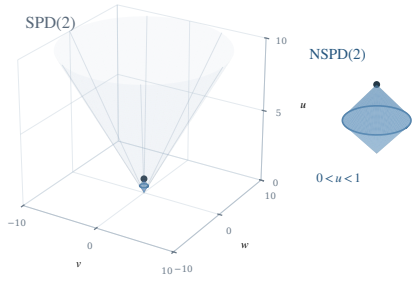


Figure 1: NSPD denotes normalized SPD. The gray cone bounds $\mathrm{SPD}(2)$ over $u \in [0, 10]$. The right view enlarges the blue boundary of $\mathrm{NSPD}(2)$ over $u \in [0, 1]$.

We use the *normalized SPD (NSPD) domain*, $\mathrm{NSPD}(n) = \mathrm{SPD}_{(0,1)}(n)$, as a common domain for both normalizations. A scalar diffeomorphism from $(0, 1)$ to $\mathbb{R}$, applied to each eigenvalue, maps NSPD onto symmetric matrices. Pulling back the Euclidean metric yields the *Deformed-Euclidean Metric (DEM) family*, whose distances, geodesics, and weighted Fréchet means follow in closed form from Euclidean operations (Arsigny et al., 2006; Chen et al., 2024b). This makes geometric learning convenient to implement, with no iterative manifold optimization for these operators. We provide five scalar deformations and fast polynomial and rational approximations for large matrices (Higham, 2008; Trefethen, 2013).

Our contributions are threefold. (1) We identify NSPD as a common bounded-spectrum domain for normalized SPD representations and design geometry directly on it. (2) We construct the DEM

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

family through scalar diffeomorphisms, providing closed-form Riemannian operators that make geometric learning computationally convenient. (3) We enable fast evaluation through polynomial and rational approximations without eigendecomposition, and validate the framework in visual recognition, Riemannian classification, and fMRI generation against matched baselines.

2 Related Work

SPD representations and learning. Second-order and covariance representations have been widely used in visual recognition Lin et al. (2015); Li et al. (2017); Acharya et al. (2018); Li et al. (2018); Wang et al. (2021), while Riemannian neural networks directly exploit the manifold structure of SPD matrices Huang & Gool (2017); Brooks et al. (2019); Katsman et al. (2023); Chen et al. (2024a). Recent work further analyzes matrix-function normalizations in covariance pooling through a Riemannian-geometric lens Chen et al. (2025). Metric learning, kernel construction, and geometry-aware dimensionality reduction provide complementary approaches for SPD-valued data Huang et al. (2015); Jayasumana et al. (2015); Harandi et al. (2018). Covariance descriptors have also been studied for visual recognition and functional-connectivity analysis Ilea et al. (2018); Dodero et al. (2015); You & Park (2021). Most of these methods operate on the full SPD cone or use a fixed matrix function as the representation map. A related line of work instead controls scale, notably through full-rank correlation matrices, where variable-wise scales are removed by fixing the diagonal Thanwerdas & Pennec (2022); Chen et al. (2026). Our normalized representation is different in that it removes only one global positive scale, thereby retaining eigendirections and eigenvalue ratios.

Riemannian geometries on SPD matrices. The affine-invariant and log-Euclidean metrics are classical approaches to SPD geometry, within the broader Riemannian framework for tensor-valued data Pennec et al. (2006); Arsigny et al. (2005; 2006); Thanwerdas & Pennec (2023). Alternative coordinate geometries, including Cholesky-based constructions, provide different computational trade-offs Lin (2019); Ilea et al. (2018). More general pullback-Euclidean constructions retain this computational advantage while allowing alternative representation maps Chen et al. (2024b). Pullback geometry has also been used for generative modeling of brain-connectivity matrices Chen & Lipman (2024); Collas et al. (2025). DEMs belong to this pullback-Euclidean family, but specialize the representation-map design to scalar spectral deformations on a normalized, bounded spectral domain.

3 Preliminaries

This section introduces the notation and geometric objects used throughout the paper. Additional background on smooth and Riemannian manifolds is provided in Section A.

Notation. Scalars, vectors, and matrices are denoted by lowercase letters, bold lowercase letters, and uppercase letters, respectively. We write $\mathbb{R}$, $\mathbb{R}_{++}$, and $O(n)$ for the real numbers, the positive real numbers, and the orthogonal group. The identity matrix is denoted by I. For a matrix X, $X^\top$, $\mathrm{tr}(X)$, $\|X\|_F$, and $\lambda_i(X)$ denote its transpose, trace, Frobenius norm, and i-th eigenvalue. The largest eigenvalue is denoted by $\lambda_{\max}(X)$, and $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. For a Riemannian manifold $(\mathcal{M}, g)$, $T_p\mathcal{M}$ denotes the tangent space at $p \in \mathcal{M}$. We use Exp_p, Log_p, d_g, and $\Gamma_{p\to q}$ for the Riemannian exponential map, logarithmic map, geodesic distance, and parallel transport from p to q, respectively.

Pullback metrics. Suppose $\mathcal{M}$ and $\mathcal{N}$ are smooth manifolds, g is a Riemannian metric on $\mathcal{N}$, and $f : \mathcal{M} \to \mathcal{N}$ is smooth. The pullback of the tensor field g by f is defined pointwise as

$$(f^*g)_p(V_1, V_2) = g_{f(p)}\left(Df_p[V_1], Df_p[V_2]\right), \tag{1}$$

where $p \in \mathcal{M}$, Df_p is the differential of f at p, and $V_i \in T_p\mathcal{M}$. If f^*g is positive definite, it defines a Riemannian metric on $\mathcal{M}$ and is called the pullback metric induced by f. The most common pullback metrics are induced by diffeomorphisms. In this case, f is an isometry between $(\mathcal{M}, f^*g)$ and $(\mathcal{N}, g)$. Geometric operations on $\mathcal{M}$ can therefore be obtained by transferring the corresponding operations on $\mathcal{N}$ through f. This principle underlies the log-Euclidean metric Arsigny et al. (2006) and the broader class of pullback-Euclidean metrics on SPD matrices Chen et al. (2024b).

SPD spaces. The spaces of real symmetric and symmetric positive definite matrices are denoted by $\mathrm{Sym}(n) = \{X \in \mathbb{R}^{n\times n} \mid X = X^\top\}$ and $\mathrm{SPD}(n) = \{X \in \mathrm{Sym}(n) \mid X \succ 0\}$, respectively. The vector space $\mathrm{Sym}(n)$ has dimension $n(n+1)/2$. Since $\mathrm{SPD}(n)$ is an open subset of $\mathrm{Sym}(n)$, its tangent space satisfies $T_X \,\mathrm{SPD}(n) = \mathrm{Sym}(n)$ for every $X \in \mathrm{SPD}(n)$. Each $X \in \mathrm{SPD}(n)$ admits an eigendecomposition $X = U\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)U^\top$, where $U \in O(n)$ and $\lambda_i > 0$. Accordingly, the spectrum of the full SPD cone ranges over $(0, \infty)$ and is unbounded.

The bounded-spectrum domain used by our method and the normalization that produces the observed matrices are introduced separately in Section 4.1.

4 Deformed-Euclidean Geometry for Normalized SPD Representations

4.1 Normalized SPD Representations

The NSPD domain consists of matrices $X \in \mathrm{Sym}(n)$ with $0 \prec X \prec I$. Its matrices have eigenvalues strictly between zero and one. We obtain inputs by Frobenius normalization: for $A \in \mathrm{SPD}(n)$ with $n \geq 2$, set

$$X = \frac{A}{\|A\|_F}. \tag{2}$$

Since $\|A\|_F > \lambda_{\max}(A)$, X lies in the strict open domain. This normalization preserves positive definiteness, eigenvectors, and eigenvalue ratios, and is invariant under positive rescaling of A. Trace normalization also yields matrices in NSPD for $n \geq 2$. We use Frobenius normalization by default, while defining the metrics on the full domain. Figure 2 shows the normalized region and the unit-Frobenius-norm section. We next construct the geometry on NSPD.

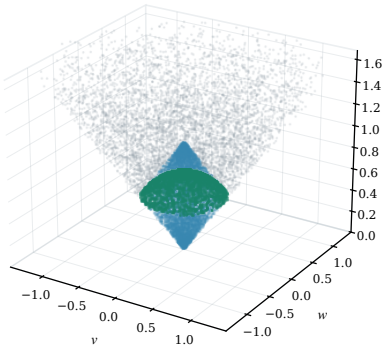


Figure 2: $\mathrm{SPD}(2)$ (gray), $\mathrm{NSPD}(2)$ (blue), and its unit-Frobenius-norm section (green).

4.2 Scalar Diffeomorphisms for NSPD Geometry

Inspired by the log-Euclidean metric (LEM) (Arsigny et al., 2006) and adaptive log-Euclidean metrics (ALEM) (Chen et al., 2024b), we use the spectral viewpoint: the matrix logarithm applies the scalar diffeomorphism $\log : (0, \infty) \to \mathbb{R}$ to each eigenvalue. The intervals $(0, 1)$ and $\mathbb{R}$ are homeomorphic (Munkres, 2000) and admit smooth diffeomorphisms. We choose one, $f : (0, 1) \to \mathbb{R}$, for NSPD. For $X = U\,\mathrm{diag}(\mu_1, \ldots, \mu_n)U^\top \in \mathrm{NSPD}(n)$, define its matrix function :

$$f(X) := U\,\mathrm{diag}\left(f(\mu_1), \ldots, f(\mu_n)\right)U^\top. \tag{3}$$

For $Z = V\,\mathrm{diag}(z_1, \ldots, z_n)V^\top \in \mathrm{Sym}(n)$, its inverse is

$$f^{-1}(Z) = V\,\mathrm{diag}\left(f^{-1}(z_1), \ldots, f^{-1}(z_n)\right)V^\top. \tag{4}$$

Table 1 gives five scalar diffeomorphisms and their inverses, with IAS denoting Inverse Algebraic Sigmoid.

Table 1: Representative scalar deformations and their inverses. The Cauchy map is named for the standard Cauchy quantile function, whose inverse is the cumulative distribution function (CDF).

	f_{Logit}	f_{IAS}	f_{Rational}	f_{Cauchy}	f_{Gumbel}
$f(x)$	$\log\frac{x}{1-x}$	$\frac{2x-1}{\sqrt{x(1-x)}}$	$\frac{2x-1}{x(1-x)}$	$\tan\left(\pi\left(x-\frac{1}{2}\right)\right)$	$-\log(-\log x)$
$f^{-1}(y)$	$\frac{1}{1+e^{-y}}$	$\frac{1}{2}\left(1+\frac{y}{\sqrt{y^2+4}}\right)$	$\frac{1}{2}\left(1+\frac{y}{\sqrt{y^2+4}+2}\right)$	$\frac{1}{2}+\frac{1}{\pi}\arctan(y)$	$\exp(-e^{-y})$

Theorem 4.1 (Global matrix diffeomorphism)**.** [↓] *For every smooth, strictly monotone diffeomorphism $f : (0, 1) \to \mathbb{R}$, the matrix function $f : \mathrm{NSPD}(n) \to \mathrm{Sym}(n)$ is a global diffeomorphism with inverse f^{-1} given by* (4)*.*

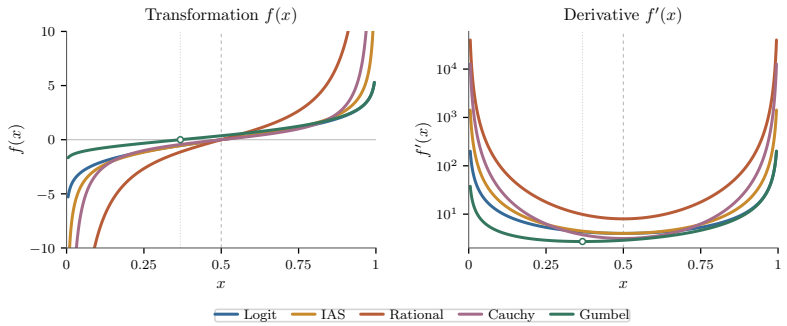


Figure 3: Scalar deformations and their derivatives. The first four satisfy $f(1-x) = -f(x)$, whereas Gumbel is asymmetric.

4.3 DEM Metric

Let g^{E} be the Euclidean metric induced by the Frobenius inner product on $\mathrm{Sym}(n)$. The spectral diffeomorphism f induces the metric on $\mathrm{NSPD}(n)$ by pullback:

$$g^f := f^* g^{\mathrm{E}}. \tag{5}$$

Varying f yields the *Deformed-Euclidean Metric (DEM) family*. The Rational, Cauchy, Logit, Inverse Algebraic Sigmoid (IAS), and Gumbel maps in Table 1 induce REM, CEM, LEM, ISAM, and GEM, respectively. Since $\mathrm{NSPD}(n)$ is open in $\mathrm{Sym}(n)$, its tangent space is $T_X \mathrm{NSPD}(n) = \mathrm{Sym}(n)$ for every $X \in \mathrm{NSPD}(n)$.

Theorem 4.2 (Deformed-Euclidean metric)**.** [↓] *For $X \in \mathrm{NSPD}(n)$ and $V, W \in T_X \mathrm{NSPD}(n)$, the metric is*

$$g^f_X(V, W) = \langle Df_X[V], Df_X[W] \rangle_F . \tag{6}$$

The map $f : (\mathrm{NSPD}(n), g^f) \to (\mathrm{Sym}(n), \langle \cdot, \cdot \rangle_F)$ is a Riemannian isometry. For $X, Y \in \mathrm{NSPD}(n)$ and $V \in T_X \mathrm{NSPD}(n)$, the induced distance and geodesic are

$$d_f(X, Y) = \|f(X) - f(Y)\|_F , \tag{7}$$

$$\gamma^f_{X \to Y}(t) = f^{-1}\left((1-t)f(X) + tf(Y)\right), \qquad t \in [0, 1]. \tag{8}$$

The exponential map, logarithmic map, and parallel transport are

$$\mathrm{Exp}^f_X(V) = f^{-1}\left(f(X) + Df_X[V]\right), \tag{9}$$

$$\mathrm{Log}^f_X(Y) = D(f^{-1})_{f(X)}\left[f(Y) - f(X)\right], \tag{10}$$

$$\Gamma^f_{X \to Y}(V) = D(f^{-1})_{f(Y)}\left[Df_X[V]\right]. \tag{11}$$

The induced operations

$$X \oplus_f Y = f^{-1}\left(f(X) + f(Y)\right), \qquad a \odot_f X = f^{-1}\left(af(X)\right) \tag{12}$$

for $a \in \mathbb{R}$ give $\mathrm{NSPD}(n)$ a real vector-space structure. In particular, $(\mathrm{NSPD}(n), \oplus_f)$ is an Abelian Lie group.

For $X_1, \ldots, X_m \in \mathrm{NSPD}(n)$ and positive weights $w_1, \ldots, w_m$, the weighted Fréchet mean exists uniquely and is

$$\mathrm{FM}^w_f(X_1, \ldots, X_m) = \operatorname*{arg\,min}_{X \in \mathrm{NSPD}(n)} \sum_{i=1}^m w_i d^2_f(X, X_i) = f^{-1}\left(\frac{\sum_{i=1}^m w_i f(X_i)}{\sum_{i=1}^m w_i}\right). \tag{13}$$

For $X = U \,\mathrm{diag}(\mu) U^\top$ and $V \in \mathrm{Sym}(n)$, the differential used above is computed by the standard Loewner formula Higham (2008):

$$Df_X[V] = U\left(L_f(\mu) \circ \widetilde{V}\right) U^\top, \tag{14}$$

where $\widetilde{V} = U^\top V U$, $\circ$ is the Hadamard product, and $L_{f,ij} = (f(\mu_i) - f(\mu_j))/(\mu_i - \mu_j)$ for distinct eigenvalues, with continuous extension $L_{f,ij} = f'(\mu_i)$ when they coincide. The inverse differential $D(f^{-1})_{f(X)}$ follows from the same formula using f^{-1} at the eigenvalues of $f(X)$.

4.4 Properties of Deformed-Euclidean Metrics

Every exact DEM is globally isometric to Euclidean space. The choice of f determines how the metric weights perturbations at each point.

Proposition 4.3 (Global geometry)**.** [↓] *Every exact DEM on* $\mathrm{NSPD}(n)$ *is flat and geodesically complete, and every pair of points is joined by a unique minimizing geodesic.*

Proposition 4.4 (Bi-invariance)**.** [↓] *The metric* g^f *is bi-invariant on the Abelian Lie group* $(\mathrm{NSPD}(n), \oplus_f)$*, whose identity is* $f^{-1}(0)I$*. Group inversion* $\ominus_f X = f^{-1}(-f(X))$ *is also an isometry. Hence, for* $A, X, Y \in \mathrm{NSPD}(n)$*,*

$$\begin{aligned} d_f(A \oplus_f X, A \oplus_f Y) &= d_f(X \oplus_f A, Y \oplus_f A) \\ &= d_f(\ominus_f X, \ominus_f Y) = d_f(X, Y). \end{aligned} \tag{15}$$

Proposition 4.5 (Orthogonal and scale invariance)**.** [↓] *For* $Q \in O(n)$*, conjugation* $X \mapsto QXQ^\top$ *is an isometry of* $(\mathrm{NSPD}(n), g^f)$*. In particular, for* $X, Y, X_1, \ldots, X_m \in \mathrm{NSPD}(n)$ *and positive weights* w_i*,*

$$d_f(QXQ^\top, QYQ^\top) = d_f(X, Y), \tag{16}$$

$$\mathrm{FM}_f^w(QX_1Q^\top, \ldots, QX_mQ^\top) = Q\,\mathrm{FM}_f^w(X_1, \ldots, X_m)Q^\top. \tag{17}$$

Proposition 4.6 (Power invariance of GEM)**.** [↓] *For* $f = f_{\mathrm{Gumbel}}$ *and* $\beta > 0$*, the power map* $X \mapsto X^\beta$ *on* $\mathrm{NSPD}(n)$ *is a deformed translation:*

$$X^\beta = (e^{-\beta} I) \oplus_f X. \tag{18}$$

It is therefore an isometry. For $X, Y, X_1, \ldots, X_m \in \mathrm{NSPD}(n)$ *and positive weights* w_i*,*

$$d_f(X^\beta, Y^\beta) = d_f(X, Y), \tag{19}$$

$$\mathrm{FM}_f^w(X_1^\beta, \ldots, X_m^\beta) = \left(\mathrm{FM}_f^w(X_1, \ldots, X_m)\right)^\beta. \tag{20}$$

Corollary 4.7 (Complement symmetry)**.** [↓] *For Logit, IAS, Rational, and Cauchy, the scalar symmetry* $f(1-x) = -f(x)$ *induces, for* $X, Y \in \mathrm{NSPD}(n)$*,*

$$f(I - X) = -f(X), \qquad d_f(I - X, I - Y) = d_f(X, Y). \tag{21}$$

For these four maps, $f^{-1}(0) = 1/2$ *and* $\ominus_f X = I - X$*.*

As illustrated in Figure 3, Logit has the slowest endpoint derivative growth among the four symmetric choices, followed by IAS. Rational and Cauchy have the same asymptotic order but different interior profiles, while Gumbel treats the two endpoints asymmetrically. The complement isometry holds on $\mathrm{NSPD}(n)$ and need not preserve the subset of Frobenius-normalized inputs. The proofs of the invariances are in Section A.5; endpoint analysis and the corresponding Fréchet-mean symmetry are in Section A.4. The closed-form group operations also support Lie-group normalization methods such as LieBN (Chen et al., 2024a).

5 Scalable Computation for Large SPD Matrices

Spectral evaluation of the DEM maps is practical at moderate dimensions, but eigendecomposition becomes the dominant cost for large covariance matrices. We therefore evaluate the forward transformation using polynomial and rational matrix-function approximations that avoid explicit eigendecomposition Higham (2008); Trefethen (2013); Sastre (2012); Al-Mohy & Higham (2012); Jarlebring et al. (2025); Li et al. (2018); Song et al. (2021; 2022b). Following the pre-normalization and matrix-arithmetic approach of Rahman et al. (2026), we approximate the target scalar map on a specified spectral interval. Our classification experiments use Frobenius normalization to construct the bounded-spectrum input for DEMs. We evaluate the approximations directly at $B = A/\|A\|_F$ and fit the scalar coefficients on a prescribed interval $[a, b] \subset (0, 1)$. This fitting interval is specified separately from the input normalization. For example, a degree-K Chebyshev approximation gives

$$\widehat{f}_K(B) = \sum_{k=0}^{K} c_k T_k(S), \qquad S = \frac{2B - (a+b)I}{b-a}. \tag{22}$$

The coefficients depend on the deformation, working interval, and degree and are precomputed before training. The basis matrices are evaluated by

$$T_0(S) = I, \qquad T_1(S) = S, \qquad T_{k+1}(S) = 2ST_k(S) - T_{k-1}(S). \tag{23}$$

Thus, a degree-K evaluation uses $O(K)$ dense matrix multiplications without an eigendecomposition. The other polynomial families likewise use matrix multiplications and additions, while Padé additionally requires a linear solve. These implementations change the computational primitives, not the cubic dependence on matrix dimension. Accuracy and runtime are evaluated empirically in Section B.5. The coefficient constructions, recurrences, and specialized matrix formulas are provided in Section B.4. These schemes approximate the forward transformation on the working interval. Applications requiring the global inverse, including our generative experiments, retain the exact spectral maps.

6 Experiments

Experimental overview. Synthetic experiments assess approximation accuracy and runtime. Classification and fMRI generation evaluate DEM representations, Riemannian operators, and exact inverse maps, with matched log-based controls. Training protocols are provided in Section D.

6.1 Synthetic Accuracy and Runtime

Approximations matched to each deformation closely reproduce exact spectral evaluation. Using batch size 32, FP64, Frobenius normalization, fitting interval $[0.05, 0.95]$, and degree $K = 8$, Chebyshev attains relative Frobenius error 2.740×10^{-3} for LEM at $n = 1024$, while Padé reaches 4.424×10^{-6} for REM. These choices give the lowest errors for the respective maps across all four tested dimensions (Table 7). The preferred approximation depends on the scalar deformation, while its error changes little across the tested matrix dimensions. The scalar map is therefore a key consideration when selecting an approximation family. Table 2 shows all five approximation families at $n = 1024$; results for all four dimensions remain in Table 7.

Table 2: Relative Frobenius errors at $n = 1024$ (FP64, $K = 8$). The log row is the reference; bold marks the lowest error in each row.

Map	Taylor	Chebyshev	Legendre	Padé	Laguerre
log	6.852e-02	**2.062e-03**	3.863e-03	3.633e-03	1.094e-02
LEM	7.719e-02	**2.740e-03**	5.011e-03	2.065e-02	1.450e-02
ISAM	1.697e-01	**7.600e-03**	1.362e-02	3.514e-02	3.497e-02
REM	3.718e-01	2.156e-02	3.744e-02	**4.424e-06**	8.568e-02
CEM	3.525e-01	2.042e-02	3.553e-02	**9.169e-04**	8.131e-02
GEM	6.158e-02	**2.891e-03**	5.067e-03	1.677e-02	1.550e-02

Matrix-arithmetic evaluation also reduces computational cost. At $n = 1024$, the tested Chebyshev and Padé implementations take approximately 109 and 115 ms for forward and backward evaluation, compared with 535–547 ms for eigendecomposition-based log. Section B.5 specifies the benchmark and comparison. Together with the accuracy results, this shows that matrix-arithmetic evaluation can approximate the spectral map while avoiding eigendecomposition in the tested large-matrix setting.

6.2 Classification

Classification protocol. Fine-grained and ImageNet-1K experiments use Frobenius-normalized covariance pooling, followed by a DEM and a Euclidean classifier. Fine-grained experiments fine-tune ImageNet-pretrained backbones, whereas ImageNet-1K trains ResNet-50 from scratch. DEMs and MLN use matched approximation families and orders. Full settings are given in Section D.

Fine-grained recognition with pretrained backbones. REM and CEM deliver consistent gains across convolutional and transformer backbones in Figure 4. Each achieves higher mean accuracy than the matched MLN reference in all 45 configurations formed by nine backbone–dataset settings and five approximation families. For example, Chebyshev REM improves ResNet-101 on Aircraft

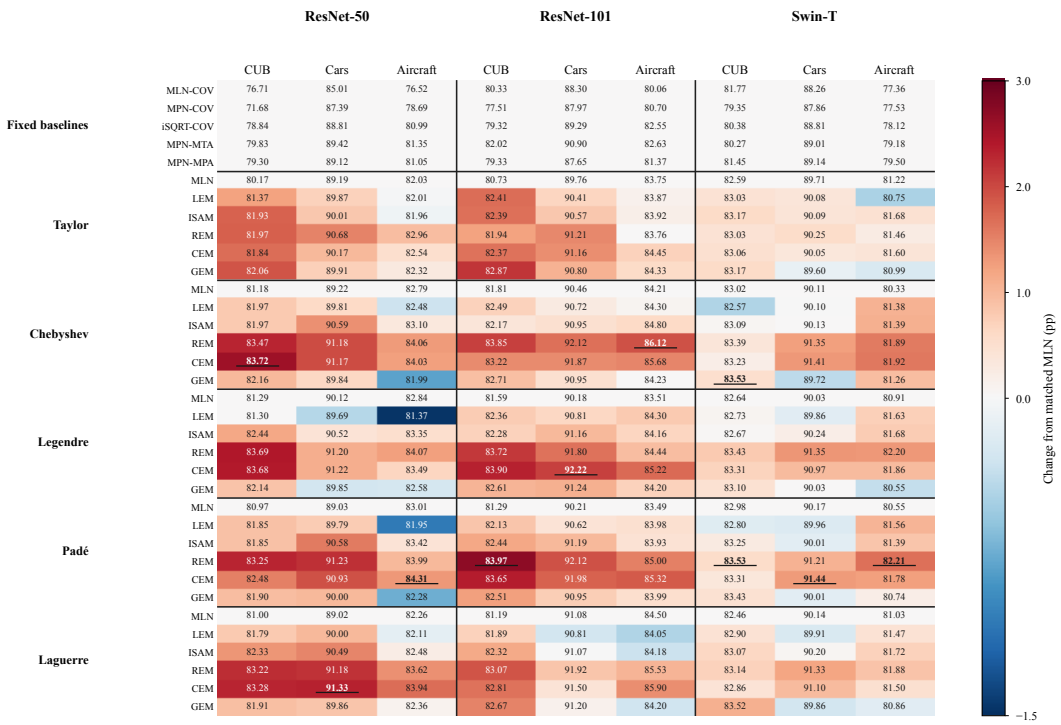


Figure 4: Fine-grained Top-1 accuracy (%). MLN-COV and MPN-COV use exact log and square root, respectively. REM, CEM, LEM, ISAM, and GEM are the five DEMs. Each group matches MLN and DEM approximations. Colours show gains over matched MLN.

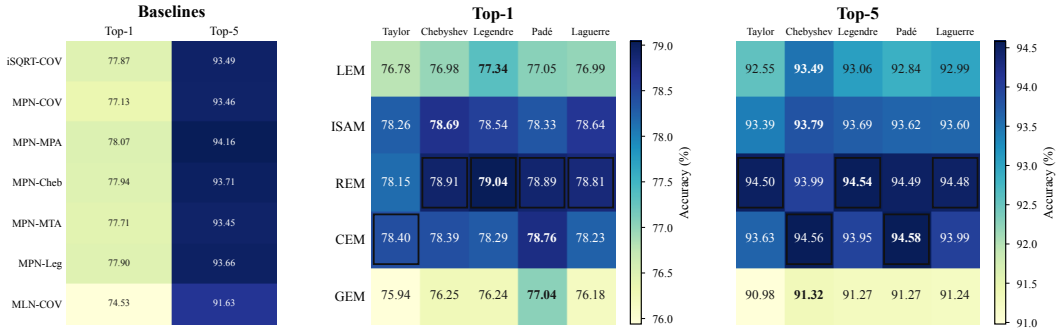


Figure 5: ImageNet-1K Top-1/Top-5 accuracy (%) with covariance-pooling baselines (left) and DEM approximation grids (middle/right). Grid rows are LEM, ISAM, REM, CEM, and GEM, from top to bottom.

from 84.21% to 86.12%. These gains span both polynomial and rational approximations across the tested architectures. Within each matched block, the backbone, normalization, and approximation family are fixed. The persistent gains therefore link the improved covariance representation to the choice of scalar deformation.

ImageNet-1K training from scratch. The benefits extend to large-scale training from scratch (Figure 5). REM with Legendre reaches 79.04% Top-1 accuracy, while CEM with Padé reaches 94.58% Top-5 accuracy. Both outperform all plotted baselines in Top-1 accuracy with every approximation family. The ordering also persists when the backbone is trained from scratch, so the fine-grained pattern does not depend on fine-tuning an ImageNet-pretrained feature extractor. REM and CEM remain strong across all five approximation bases, indicating that the choice of deformation transfers across numerical realizations.

Table 3: Classification accuracy (%) in Riemannian ResNet and BatchNorm. The log and square-root rows are baselines; the remaining rows are DEMs.

Method	FPHA	HINSS cross-session	HINSS cross-subject	Cora
log	64.58 ± 0.98	47.95 ± 4.88	45.59 ± 4.36	77.60 ± 2.78
square root	62.51 ± 0.36	45.07 ± 3.89	41.72 ± 4.00	76.11 ± 2.09
LEM *(ours)*	67.74 ± 1.15	$\mathbf{49.05 \pm 4.93}$	46.34 ± 5.44	$\mathbf{81.00 \pm 0.50}$
ISAM *(ours)*	$\mathbf{69.85 \pm 1.29}$	48.45 ± 4.85	$\mathbf{46.47 \pm 4.42}$	78.20 ± 3.72
REM *(ours)*	68.41 ± 1.18	46.55 ± 6.59	43.37 ± 5.40	71.00 ± 8.07
CEM *(ours)*	68.88 ± 0.98	47.00 ± 5.64	46.39 ± 4.66	72.90 ± 8.82
GEM *(ours)*	40.37 ± 1.16	48.61 ± 4.72	45.09 ± 5.74	79.20 ± 2.84

Riemannian ResNet and BatchNorm classification. We instantiate DEM operators in an RResNet with Riemannian BatchNorm Brooks et al. (2019); Chen et al. (2024a). We compare them with log and square-root maps on FPHA, HINSS cross-session, HINSS cross-subject, and Cora.

ISAM improves FPHA accuracy over log by 5.27 percentage points and leads HINSS cross-subject in mean accuracy. LEM leads HINSS cross-session and improves Cora over log by 3.40 points (Table 3). The leading DEM differs between protocols, consistent with distinct spectral weighting being useful for different covariance statistics. More broadly, these gains arise when the map is used inside a Riemannian network and normalization pipeline, extending the evidence beyond terminal covariance pooling.

6.3 fMRI Connectivity Generation

fMRI protocol. Following the pullback-flow formulation of DiffeoCFM, each scan is converted to an SPD connectivity matrix and mapped to a Euclidean representation before flow matching:

$$A \xrightarrow{\mathcal{N}} X \xrightarrow{f} Z \xrightarrow{\mathrm{CFM}} \widehat{Z} \xrightarrow{f^{-1}} \widehat{G} \xrightarrow{\mathrm{Corr}} \widehat{C}. \tag{24}$$

Here $\mathcal{N}(A) = A/\operatorname{tr}(A)$. This trace normalization keeps the connectivity inputs in the NSPD domain and fixes total variance across scans. The forward and inverse maps are evaluated exactly. Generated matrices are returned to the correlation space only for final quality and downstream-utility evaluation.

We evaluate the framework in DiffeoCFM for fMRI connectivity generation on ABIDE, ADNI, and OASIS3 Collas et al. (2025). The DiffeoGauss, TriangDDPM, TriangCFM, and DiffeoCFM baselines, together with all DEM variants, are rerun under the same subject-level splits and training protocol. We report distributional fidelity through α, β, and their harmonic score H, and downstream utility through CAS-AUC and CAS-F1. The complete results are reported in Table 4.

DEMs lead the compared generative methods in distributional fidelity on ABIDE and OASIS3. Relative to the matched DiffeoCFM baseline rerun under the same protocol, REM raises H on ABIDE from 0.590 to 0.715, and GEM raises it on OASIS3 from 0.440 to 0.684. LEM achieves the highest generative CAS-F1 on ADNI, and CEM leads this metric on OASIS3. On ABIDE and OASIS3, the higher H scores accompany gains in both α and β relative to matched DiffeoCFM. Because the splits, flow model, and exact forward and inverse evaluation are shared, the comparison locates these differences in the representation geometry. The distinct leaders in fidelity and CAS also show the value of offering several deformations within the family.

6.4 Ablation Studies

Normalization and approximation interaction. The classification gains persist under both trace and Frobenius normalization. On Cars, all five DEMs improve over the corresponding MLN reference when accuracy is averaged over the five degree-8 approximation families (Table 11). REM leads under trace normalization, and LEM leads under Frobenius normalization. Thus, the benefit of replacing the logarithm extends across both normalization choices in this experiment. The ordering among DEMs changes with normalization, which affects the input spectrum. Their shared advantage over MLN persists after changing the normalization.

Normalized exact-log control for fMRI. The fidelity gains also hold against exact log with the same normalization, $X = A/\operatorname{tr}(A)$, on ten subject-group splits (Table 12). REM improves H on

Table 4: fMRI connectivity-generation results. REM, CEM, LEM, ISAM, and GEM are the five DEMs. All generative methods are rerun under the same protocol. Entries are mean ± standard deviation. Bold and underlined entries denote the best and second-best generative results. A thin rule separates DEMs from the baselines within each dataset.

Method	α	β	H	CAS-AUC	CAS-F1
			ABIDE		
Real Data	0.800 ± 0.080	0.790 ± 0.080	0.790 ± 0.030	0.670 ± 0.060	0.590 ± 0.070
DiffeoGauss	0.560 ± 0.060	0.290 ± 0.060	0.380 ± 0.060	$\mathbf{0.660 \pm 0.040}$	$\underline{0.530 \pm 0.060}$
TriangDDPM	0.040 ± 0.020	0.000 ± 0.000	0.000 ± 0.000	0.530 ± 0.060	0.470 ± 0.120
TriangCFM	0.040 ± 0.020	0.000 ± 0.000	0.000 ± 0.000	0.520 ± 0.050	0.400 ± 0.180
DiffeoCFM	0.770 ± 0.090	0.480 ± 0.070	0.590 ± 0.080	$\underline{0.640 \pm 0.060}$	$\mathbf{0.580 \pm 0.070}$
LEM *(ours)*	$\mathbf{0.863 \pm 0.039}$	$\underline{0.604 \pm 0.069}$	$\underline{0.708 \pm 0.047}$	0.597 ± 0.055	0.469 ± 0.084
ISAM *(ours)*	0.136 ± 0.039	0.413 ± 0.056	0.204 ± 0.049	0.599 ± 0.056	0.527 ± 0.106
REM *(ours)*	0.857 ± 0.024	$\mathbf{0.617 \pm 0.079}$	$\mathbf{0.715 \pm 0.053}$	0.604 ± 0.053	0.519 ± 0.065
CEM *(ours)*	$\underline{0.861 \pm 0.030}$	0.592 ± 0.076	0.700 ± 0.058	0.601 ± 0.052	0.521 ± 0.068
GEM *(ours)*	0.854 ± 0.043	0.582 ± 0.083	0.689 ± 0.062	0.584 ± 0.050	0.513 ± 0.097
			ADNI		
Real Data	0.910 ± 0.030	0.850 ± 0.060	0.880 ± 0.030	0.620 ± 0.050	0.620 ± 0.050
DiffeoGauss	0.020 ± 0.010	0.510 ± 0.080	0.040 ± 0.020	$\underline{0.600 \pm 0.050}$	0.290 ± 0.130
TriangDDPM	0.020 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.530 ± 0.060	0.180 ± 0.110
TriangCFM	0.020 ± 0.000	0.000 ± 0.010	0.010 ± 0.010	0.560 ± 0.040	0.340 ± 0.130
DiffeoCFM	$\underline{0.620 \pm 0.110}$	$\mathbf{0.770 \pm 0.020}$	$\mathbf{0.680 \pm 0.060}$	$\mathbf{0.630 \pm 0.040}$	0.470 ± 0.100
LEM *(ours)*	0.586 ± 0.047	0.331 ± 0.035	0.422 ± 0.039	0.599 ± 0.062	$\mathbf{0.631 \pm 0.065}$
ISAM *(ours)*	0.388 ± 0.057	$\underline{0.594 \pm 0.064}$	0.470 ± 0.061	0.573 ± 0.056	$\underline{0.616 \pm 0.052}$
REM *(ours)*	0.379 ± 0.059	0.167 ± 0.036	0.232 ± 0.045	0.581 ± 0.043	0.609 ± 0.068
CEM *(ours)*	0.383 ± 0.055	0.173 ± 0.076	0.238 ± 0.042	0.571 ± 0.045	0.609 ± 0.066
GEM *(ours)*	$\mathbf{0.764 \pm 0.045}$	0.551 ± 0.046	$\underline{0.640 \pm 0.046}$	0.593 ± 0.052	0.600 ± 0.082
			OASIS3		
Real Data	0.880 ± 0.040	0.870 ± 0.030	0.880 ± 0.020	0.730 ± 0.050	0.630 ± 0.060
DiffeoGauss	0.510 ± 0.040	0.300 ± 0.040	0.380 ± 0.040	$\mathbf{0.700 \pm 0.050}$	0.410 ± 0.070
TriangDDPM	0.030 ± 0.010	0.000 ± 0.000	0.000 ± 0.000	0.540 ± 0.060	0.410 ± 0.140
TriangCFM	0.060 ± 0.010	0.000 ± 0.000	0.010 ± 0.010	0.520 ± 0.060	0.410 ± 0.140
DiffeoCFM	0.600 ± 0.050	0.350 ± 0.040	0.440 ± 0.040	$\underline{0.670 \pm 0.060}$	0.530 ± 0.070
LEM *(ours)*	$\underline{0.691 \pm 0.062}$	0.432 ± 0.052	$\underline{0.531 \pm 0.057}$	0.652 ± 0.049	0.526 ± 0.062
ISAM *(ours)*	0.302 ± 0.046	$\underline{0.492 \pm 0.048}$	0.373 ± 0.049	0.661 ± 0.046	0.515 ± 0.085
REM *(ours)*	0.603 ± 0.036	0.370 ± 0.030	0.458 ± 0.033	0.655 ± 0.050	$\underline{0.544 \pm 0.060}$
CEM *(ours)*	0.593 ± 0.038	0.355 ± 0.031	0.444 ± 0.035	0.660 ± 0.048	$\mathbf{0.553 \pm 0.061}$
GEM *(ours)*	$\mathbf{0.831 \pm 0.052}$	$\mathbf{0.582 \pm 0.068}$	$\mathbf{0.684 \pm 0.062}$	0.652 ± 0.039	0.531 ± 0.043

ABIDE, and GEM improves H on ADNI and OASIS3. With input normalization shared and all maps evaluated exactly, this control attributes these gains to the choice of spectral deformation within the generative pipeline. Together, the two controls show that the gains survive a change in normalization and persist when the spectral maps are evaluated exactly.

7 Conclusion

We presented a Riemannian framework for learning on normalized SPD matrices. Trace and Frobenius normalization place these representations in NSPD, a bounded-spectrum domain that provides a common setting for designing their geometry. The DEM family constructs this geometry through scalar spectral diffeomorphisms, yielding closed-form distances, geodesics, tangent-space operators, and weighted Fréchet means. Its five members offer different local weighting while sharing one operator construction, making geometric learning convenient to implement. Polynomial and rational approximations further enable fast forward evaluation without eigendecomposition for large matrices. Experiments in visual recognition, Riemannian classification, and fMRI generation demonstrate gains over matched logarithmic controls.

REFERENCES

Dinesh Acharya, Zhiwu Huang, Danda Pani Paudel, and Luc Van Gool. Covariance pooling for facial expression recognition. In *CVPR Workshops*, 2018.

Awad H. Al-Mohy and Nicholas J. Higham. Improved inverse scaling and squaring algorithms for the matrix logarithm. *SIAM Journal on Scientific Computing*, 34(4):C153–C169, 2012. doi: 10.1137/110852553. URL https://eprints.maths.manchester.ac.uk/1851/.

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Fast and simple calculus on tensors in the Log-Euclidean framework. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 115–122, 2005. doi: 10.1007/11566465_15.

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006. doi: 10.1002/mrm.20965.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012. doi: 10.1109/TBME.2011.2172210.

Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for SPD neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/6e69ebbfad976d4637bb4b39de261bf7-Abstract.html.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=g7ohDlTITL.

Ziheng Chen, Yue Song, Yunmei Liu, and Nicu Sebe. A Lie group approach to Riemannian batch normalization. In *International Conference on Learning Representations*, 2024a. URL https://openreview.net/forum?id=okYdj8Ysru.

Ziheng Chen, Yue Song, Tianyang Xu, Zhiwu Huang, Xiao-Jun Wu, and Nicu Sebe. Adaptive Log-Euclidean metrics for SPD matrix learning. *IEEE Transactions on Image Processing*, 33: 5194–5205, 2024b. doi: 10.1109/TIP.2024.3451930. URL https://arxiv.org/abs/2303.15477.

Ziheng Chen, Yue Song, Xiaojun Wu, Gaowen Liu, and Nicu Sebe. Understanding matrix function normalizations in covariance pooling through the lens of Riemannian geometry. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=q1t0Lmvhty.

Ziheng Chen, Xiaojun Wu, Bernhard Schölkopf, and Nicu Sebe. Riemannian networks over full-rank correlation matrices. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2605.19073.

Antoine Collas, Ce Ju, Nicolas Salvy, and Bertrand Thirion. Riemannian flow matching for brain connectivity matrices via pullback geometry. In *Advances in Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=NY3LzmUXl7.

Luca Dodero, Ha Quang Minh, Marco San Biagio, Vittorio Murino, and Diego Sona. Kernel-based classification for brain connectivity graphs on the Riemannian manifold of positive definite matrices. In *2015 IEEE 12th International Symposium on Biomedical Imaging*, pp. 42–45, 2015. doi: 10.1109/ISBI.2015.7163812.

Mehrtash Harandi, Mathieu Salzmann, and Richard Hartley. Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):48–62, 2018. doi: 10.1109/TPAMI.2017.2655048. URL https://arxiv.org/abs/1605.06182.

Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, 2008. doi: 10.1137/1.9780898717778. URL https://doi.org/10.1137/1.9780898717778.

Lei Huang, Yi Zhou, Fan Zhu, Li Liu, and Ling Shao. Iterative normalization: Beyond standardization towards efficient whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4869–4878, 2019. doi: 10.1109/CVPR.2019.00501.

Zhiwu Huang and Luc Van Gool. A Riemannian network for SPD matrix learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017. doi: 10.1609/aaai.v31i1.10866.

Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xianqiu Li, and Xilin Chen. Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 720–729, 2015. URL https://proceedings.mlr.press/v37/huanga15.html.

Ioana Ilea, Lionel Bombrun, Salem Said, and Yannick Berthoumieu. Fisher vector coding for covariance matrix descriptors based on the log-euclidean and affine invariant Riemannian metrics. *Journal of Imaging*, 4(7):85, 2018. doi: 10.3390/jimaging4070085.

Catalin Ionescu, Orestis Vantzos, and Cristian Sminchisescu. Matrix backpropagation for deep networks with structured layers. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2965–2973, 2015. doi: 10.1109/ICCV.2015.339. URL https://openaccess.thecvf.com/content_iccv_2015/html/Ionescu_Matrix_Backpropagation_for_ICCV_2015_paper.html.

Elias Jarlebring, Jorge Sastre, and J. Javier Ibáñez. Polynomial approximations for the matrix logarithm with computation graphs. *Linear Algebra and its Applications*, 721:692–714, 2025. doi: 10.1016/j.laa.2024.10.024. URL https://arxiv.org/abs/2401.10089.

Sadeep Jayasumana, Richard Hartley, Mathieu Salzmann, Hongdong Li, and Mehrtash Harandi. Kernel methods on Riemannian manifolds with Gaussian RBF kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(12):2464–2477, 2015. doi: 10.1109/TPAMI.2015.2414422. URL https://arxiv.org/abs/1412.0265.

Isay Katsman, Eric Chen, Sidhanth Holalkere, Anna Asch, Aaron Lou, Ser Nam Lim, and Christopher M. De Sa. Riemannian residual neural networks. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c868aa7437dc9b29e674cd2e25689021-Abstract-Conference.html.

Peihua Li, Jiangtao Xie, Qilong Wang, and Wangmeng Zuo. Is second-order information helpful for large-scale visual recognition? In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2070–2078, 2017. URL https://openaccess.thecvf.com/content_iccv_2017/html/Li_Is_Second-Order_Information_ICCV_2017_paper.html.

Peihua Li, Jiangtao Xie, Qilong Wang, and Zilin Gao. Towards faster training of global covariance pooling networks by iterative matrix square root normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 947–955, 2018. doi: 10.1109/CVPR.2018.00105. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Li_Towards_Faster_Training_CVPR_2018_paper.html.

Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. Bilinear CNN models for fine-grained visual recognition. In *ICCV*, pp. 1449–1457, 2015.

Zhenhua Lin. Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM Journal on Matrix Analysis and Applications*, 40(4):1353–1370, 2019. doi: 10.1137/18M1221084.

James R. Munkres. *Topology*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 2000. ISBN 0-13-181629-2.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. MAtt: A manifold attention network for EEG decoding. In *Advances in Neural Information Processing Systems*, volume 35, pp. 31116–31129, 2022. doi: 10.52202/068431-2256.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. doi: 10.1007/s11263-005-3222-z.

Md Rifat Ur Rahman, Md Raihan Khan, Md Sakib Hossain Shovon, Pietro Liò, and Mohammad Ali Moni. Orthogonal polynomial approximation for matrix log normalization in global covariance pooling, 2026. URL https://arxiv.org/abs/2608.19021.

H. Ramoser, J. M"uller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446, 2000. doi: 10.1109/86.895946.

Wojciech Samek, Carmen Vidaurre, Klaus-Robert M"uller, and Motoaki Kawanabe. Stationary common spatial patterns for brain–computer interfacing. *Journal of Neural Engineering*, 9(2): 026013, 2012. doi: 10.1088/1741-2560/9/2/026013.

Jorge Sastre. Efficient mixed rational and polynomial approximation of matrix functions. *Applied Mathematics and Computation*, 218(24):11938–11946, 2012. doi: 10.1016/j.amc.2012.05.064.

Yue Song, Nicu Sebe, and Wei Wang. Why approximate matrix square root outperforms accurate SVD in global covariance pooling? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1115–1123, 2021. doi: 10.1109/ICCV48922.2021.00115. URL https://openaccess.thecvf.com/content/ICCV2021/html/Song_Why_Approximate_Matrix_Square_Root_Outperforms_Accurate_SVD_in_Global_ICCV_2021_paper.html.

Yue Song, Nicu Sebe, and Wei Wang. On the eigenvalues of global covariance pooling for fine-grained visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022a. doi: 10.1109/TPAMI.2022.3178802. URL https://arxiv.org/abs/2205.13282.

Yue Song, Nicu Sebe, and Wei Wang. Fast differentiable matrix square root. In *International Conference on Learning Representations*, 2022b. URL https://openreview.net/forum?id=-AOEi-5VTU8.

Yann Thanwerdas and Xavier Pennec. Theoretically and computationally convenient geometries on full-rank correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 43(4):1851–1872, 2022. doi: 10.1137/22M1471729. URL https://arxiv.org/abs/2201.06282.

Yann Thanwerdas and Xavier Pennec. $O(n)$-invariant Riemannian metrics on SPD matrices. *Linear Algebra and its Applications*, 661:163–201, 2023. doi: 10.1016/j.laa.2022.12.009. URL https://arxiv.org/abs/2109.05768.

Lloyd N. Trefethen. *Approximation Theory and Approximation Practice*. Society for Industrial and Applied Mathematics, 2013. doi: 10.1137/1.9781611972405.

Qilong Wang, Jiangtao Xie, Wangmeng Zuo, Lei Zhang, and Peihua Li. Deep CNNs meet global covariance pooling: Better representation and generalization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(8):2582–2597, 2021. doi: 10.1109/TPAMI.2020.2974833.

Kisung You and Hae-Jeong Park. Re-visiting Riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity. *NeuroImage*, 225:117464, 2021. doi: 10.1016/j.neuroimage.2020.117464.

A Additional Geometric Background

A.1 Differentials, Pullbacks, and Isometries

Let $f : \mathcal{M} \to \mathcal{N}$ be a smooth map between smooth manifolds. At each $p \in \mathcal{M}$, its differential is the linear map

$$Df_p : T_p\mathcal{M} \to T_{f(p)}\mathcal{N}. \tag{25}$$

If f is smooth, bijective, and has a smooth inverse, then f is a diffeomorphism. Given a Riemannian metric $g^{\mathcal{N}}$ on $\mathcal{N}$, the pullback tensor is

$$\left(f^* g^{\mathcal{N}}\right)_p (V, W) = g^{\mathcal{N}}_{f(p)} \left(Df_p[V], Df_p[W]\right). \tag{26}$$

When f is a diffeomorphism, this tensor is a Riemannian metric on $\mathcal{M}$, and f is an isometry between $\left(\mathcal{M}, f^* g^{\mathcal{N}}\right)$ and $\left(\mathcal{N}, g^{\mathcal{N}}\right)$. Thus the geometry on $\mathcal{M}$ can be computed in $\mathcal{N}$ and transferred back through f^{-1}.

A.2 Euclidean and Riemannian Operations

The isometry induced by a diffeomorphism turns standard Euclidean operations into their Riemannian counterparts. Table 5 summarizes the correspondence used in the main text Chen et al. (2024b).

Table 5: Correspondence between Euclidean and Riemannian operations under a pullback isometry.

Euclidean operation	Riemannian counterpart
Straight line	Geodesic
Difference $y - x$	Logarithmic map $\mathrm{Log}_x(y)$
Addition $x + v$	Exponential map $\mathrm{Exp}_x(v)$
Fixed tangent vector	Parallel transport $\Gamma_{x \to y}(v)$

A.3 Pullback Metric Induced by the Spectral Map

Let $f : \mathrm{NSPD}(n) \to \mathrm{Sym}(n)$ be the global spectral diffeomorphism in Theorem 4.1, and let g^f be the pullback metric in Equation (6).

Proof of Theorem 4.2. The form g^f_X is bilinear and symmetric because the Frobenius inner product is bilinear and symmetric. Since f is a diffeomorphism by Theorem 4.1, its differential Df_X is a linear isomorphism. Hence, for every nonzero $V \in T_X \,\mathrm{NSPD}(n)$, we have $Df_X[V] \neq 0$ and

$$g^f_X(V, V) = \|Df_X[V]\|_F^2 > 0. \tag{27}$$

Therefore, g^f is a Riemannian metric. Its definition states that f preserves tangent inner products, which proves that f is a Riemannian isometry.

An isometry preserves curve lengths. The Euclidean distance and straight segment between $f(X)$ and $f(Y)$ therefore pull back to (7) and (8). A tangent vector V at X is represented in $\mathrm{Sym}(n)$ by $Df_X[V]$. Euclidean addition followed by f^{-1} gives (9). Applying $D(f^{-1})_{f(X)}$ to the Euclidean difference gives (10), while applying $D(f^{-1})_{f(Y)}$ to the unchanged vector $Df_X[V]$ gives (11).

Transporting Euclidean addition and scalar multiplication through the diffeomorphism gives (12). All vector-space and Abelian Lie-group axioms follow from those of $\mathrm{Sym}(n)$. In particular, the identity is $0_f = f^{-1}(0)I$ and the inverse of X is $\ominus_f X = f^{-1}(-f(X))$.

Finally, set $Z = f(X)$ and $Z_i = f(X_i)$. The weighted Fréchet objective becomes

$$\sum_{i=1}^{m} w_i \|Z - Z_i\|_F^2. \tag{28}$$

It is strictly convex for positive weights and has the unique minimizer $Z = (\sum_i w_i Z_i)/(\sum_i w_i)$. Applying f^{-1} proves (13). □

Table 6: Leading asymptotic behavior of the scalar derivatives, where $\delta = 1 - x$.

Deformation	$x \to 0^+$	$\delta \to 0^+$
Logit	x^{-1}	δ^{-1}
IAS	$\frac{1}{2}x^{-3/2}$	$\frac{1}{2}\delta^{-3/2}$
Rational	x^{-2}	δ^{-2}
Cauchy	$(\pi x^2)^{-1}$	$(\pi\delta^2)^{-1}$
Gumbel	$(x\lvert\log x\rvert)^{-1}$	δ^{-1}

A.4 Geometric Properties of DEMs

Let $X = U\,\mathrm{diag}(\mu_1, \ldots, \mu_n)U^\top \in \mathrm{NSPD}(n)$. Define the Loewner divided-difference matrix by

$$(L_f(\mu))_{ij} = \begin{cases} \dfrac{f(\mu_i) - f(\mu_j)}{\mu_i - \mu_j}, & \mu_i \neq \mu_j, \\ f'(\mu_i), & \mu_i = \mu_j. \end{cases} \tag{29}$$

The differential in (14) uses the continuous extension $f'(\mu_i)$ at repeated eigenvalues.

The first four scalar deformations satisfy $f(1-x) = -f(x)$, which gives

$$f(I - X) = -f(X), \qquad d_f(I - X, I - Y) = d_f(X, Y). \tag{30}$$

Their weighted Fréchet means satisfy

$$\mathrm{FM}_f^w(I - X_1, \ldots, I - X_m) = I - \mathrm{FM}_f^w(X_1, \ldots, X_m). \tag{31}$$

Derivation of the local spectral metric. By (14) and orthogonal invariance of the Frobenius norm, for $\widetilde{V} = U^\top V U$ the pullback metric is

$$g_X^f(V, V) = \sum_{i,j} (L_f(\mu))_{ij}^2 \widetilde{V}_{ij}^2. \tag{32}$$

The matrices $L_f(\mu)$ and $\widetilde{V}$ are symmetric, so separating the diagonal and off-diagonal terms gives

$$g_X^f(V, V) = \sum_i f'(\mu_i)^2 \widetilde{V}_{ii}^2 + 2\sum_{i<j} L_{f,ij}^2 \widetilde{V}_{ij}^2. \tag{33}$$

□

Proof of complement symmetry. For Logit, IAS, Rational, and Cauchy, the scalar identity $f(1-x) = -f(x)$ applied spectrally gives $f(I - X) = -f(X)$. Substitution into (7) proves (30). The closed-form weighted mean in (13) then gives

$$\begin{aligned} \mathrm{FM}_f^w(I - X_1, \ldots, I - X_m) &= f^{-1}\left(-\frac{\sum_i w_i f(X_i)}{\sum_i w_i}\right) \\ &= I - \mathrm{FM}_f^w(X_1, \ldots, X_m), \end{aligned} \tag{34}$$

which proves (31). □

A.5 Shared Global Properties

Proof of Theorem 4.3. The global isometry f maps $\mathrm{NSPD}(n)$ onto Euclidean space $\mathrm{Sym}(n)$. Completeness, flatness, and uniqueness of minimizing geodesics follow by pulling back the corresponding Euclidean properties. □

Proof of Theorem 4.4. For $A \in \mathrm{NSPD}(n)$, write $L_A(X) = A \oplus_f X$. By (12), $f(L_A(X)) = f(A) + f(X)$, so

$$Df_{L_A(X)}[D(L_A)_X[V]] = Df_X[V]. \tag{35}$$

The pullback metric (6) therefore satisfies $g^f_{L_A(X)}(D(L_A)_X[V], D(L_A)_X[W]) = g^f_X(V, W)$. Thus left translations are isometries. Right translations coincide with left translations because $\oplus_f$ is commutative. The identity is $f^{-1}(0)I$, and $f(\ominus_f X) = -f(X)$. Differentiating the latter identity shows that inversion changes the sign of $Df_X[V]$ and hence preserves the Frobenius inner product. The distance identities follow. □

Proof of Theorem 4.5. The matrix function commutes with orthogonal conjugation. Differentiating $f(QXQ^\top) = Qf(X)Q^\top$ gives

$$Df_{QXQ^\top}[QVQ^\top] = QDf_X[V]Q^\top. \tag{36}$$

Orthogonal invariance of the Frobenius inner product in (6) proves that conjugation is an isometry. The distance and geodesic identities follow from (7) and (8). The Fréchet-mean identity follows from (13) and $f^{-1}(QZQ^\top) = Qf^{-1}(Z)Q^\top$.

□

Proof of Theorem 4.6. For $f = f_{\mathrm{Gumbel}}$ and $x \in (0, 1)$,

$$f(x^\beta) = -\log(-\beta \log x) = f(x) - \log \beta. \tag{37}$$

Applying this identity to the eigenvalues of X yields $f(X^\beta) = f(X) - (\log \beta)I$. Also, $f(e^{-\beta}I) = -(\log \beta)I$, which gives (18). The power map is a diffeomorphism of $\mathrm{NSPD}(n)$ with inverse $X \mapsto X^{1/\beta}$, so Theorem 4.4 gives its distance invariance.

For every $Z \in \mathrm{Sym}(n)$, the same scalar identity gives

$$f^{-1}\left(Z - (\log \beta)I\right) = \left(f^{-1}(Z)\right)^\beta. \tag{38}$$

Substitution into (8) proves geodesic equivariance. For the weighted mean, the weights in (13) sum to one after normalization. Thus powering every input subtracts $(\log \beta)I$ from the weighted average of its images under f. Applying the displayed inverse identity proves the Fréchet-mean formula. □

B Additional Method Details

B.1 Additional Normalization Details

The Frobenius normalization in (2) is invariant to positive scaling and orthogonal conjugation.

Proposition B.1 (Spectral boundedness and scale invariance)**.** *For $A \in \mathrm{SPD}(n)$, $n \geq 2$, and $c > 0$, the matrix $X = A/\|A\|_F$ belongs to $\mathrm{NSPD}(n)$ and satisfies $cA/\|cA\|_F = X$.*

Proof. All eigenvalues of A are positive, so $\|A\|_F^2 = \sum_i \lambda_i(A)^2 > \lambda_{\max}(A)^2$ when $n \geq 2$. Thus $0 \prec X \prec I$. The scale identity follows from $\|cA\|_F = c\|A\|_F$. □

Proposition B.2 (Normalized section)**.** *The image of Frobenius normalization is $\{X \in \mathrm{SPD}(n) \mid \|X\|_F = 1\}$, a proper subset of $\mathrm{NSPD}(n)$. Every positive scaling ray intersects this subset once.*

Proof. For $X \in \mathrm{SPD}(n)$, $\|X/\|X\|_F\|_F = 1$. Conversely, every SPD matrix of unit Frobenius norm is unchanged by normalization. The unique unit-norm point on $\{cX \mid c > 0\}$ has $c = 1/\|X\|_F$. The image is a proper subset because $\mathrm{NSPD}(n)$ also contains SPD matrices with other Frobenius norms. □

For a symmetric perturbation H, the differential of Frobenius normalization is

$$D\left(\frac{A}{\|A\|_F}\right)[H] = \frac{H}{\|A\|_F} - \frac{A\langle A, H\rangle_F}{\|A\|_F^3}. \tag{39}$$

The numerical implementations evaluate the spectral transformation at the normalized matrix. Trace normalization is used in the fMRI experiment and the normalization ablation.

B.2 Scalar Diffeomorphism Analysis

The candidate functions in Table 1 have derivatives

$$f'_{\mathrm{Logit}}(x) = \frac{1}{x(1-x)}, \qquad f'_{\mathrm{IAS}}(x) = \frac{1}{2[x(1-x)]^{3/2}},$$
$$f'_{\mathrm{Rational}}(x) = \frac{1}{x^2} + \frac{1}{(1-x)^2}, \qquad f'_{\mathrm{Cauchy}}(x) = \frac{\pi}{\cos^2(\pi(x-\frac{1}{2}))},$$
$$f'_{\mathrm{Gumbel}}(x) = -\frac{1}{x\log x}. \tag{40}$$

Every expression in (40) is positive on $(0,1)$. Each function is smooth on this interval and has limits $-\infty$ and $+\infty$ at the lower and upper endpoints, respectively. The inverse function theorem and the endpoint limits therefore show that every candidate is a smooth increasing diffeomorphism from $(0,1)$ onto $\mathbb{R}$.

Their leading endpoint behavior is summarized below. Let $\delta = 1 - x \to 0^+$.

$$\begin{array}{ll} f_{\mathrm{Logit}}(x) \sim \log x & f_{\mathrm{Logit}}(1-\delta) \sim -\log\delta, \\ f_{\mathrm{IAS}}(x) \sim -x^{-1/2} & f_{\mathrm{IAS}}(1-\delta) \sim \delta^{-1/2}, \\ f_{\mathrm{Rational}}(x) \sim -x^{-1} & f_{\mathrm{Rational}}(1-\delta) \sim \delta^{-1}, \\ f_{\mathrm{Cauchy}}(x) \sim -(\pi x)^{-1} & f_{\mathrm{Cauchy}}(1-\delta) \sim (\pi\delta)^{-1}, \\ f_{\mathrm{Gumbel}}(x) \sim -\log|\log x| & f_{\mathrm{Gumbel}}(1-\delta) \sim -\log\delta. \end{array} \tag{41}$$

The corresponding derivative growth is

$$\begin{array}{ll} f'_{\mathrm{Logit}}(x) \sim x^{-1} & f'_{\mathrm{Logit}}(1-\delta) \sim \delta^{-1}, \\ f'_{\mathrm{IAS}}(x) \sim \frac{1}{2}x^{-3/2} & f'_{\mathrm{IAS}}(1-\delta) \sim \frac{1}{2}\delta^{-3/2}, \\ f'_{\mathrm{Rational}}(x) \sim x^{-2} & f'_{\mathrm{Rational}}(1-\delta) \sim \delta^{-2}, \\ f'_{\mathrm{Cauchy}}(x) \sim (\pi x^2)^{-1} & f'_{\mathrm{Cauchy}}(1-\delta) \sim (\pi\delta^2)^{-1}, \\ f'_{\mathrm{Gumbel}}(x) \sim \frac{1}{x|\log x|} & f'_{\mathrm{Gumbel}}(1-\delta) \sim \delta^{-1}. \end{array} \tag{42}$$

The first four functions satisfy $f(1-x) = -f(x)$. Their zeros and minimum derivative points occur at $x = 1/2$, and their minimum derivative values are 4, 4, 8, and π, respectively. Gumbel is not antisymmetric. Its zero and derivative minimum both occur at $x = e^{-1}$, where its minimum derivative equals e.

B.3 Proof of the Global Matrix Diffeomorphism

Proof of Theorem 4.1. The expression in (3) is independent of the selected eigenbasis because the same scalar function is applied to every eigenvalue. On a repeated eigenspace with eigenvalue λ, the output is $f(\lambda)$ times the identity, so changing the orthonormal basis within that eigenspace leaves the matrix unchanged.

Because f is a bijection from $(0,1)$ onto $\mathbb{R}$, applying it to each eigenvalue maps $\mathrm{NSPD}(n)$ bijectively onto $\mathrm{Sym}(n)$. Applying f^{-1} spectrally gives (4). Both maps are smooth by the standard smoothness of primary matrix functions generated by smooth scalar functions on open spectral domains Higham (2008). Hence f is a global diffeomorphism. □

The Frobenius-normalized subset has one fewer degree of freedom than $\mathrm{NSPD}(n)$. The global diffeomorphisms are defined on the full domain, rather than only on the normalized subset.

B.4 Coefficient Construction for Scalable Computation

This appendix gives the coefficient details used by the decomposition-free matrix-function approximations in Section 5. Let $\sigma(B) \subset [a,b] \subset (0,1)$ and define

$$M(x) = \frac{2x-(a+b)}{b-a}, \qquad g(t) = \frac{a+b}{2} + \frac{b-a}{2}t, \qquad \widetilde{f}(t) = f(g(t)). \tag{43}$$

Weighted projection. For a basis $\{q_k\}_{k=0}^{K}$ and a positive weight w on $[-1, 1]$, define $\langle u, v\rangle_w = \int_{-1}^{1} u(t)v(t)w(t)dt$. The coefficients of the degree-K projection satisfy

$$Gc = \beta, \qquad G_{ij} = \langle q_i, q_j\rangle_w, \qquad \beta_i = \langle \widetilde{f}, q_i\rangle_w. \tag{44}$$

If the basis is orthogonal under w, then $c_k = \langle \widetilde{f}, q_k\rangle_w / \langle q_k, q_k\rangle_w$. This orthogonal-polynomial viewpoint follows standard approximation theory Trefethen (2013). The coefficients depend only on the scalar map, interval, and approximation degree, so they are precomputed before training.

Chebyshev coefficients. For $q_k = T_k$ and $w(t) = (1 - t^2)^{-1/2}$, the projection coefficients are

$$c_k = \frac{2 - \mathbf{1}_{\{k=0\}}}{\pi} \int_0^{\pi} \widetilde{f}(\cos\theta)\cos(k\theta)d\theta. \tag{45}$$

They can be evaluated by one-dimensional quadrature or from Chebyshev nodes $t_j = \cos(\pi(j + 1/2)/(K + 1))$. The basis matrices are generated from $T_0(S) = I$, $T_1(S) = S$, and $T_{k+1}(S) = 2ST_k(S) - T_{k-1}(S)$.

Legendre coefficients. For $q_k = P_k$ and the uniform weight, orthogonality gives

$$c_k = \frac{2k + 1}{2} \int_{-1}^{1} \widetilde{f}(t)P_k(t)dt. \tag{46}$$

The integral is evaluated by Gauss–Legendre quadrature. Matrix evaluation uses $P_0(S) = I$, $P_1(S) = S$, and $(k + 1)P_{k+1}(S) = (2k + 1)SP_k(S) - kP_{k-1}(S)$.

Taylor and Padé instances. Let $c = (a + b)/2$ and $Y = B - cI$. The degree-K Taylor instance is $\sum_{k=0}^{K} d_k Y^k$, where $d_k = f^{(k)}(c)/k!$. The $[m/n]$ Padé instance has the form

$$R_{m,n}(Y) = N_m(Y)D_n(Y)^{-1}. \tag{47}$$

Here, N_m and D_n are the numerator and denominator polynomials of degrees m and n. Their coefficients are obtained by matching the scalar series of $f(c+y)$ through order $m+n$. In implementation, $D_n(Y)^{-1}$ is not formed explicitly. The output is obtained by solving the corresponding linear system.

Laguerre instance. The Laguerre basis is generated by $L_0(B) = I$, $L_1(B) = I - B$, and

$$(k + 1)L_{k+1}(B) = ((2k + 1)I - B)\, L_k(B) - kL_{k-1}(B). \tag{48}$$

For a target interval $[a, b]$, its coefficients are obtained from the same projection system in (44) after restricting the basis to that interval. Once the coefficients are fixed, all polynomial instances are evaluated as finite sums of recursively generated matrix basis terms.

B.5 Numerical Evaluation of the Approximations

We compare five approximation families against exact spectral evaluation using Frobenius-normalized synthetic SPD matrices, batch size 32, fitting interval $[0.05, 0.95]$, degree $K = 8$, and FP64 on one NVIDIA A100 GPU (80 GB). Figure 6 shows $n = 1024$ results. Table 7 reports relative Frobenius errors across four dimensions, including the logarithmic reference.

Across all four dimensions, Chebyshev gives the lowest errors for LEM, ISAM, and GEM, while Padé leads for REM and CEM. Padé keeps REM errors below 4.5×10^{-6} and CEM errors below 10^{-3}, supporting accurate evaluation as matrix size increases in this benchmark.

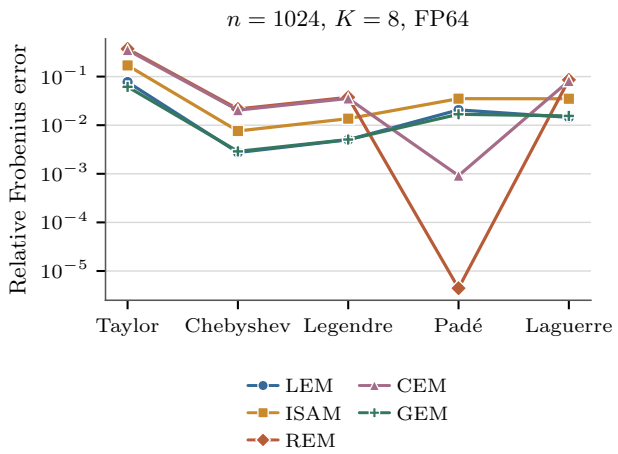


Figure 6: Relative FP64 Frobenius errors for the five DEMs at $n = 1024$ and $K = 8$. Lower is better.

Table 7: Relative FP64 Frobenius errors reported for the synthetic-spectrum experiment. Each row uses Frobenius normalization, fitting interval $[0.05, 0.95]$, and degree $K = 8$. The log row is included as the reference, while the remaining rows are the five DEMs.

n	Metric	Taylor	Chebyshev	Legendre	Padé	Laguerre
128	log	6.056e-02	2.120e-03	3.638e-03	3.242e-03	1.027e-02
128	LEM	6.895e-02	2.854e-03	4.748e-03	1.847e-02	1.397e-02
128	ISAM	1.535e-01	8.088e-03	1.305e-02	3.180e-02	3.394e-02
128	REM	3.431e-01	2.426e-02	3.671e-02	4.185e-06	8.470e-02
128	CEM	3.242e-01	2.275e-02	3.469e-02	8.417e-04	8.004e-02
128	GEM	5.541e-02	3.077e-03	4.830e-03	1.509e-02	1.548e-02
256	log	6.292e-02	2.021e-03	3.612e-03	3.348e-03	1.062e-02
256	LEM	7.134e-02	2.650e-03	4.734e-03	1.906e-02	1.434e-02
256	ISAM	1.584e-01	7.445e-03	1.294e-02	3.274e-02	3.481e-02
256	REM	3.527e-01	2.164e-02	3.592e-02	4.263e-06	8.674e-02
256	CEM	3.335e-01	2.043e-02	3.402e-02	8.646e-04	8.206e-02
256	GEM	5.715e-02	2.790e-03	4.860e-03	1.553e-02	1.572e-02
512	log	6.560e-02	1.985e-03	3.745e-03	3.477e-03	1.087e-02
512	LEM	7.412e-02	2.649e-03	4.890e-03	1.979e-02	1.454e-02
512	ISAM	1.639e-01	7.377e-03	1.334e-02	3.385e-02	3.519e-02
512	REM	3.623e-01	2.110e-02	3.694e-02	4.342e-06	8.703e-02
512	CEM	3.430e-01	1.996e-02	3.501e-02	8.897e-04	8.246e-02
512	GEM	5.925e-02	2.822e-03	4.993e-03	1.609e-02	1.571e-02
1024	log	6.852e-02	2.062e-03	3.863e-03	3.633e-03	1.094e-02
1024	LEM	7.719e-02	2.740e-03	5.011e-03	2.065e-02	1.450e-02
1024	ISAM	1.697e-01	7.600e-03	1.362e-02	3.514e-02	3.497e-02
1024	REM	3.718e-01	2.156e-02	3.744e-02	4.424e-06	8.568e-02
1024	CEM	3.525e-01	2.042e-02	3.553e-02	9.169e-04	8.131e-02
1024	GEM	6.158e-02	2.891e-03	5.067e-03	1.677e-02	1.550e-02

At $n = 1024$ in FP64, median forward and backward times total approximately 109 ms for the representative Chebyshev implementation and 115 ms for Padé, compared with 535–547 ms across repeated exact-log runs. The lower costs characterize these matrix-arithmetic implementations in this benchmark rather than matched speedups for individual DEMs.

C Forward and Backward Algorithms for NSPD Matrix Functions

Let $A \in \mathrm{SPD}(n)$, with $n \geq 2$. Set

$$B = \frac{A}{\|A\|_F}, \qquad Z = f(B), \qquad B \in \mathrm{NSPD}(n),$$

where f is one of the scalar diffeomorphisms $\{f_{\mathrm{Logit}}, f_{\mathrm{IAS}}, f_{\mathrm{Rational}}, f_{\mathrm{Cauchy}}, f_{\mathrm{Gumbel}}\}$. The optional log baseline is obtained by adding $f_{\log}(x) = \log x$ to the same interface. All coefficients are computed offline in CPU double precision. The forward pass only evaluates matrix products,

additions, and (for Padé) one linear solve. The exact spectral backward pass uses the standard matrix-function backpropagation pattern with Loewner divided differences Ionescu et al. (2015); Higham (2008). For polynomial and rational evaluation, let $[a, b] \subset (0, 1)$ denote the coefficient-fitting interval. Its endpoints are supplied with the cached coefficients and do not modify the normalized input B.

C.1 Shared pre-normalization

Algorithm 1 PRE-NORMALIZATION for NSPD matrix functions

Require: $A \in \mathrm{SPD}(n)$, $n \geq 2$
Ensure: $B \in \mathrm{NSPD}(n)$ with $\|B\|_F = 1$
1: $t \leftarrow \|A\|_F$
2: $B \leftarrow A/t$
3: **return** B

For the trace-normalization ablation, replace $\|A\|_F$ by $\mathrm{tr}(A)$ in Algorithm 1. The five matrix-function evaluators below remain unchanged.

C.2 Taylor approximation

Let $c = (a + b)/2$ and let $a_k^{(f)}$ be the Taylor coefficients of f at c. The approximation is $f(B) \approx \sum_{k=0}^{K} a_k^{(f)} (B - cI)^k$.

Algorithm 2 NSPD TAYLOR FORWARD

Require: A, degree K, fitting interval $[a, b]$, cached $\{a_k^{(f)}\}_{k=0}^{K}$
1: $B \leftarrow$ PRE-NORMALIZATION(A)
2: $c \leftarrow (a + b)/2$, $Y \leftarrow B - cI$, $P \leftarrow I$
3: $Z \leftarrow a_0^{(f)} I$
4: **for** $k = 1$ to K **do**
5: $\quad P \leftarrow PY$
6: $\quad Z \leftarrow Z + a_k^{(f)} P$
7: **end for**
8: **return** Z

C.3 Chebyshev approximation

Define $S = (2B - (a + b)I)/(b - a)$ so that the fitting interval maps to $[-1, 1]$. With cached Chebyshev coefficients $\{c_k^{(f)}\}_{k=0}^{K}$, use the first-kind recurrence.

Algorithm 3 NSPD CHEBYSHEV FORWARD (recommended)

Require: A, degree K, fitting interval $[a, b]$, cached $\{c_k^{(f)}\}_{k=0}^{K}$
1: $B \leftarrow$ PRE-NORMALIZATION(A)
2: $S \leftarrow (2B - (a + b)I)/(b - a)$
3: $T_0 \leftarrow I$, $T_1 \leftarrow S$
4: $Z \leftarrow c_0^{(f)} T_0 + c_1^{(f)} T_1$
5: **for** $k = 1$ to $K - 1$ **do**
6: $\quad T_{k+1} \leftarrow 2ST_k - T_{k-1}$
7: $\quad Z \leftarrow Z + c_{k+1}^{(f)} T_{k+1}$
8: **end for**
9: **return** Z

C.4 Legendre approximation

Using the same S, let $\{\ell_k^{(f)}\}$ be the Gauss–Legendre projection coefficients.

Algorithm 4 NSPD LEGENDRE FORWARD

Require: A, degree K, fitting interval $[a, b]$, cached $\{\ell_k^{(f)}\}_{k=0}^K$
1: $B \leftarrow$ PRE-NORMALIZATION(A)
2: $S \leftarrow (2B - (a+b)I)/(b-a)$
3: $P_0 \leftarrow I$, $P_1 \leftarrow S$
4: $Z \leftarrow \ell_0^{(f)} P_0 + \ell_1^{(f)} P_1$
5: **for** $k = 1$ to $K-1$ **do**
6: $\quad P_{k+1} \leftarrow ((2k+1)SP_k - kP_{k-1})/(k+1)$
7: $\quad Z \leftarrow Z + \ell_{k+1}^{(f)} P_{k+1}$
8: **end for**
9: **return** Z

C.5 FINITE-INTERVAL LAGUERRE APPROXIMATION

Define $U = (B - aI)/(b-a)$. The finite-interval Laguerre coefficients approximate $u \mapsto f(a + (b-a)u)$ on $u \in [0, 1]$.

Algorithm 5 NSPD LAGUERRE FORWARD

Require: A, degree K, fitting interval $[a, b]$, cached $\{g_k^{(f)}\}_{k=0}^K$
1: $B \leftarrow$ PRE-NORMALIZATION(A)
2: $U \leftarrow (B - aI)/(b-a)$
3: $L_0 \leftarrow I$, $L_1 \leftarrow I - U$
4: $Z \leftarrow g_0^{(f)} L_0 + g_1^{(f)} L_1$
5: **for** $k = 1$ to $K-1$ **do**
6: $\quad L_{k+1} \leftarrow (((2k+1)I - U)L_k - kL_{k-1})/(k+1)$
7: $\quad Z \leftarrow Z + g_{k+1}^{(f)} L_{k+1}$
8: **end for**
9: **return** Z

C.6 PADÉ APPROXIMATION

For Padé, $p_i^{(f)}$ and $q_j^{(f)}$ are obtained by matching the Taylor series of $f(c+y)$, with $c = (a+b)/2$. This is a rational approximation and therefore requires a stability check on the denominator.

Algorithm 6 NSPD PADÉ FORWARD ($[m/n]$)

Require: A, orders m, n, fitting interval $[a, b]$, cached $\{p_i^{(f)}\}$ and $\{q_j^{(f)}\}$
1: $B \leftarrow$ PRE-NORMALIZATION(A)
2: $c \leftarrow (a+b)/2$, $Y \leftarrow B - cI$, $P_0 \leftarrow I$
3: **for** $k = 1$ to $\max(m, n)$ **do**
4: $\quad P_k \leftarrow P_{k-1} Y$
5: **end for**
6: $N \leftarrow \sum_{i=0}^m p_i^{(f)} P_i$
7: $D \leftarrow \sum_{j=0}^n q_j^{(f)} P_j$
8: Solve $DZ = N$, with D nonsingular
9: **return** Z

C.7 EXACT SPECTRAL FORWARD PASS

The non-approximated route uses one eigendecomposition. Let λ_i and u_i denote the eigenvalues and eigenvectors of the normalized matrix B. For an SPD input with $n \geq 2$, Frobenius normalization gives $\sigma(B) \subset (0, 1)$. The scalar map is evaluated directly at these eigenvalues.

Algorithm 7 EXACT SPECTRAL f FORWARD

Require: $A \in \mathrm{SPD}(n)$, $n \geq 2$, scalar map f
$B \leftarrow$ PRE-NORMALIZATION(A)
$(\lambda, U) \leftarrow \mathrm{eigh}(B)$
$v_i \leftarrow f(\lambda_i)$
$Z \leftarrow U\,\mathrm{diag}(v)U^{\mathsf{T}}$
return Z

For the exact route, the eigendecomposition is used for the forward value, but its generic automatic backward is not used. The following custom backward evaluates the Fréchet derivative through the Loewner matrix.

C.8 EXACT SPECTRAL BACKWARD PASS WITH LOEWNER DIVIDED DIFFERENCES

Let $G = \partial\mathcal{L}/\partial Z$ be the upstream gradient. Define

$$L_{ij} = \begin{cases} \dfrac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}, & \lambda_i \neq \lambda_j, \\ f'(\lambda_i), & \lambda_i = \lambda_j. \end{cases}$$

For numerically close eigenvalues, the second branch is used with the derivative-limit approximation $\frac{1}{2}(f'(\lambda_i) + f'(\lambda_j))$. This removes the $1/(\lambda_i - \lambda_j)$ singularity of the generic eigendecomposition backward.

Algorithm 8 EXACT SPECTRAL f BACKWARD

Require: Upstream gradient G, saved (λ, U), scalar derivative f'
$G_s \leftarrow \frac{1}{2}(G + G^{\mathsf{T}})$
$H \leftarrow U^{\mathsf{T}} G_s U$
for all i, j **do**
  if $|\lambda_i - \lambda_j|$ is numerically large **then**
    $L_{ij} \leftarrow (f(\lambda_i) - f(\lambda_j))/(\lambda_i - \lambda_j)$
  else
    $L_{ij} \leftarrow \frac{1}{2}(f'(\lambda_i) + f'(\lambda_j))$
  end if
end for
$\widetilde{H} \leftarrow L \odot H$
$G_B \leftarrow U\widetilde{H}U^{\mathsf{T}}$
$G_B \leftarrow \frac{1}{2}(G_B + G_B^{\mathsf{T}})$
Backpropagate G_B through $B =$ PRE-NORMALIZATION(A)
return G_A

If the implementation also returns an auxiliary eigenvalue vector, an upstream gradient g_λ for that vector is added before the final rotation as

$$\widetilde{H} \leftarrow \widetilde{H} + \mathrm{diag}(g_\lambda).$$

For the ordinary matrix logarithm, replace f by $\log$, $f'(x)$ by $1/x$, and use the corresponding logarithmic divided difference. The same Loewner construction applies to the ordinary square root, with $f'(x) = 1/(2\sqrt{x})$.

C.9 SCALAR DERIVATIVES

The derivatives needed by Algorithm 8 are

Table 8: Scalar diffeomorphisms and derivatives.

Map	$f(x)$	$f'(x)$
Logit	$\log(x/(1-x))$	$1/(x(1-x))$
IAS	$(2x-1)/\sqrt{x(1-x)}$	$1/[2(x(1-x))^{3/2}]$
Rational	$(2x-1)/(x(1-x))$	$1/x^2 + 1/(1-x)^2$
Cauchy	$\tan(\pi(x-1/2))$	$\pi/\cos^2(\pi(x-1/2))$
Gumbel	$-\log(-\log x)$	$-1/(x\log x)$

C.10 Forward cost summary

Table 9: Forward cost of the five approximation families for $f(B)$.

Family	Forward computation
Taylor	K matrix multiplications
Chebyshev	K matrix multiplications
Legendre	K matrix multiplications
Laguerre	K matrix multiplications
Padé	$\max(m,n)$ matrix multiplications plus one linear solve
Exact spectral	One eigendecomposition plus spectral reconstruction

The five approximation algorithms share the same forward interface. The choice of f changes only the offline coefficients and, for the exact route, the scalar derivative used in the Loewner matrix. Thus the paper can report one algorithm per approximation family while evaluating all five choices of f under the same implementation.

D Additional Experimental Details

D.1 Classification Protocols

Fine-grained recognition. We fine-tune ImageNet-pretrained ResNet-50, ResNet-101, and Swin-T backbones on CUB-200, Stanford Cars, and FGVC-Aircraft. Every configuration uses the same global covariance-pooling pipeline with 256×256 inputs and a 256-dimensional covariance representation. Backbone features are projected to 256 channels, batch-normalized, and converted to an FP64 covariance matrix. We use Frobenius normalization, $A/\|A\|_F$, in the main classification experiments, followed by either the exact spectral map or its degree-8 matrix-function approximation. The upper-triangular entries are vectorized and passed through LayerNorm and a linear classifier. We train with AdamW, learning rate and weight decay 10^{-4}, cosine annealing, 100 epochs, and batch size 32. Each result is averaged over seeds 0, 1, and 2. The main result figure reports the means, while the complete seed-level records and standard deviations are retained with the experiment logs.

ImageNet-1K. We train ResNet-50 from scratch for 110 epochs, comprising 10 warmup epochs and 100 cosine-annealing epochs. We use SGD with momentum, global batch size 256, 256×256 inputs, a 256-dimensional covariance representation, and batch size 32 per replica. The main DEM comparisons use Frobenius normalization. Additional log-mean and log-trace controls are reported separately below. All ImageNet results use seed 0.

Table 10: Additional ImageNet-1K log controls. Each entry reports Top-1 and Top-5 accuracy (%). The log-mean Taylor run diverged and is omitted.

Variant	Taylor	Chebyshev	Legendre	Padé	Laguerre
log-mean	–	75.25 / 92.71	75.48 / 92.83	77.91 / 93.79	78.30 / 94.18
log-trace	76.11 / 92.41	76.64 / 92.86	76.48 / 92.70	76.34 / 92.69	76.39 / 92.74

D.2 Additional Protocols

Riemannian ResNet and BatchNorm. The FPHA experiment uses a learning rate selected for each method. HINSS cross-session and cross-subject use learning rate 10^{-3} over 15 folds. The input uses the Frobenius normalization specified in the main text. Cora uses the classifier-representation replacement protocol with three seeds.

fMRI connectivity generation. We use DiffeoCFM on ABIDE, ADNI, and OASIS3 with 10 subject-group splits, seed 42, and 200 training epochs. The DiffeoGauss, TriangDDPM, TriangCFM, and DiffeoCFM baselines and all DEM variants are rerun under this protocol. The DEM experiments evaluate the forward and inverse spectral maps exactly. For each scan, the fMRI time series is converted to an SPD connectivity matrix A, and subject-level grouping is used for the train/test split. For the DEM variants, we use $X = A/\operatorname{tr}(A)$. The normalized matrix is mapped with the exact spectral DEM and its lower-triangular entries are standardized using a StandardScaler fitted on the training split. A class-conditional DiffeoCFM flow is trained on these standardized vectors and used to sample representations for each test label. We then undo the StandardScaler, restore symmetry, and apply the exact inverse DEM map to obtain generated SPD matrices. Correlation conversion is performed only for the final quality and CAS evaluation. The representation-specific input preparation and metric definitions follow the released experiment configuration.

D.3 Ablation Studies

Normalization and approximation interaction. On Cars with pretrained ResNet-50, we compare normalization choices using degree-8 Taylor, Chebyshev, Legendre, Padé, and Laguerre approximations. The two arms use $B = A/\operatorname{tr}(A)$ and $B = A/\|A\|_F$, respectively. Table 11 averages final test Top-1 accuracy over the five families. Per-family results are retained with the experiment logs.

Table 11: Cars normalization ablation. Entries are final test Top-1 accuracy averaged over the five approximation bases.

Normalization	MLN	LEM	ISAM	REM	CEM	GEM
Trace	89.32	89.83	90.44	91.09	90.96	89.89
Frobenius	90.17	90.68	90.39	90.59	90.46	90.53

All five DEMs improve the mean accuracy over matched MLN under both normalization choices. REM gives the largest gain with trace normalization, improving accuracy by 1.77 percentage points, while LEM leads with Frobenius normalization at a gain of 0.51 points. The gains across both arms support the value of spectral deformation under either normalization in this setting.

Normalized exact-log control for fMRI. This control compares the DEMs with an exact logarithmic map under the same input normalization and ten subject-group splits used in the fMRI study. Both use $X = A/\operatorname{tr}(A)$ and exact forward and inverse spectral evaluation, so the comparison isolates the choice of deformation from input normalization and numerical approximation. Table 12 reports the log reference.

Table 12: Ten-split fMRI control using exact log after trace normalization. Entries are mean $\pm$ standard deviation.

Dataset	α	β	H	CAS-AUC	CAS-F1
ABIDE	0.8394 ± 0.0679	0.4963 ± 0.0837	0.6222 ± 0.0822	0.6379 ± 0.0551	0.5708 ± 0.0687
ADNI	0.4973 ± 0.0682	0.2866 ± 0.0472	0.3631 ± 0.0547	0.5799 ± 0.0412	0.5969 ± 0.0679
OASIS3	0.6378 ± 0.0462	0.3774 ± 0.0353	0.4741 ± 0.0403	0.6601 ± 0.0413	0.5389 ± 0.0556

The matched comparison confirms fidelity gains beyond normalization alone. Relative to this log reference, REM improves H on ABIDE by 0.0928, and GEM improves H on ADNI and OASIS3 by 0.2769 and 0.2099, respectively. LEM also gives higher CAS-F1 on ADNI, while CEM gives higher CAS-F1 on OASIS3. These comparisons identify concrete gains from deformation choice in a shared generative pipeline with exact spectral maps.

E Extended Related Work

SPD representations and learning. Second-order and covariance representations have been widely used in visual recognition Lin et al. (2015); Li et al. (2017); Acharya et al. (2018); Li et al. (2018); Wang et al. (2021), while Riemannian neural networks directly exploit the manifold structure of SPD matrices Huang & Gool (2017); Brooks et al. (2019); Katsman et al. (2023); Chen et al. (2024a). Recent work further analyzes matrix-function normalizations in covariance pooling through a Riemannian-geometric lens Chen et al. (2025). Metric learning, kernel construction, and geometry-aware dimensionality reduction provide complementary approaches for SPD-valued data Huang et al. (2015); Jayasumana et al. (2015); Harandi et al. (2018). Covariance descriptors have also been studied for visual recognition and functional-connectivity analysis Ilea et al. (2018); Dodero et al. (2015); You & Park (2021). Most of these methods operate on the full SPD cone or use a fixed matrix function as the representation map. A related line of work instead controls scale, notably through full-rank correlation matrices, where variable-wise scales are removed by fixing the diagonal Thanwerdas & Pennec (2022); Chen et al. (2026). Our normalized representation is different in that it removes only one global positive scale, thereby retaining eigendirections and eigenvalue ratios.

Riemannian geometries on SPD matrices. The affine-invariant and log-Euclidean metrics are classical approaches to SPD geometry, within the broader Riemannian framework for tensor-valued data Pennec et al. (2006); Arsigny et al. (2005; 2006); Thanwerdas & Pennec (2023). Alternative coordinate geometries, including Cholesky-based constructions, provide different computational trade-offs Lin (2019); Ilea et al. (2018). More general pullback-Euclidean constructions retain this computational advantage while allowing alternative representation maps Chen et al. (2024b). Pullback geometry has also been used for generative modeling of brain-connectivity matrices Chen & Lipman (2024); Collas et al. (2025). DEMs belong to this pullback-Euclidean family, but specialize the representation-map design to scalar spectral deformations on a normalized, bounded spectral domain.

Scalable matrix functions. Explicit spectral matrix functions can be expensive for large matrices, motivating iterative and decomposition-free methods for covariance normalization Higham (2008); Li et al. (2018); Song et al. (2021; 2022b;a). General polynomial and rational approximation is another route to decomposition-free matrix-function evaluation Trefethen (2013); Sastre (2012); Al-Mohy & Higham (2012); Jarlebring et al. (2025). Polynomial approximations have also been studied for matrix-log normalization Rahman et al. (2026). These methods provide the computational foundation for our scalable implementation, where the target matrix functions are the scalar deformations defining DEMs.