UNIVERSAL DIFFEOMORPHIC REPRESENTATION LEARNING ON SPD MANIFOLDS VIA THE SPLINE-PULLBACK METRIC

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep learning on Symmetric Positive Definite (SPD) manifolds has traditionally depended on predefined algebraic Riemannian metrics. Much like hand-engineered features in early machine learning, these fixed geometric structures constrain the expressive capacity and adaptability of the metrics. While recent works have attempted to learn parameterized geometries, they frequently breach foundational matrix function axioms via rank-dependent scaling or unbounded power operations. These violations result in spatial folding, a lack of global surjectivity, and gradient vanishing near spectral singularities. To overcome these limitations, we present the Spline-Pullback Metric (SPM), comprising both Spectral-SPM and Cholesky-SPM variants, marking a paradigm shift in SPD manifold learning from static metric choices to universal approximation. By modeling the global diffeomorphism with a monotonically constrained, rank-invariant B-spline, SPM functions as a dense universal approximator for strictly increasing diffeomorphisms. Our formulation theoretically generalizes existing pullback metrics while facilitating localized, non-linear modeling of the spectral domain. SPM guarantees a globally bijective pullback geometry, preventing gradient instabilities and rank-swapping discontinuities. We empirically demonstrate that SPM achieves superior performance results across five benchmark datasets utilizing Linear Probes, SPDNets, and deep Riemannian ResNet architectures.

1 INTRODUCTION

Covariance matrices, represented as Symmetric Positive Definite (SPD) matrices, constitute a key component in numerous modern deep learning applications spanning areas such as radar signal processing (Brooks et al., 2019) and electroencephalography (Li et al., 2024), computer vision (Huang & Van Gool, 2017), and medical imaging (Chakraborty et al., 2020). From the geometrical point of view, the SPD matrix space $\mathcal{S}^n_{++}$ represents an open and convex cone within the ambient space of symmetric matrices $\mathcal{S}^n$. As a purely topological space, this cone possesses no inherent geometry or spatial curvature. Because applying standard Euclidean operations directly to SPD matrices violates the positivity boundary constraint, the manifold must be endowed with a Riemannian metric to induce a valid geometry, thereby defining geodesic distances and enabling mathematical projections

The problem of handling the positivity boundary constraint has been traditionally tackled by formulating appropriate geometry manually. Among the pioneering methods, the Affine-Invariant Riemannian Metric (AIRM) (Pennec et al., 2006) and the Log-Euclidean Metric (LEM) (Arsigny et al., 2006) were followed by the later developed Power-Euclidean (Dryden et al., 2009), Log-Cholesky (LCM) (Lin, 2019), and Bures-Wasserstein metrics (Bhatia et al., 2019) and various generalizations of this metric (Han et al., 2023). Endowing the manifold with such metrics has paved the way for extending key deep learning building blocks to the SPD domain. Examples of such include Riemannian Multinomial Logistic Regression (MLR) (Chen et al., 2024a), convolutions, Riemannian batch normalization (Brooks et al., 2019; Chen et al., 2024c), residual blocks (Katsman et al., 2023), and graph neural networks(Cruceru et al., 2021). However, relying on classical metric selection remains a manual process, akin to handcrafted feature selection in traditional Machine Learning. The central challenge in

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.

advancing SPD matrix and Riemannian metric learning is transitioning from these static mathematical priors to dynamic, computationally efficient, and completely learnable topological representations.

Literature Limitations: Although implicit pullback metrics like LEM and LCM preclude the expensive matrix inversions and gradient instability of AIRM, they enforce rigid global topological deformations indifferent to localized spectral noise. Recent advancements attempt to parameterize Riemannian metrics to improve expressivity, but often rely on mathematically ill-posed transformations. For instance, rank-dependent mappings like ALEM (Chen et al., 2024d) violate the Primary Matrix Functions definition (Bhatia, 1997) due to eigenvector non-uniqueness, causing one-to-many topological failures. Conversely, polynomial power mappings like PCM (Chen et al., 2026) lack global surjectivity onto unconstrained Euclidean space. To prevent undefined negative pre-images, PCM relies on non-injective computational clipping (e.g., $\max(x, \epsilon)$), breaking the bijectivity strictly required for a geodesically complete Riemannian metric. We provide a rigorous mathematical proof of these diffeomorphic failures in the Appendix.

Furthermore, adapting Euclidean diffeomorphism approximators (like Affine Coupling Flows (Teshima et al., 2020) or deep MLPs (Hwang, 2026)) is infeasible for Riemannian metric learning. They typically require C^2 smooth targets and only guarantee C^0 (uniform) convergence, leaving the first derivatives, essential for stable Riemannian back-propagation, structurally discontinuous, often requiring computationally heavy layer cascades.

To resolve all these limitations, we introduce the **Spline-Pullback Metric (SPM)** and its two instantiations, Spectral-SPM (S-SPM) and Cholesky-SPM (C-SPM). In contrast to ALEM or PCM, we mathematically prove that SPM metric itself is a fully learnable, geodesically complete global Riemannian diffeomorphism and a universal approximator. Requiring only a C^1 target, SPM guarantees the C^1 convergence necessary for stable back-propagation in a single, parameter-efficient layer. To the best of our knowledge, SPM is the first framework to introduce the paradigm of universal scalar-diffeomorphism approximation for Riemannian Spectral and Cholesky Pullback Euclidean metric learning learning parameterized via B-splines, capable of localized spectral modelling without sacrificing the fundamental axioms of differential geometry.

Our main contributions are summarized as follows:

- We mathematically formalize and prove the desirable properties of SPM and its instantiations S-SPM and C-SPM.
- We highlight the current metrics' topological limitations and SPM's universal solution.
- We introduce an asymmetric Float64 matrix perturbation protocol to bound the Lipschitz constant of the Daleckĭĭ-Kreĭn structural gradient, solving the numerical instability of Riemannian backpropagation.
- Under a rigorous evaluation protocol, SPM achieves superior accuracy relative to established baselines, under a rigorous and controlled, same-backbone ablation protocol across all 9 tested combinations of datasets and architectures, across 7 metrics.

2 PRELIMINARIES

We establish the notation and geometric concepts for Symmetric Positive Definite (SPD) matrices and the mechanism of Riemannian isometric pullbacks, which form the theoretical underpining of our framework.

We denote the space of $n \times n$ real symmetric matrices as $\mathcal{S}^n$, which forms a flat Euclidean vector space equipped with the standard Frobenius inner product, $\langle X, Y \rangle_F = \text{tr}(X^\top Y)$.

The set of $n \times n$ Symmetric Positive Definite (SPD) matrices, denoted as $\mathcal{S}^n_{++}$, is an open, convex cone within $\mathcal{S}^n$. It is worth noting that the topological space $\mathcal{S}^n_{++}$ possesses no inherent spatial curvature of its own; rather, its geometry is induced by the choice of the Riemannian metric tensor g placed upon it. Because $\mathcal{S}^n_{++}$ is an open cone rather than a complete vector space, applying standard Euclidean linear operations directly to SPD matrices may produce matrices that violate the strict positivity constraint (falling outside the manifold) or results in the well-documented "swelling effect" (Arsigny et al., 2006). Consequently, $\mathcal{S}^n_{++}$ must be endowed with a Riemannian metric to measure

geodesic distances and define valid tangent space projections. At any given point $P \in \mathcal{S}^n_{++}$, its tangent space $\mathcal{T}_P\mathcal{S}^n_{++}$ is isomorphic to the flat space of symmetric matrices $\mathcal{S}^n$.

To facilitate operations requiring hierarchical sensor ordering, we also define the Cholesky space. Let $\mathcal{L}^n$ denote the flat Euclidean space of $n \times n$ lower triangular matrices. The Cholesky manifold, denoted as $\mathcal{L}^n_{++}$, is defined as the open submanifold of $\mathcal{L}^n$. It consists of lower triangular matrices where all diagonal elements are strictly positive (Lin, 2019). Any SPD matrix $S \in \mathcal{S}^n_{++}$ can be uniquely mapped to the Cholesky manifold via the Cholesky decomposition, $S = LL^\top$, where $L \in \mathcal{L}^n_{++}$.

While historically popular classical metrics like the **Affine-Invariant Riemannian Metric (AIRM)** (Pennec et al., 2006) induce negative sectional curvature requiring expensive iterative approximations (Karcher, 1977), modern networks have shifted toward pullback metrics that induce a globally flat geometry.

Definition 1 (Pullback Euclidean Metric)(Thanwerdas & Pennec, 2022). Let $\mathcal{M}$ represent a smooth topological manifold (such as the SPD cone $\mathcal{S}^n_{++}$), and let $\mathcal{E}$ represent a flat Euclidean vector space (such as $\mathcal{S}^n$ or $\mathcal{L}^n$). If there exists a smooth, globally bijective mapping, a diffeomorphism, $\phi : \mathcal{M} \to \mathcal{E}$, one can equip $\mathcal{M}$ with a globally flat Riemannian geometry by pulling back the standard Frobenius inner product from $\mathcal{E}$. The Pullback Euclidean Metric (PEM) g_ϕ evaluated at point $P \in \mathcal{M}$ for tangent vectors $V_1, V_2 \in \mathcal{T}_P\mathcal{M}$ is defined point-wise as:

$$g_\phi|_P(V_1, V_2) = \langle \mathcal{D}\phi_P(V_1), \mathcal{D}\phi_P(V_2) \rangle_F \tag{1}$$

where $\mathcal{D}\phi_P : \mathcal{T}_P\mathcal{M} \to \mathcal{E}$ is the pushforward differential map of ϕ at P.

Because the target space $\mathcal{E}$ is flat, any Riemannian metric constructed via this global diffeomorphic pullback guarantees that the manifold $\mathcal{M}$ possesses zero sectional curvature globally. This eliminates the need for iterative Karcher flow approximations.

The topological deformation of a Pullback Euclidean Metric is dictated by ϕ. Two prevalent baselines enforce immutable logarithmic deformations:

Log-Euclidean Metric (LEM) (Arsigny et al., 2006): Maps to $\mathcal{S}^n$ via the eigendecomposition $S = U\Sigma U^\top$:

$$\phi_{LE}(S) = \log(S) = U \ln(\Sigma) U^\top \tag{2}$$

Log-Cholesky Metric (LCM) (Lin, 2019): Maps to the Cholesky manifold $\mathcal{L}^n_{++}$ via factorization $S = LL^\top$ to bypass $\mathcal{O}(n^3)$ eigendecompositions. It applies the natural logarithm exclusively to the strictly positive diagonal elements:

$$\phi_{LC}(S) = \lfloor L \rfloor + \log(\text{diag}(L)) \tag{3}$$

where $\lfloor L \rfloor$ isolates the strictly lower triangular elements.

While stable, LEM and LCM enforce rigid deformations. A discussion on fundamental Riemannian operators and Splines is deferred to the Appendix.

3 The Spline-Pullback Metric framework (SPM)

Classical metrics impose topological restrictions, and scalar parametric metrics often falter at boundaries. We propose the Spline-Pullback Metric (SPM) framework to address these issues and propose a universally applicable framework. SPM learns non-linear topological deformations end-to-end by drawing on the localized support properties of B-splines. In this section, we formalize the universal scalar diffeomorphism at the core of SPM, detail its matrix instantiations, and outline the resulting geometric properties. The Curry-Schoenberg Theorem (de Boor, 2001, Ch. IX, Thm. 44) establishes that B-splines provide an optimal, linearly independent basis for piecewise polynomial spaces, enabling non-linear modelling without triggering the Runge's phenomenon that mars polynomial regression. The proofs of all the subsequent theorems are provided in the Appendix.

3.1 The universal scalar diffeomorphism

SPM relies on a parameterized, strictly monotonic scalar mapping $f_\theta : \mathbb{R}_{>0} \to \mathbb{R}$. To avert the lack of global surjectivity and gradient instability that plague power metrics, and ensure full domain

coverage, $f_\theta(x)$ operates in the logarithmic domain and combines a central B-spline with bounded linear extrapolations at the edges.

Definition 2 (The SPM Scalar Generator). *Let $T = \{t_0, t_1, \ldots, t_m\}$ be a predefined, clamped knot vector. The SPM scalar generator $f_\theta : \mathbb{R}_{>0} \to \mathbb{R}$ is defined as a composite mapping, evaluated as a linear combination of B-spline basis functions $B_{i,k}(x)$ of degree k. Outside T, it extends via bounded linear extrapolation using the boundary derivatives.*

$$f_\theta(x) = \begin{cases} m_L(\log(x) - t_k) + S(t_k) & \text{if } \log(x) < t_k \\ S(\log(x)) & \text{if } t_k \le \log(x) \le t_{m-k} \\ m_R(\log(x) - t_{m-k}) + S(t_{m-k}) & \text{if } \log(x) > t_{m-k} \end{cases} \tag{4}$$

where $S(x) = \sum c_i B_{i,k}(x)$ is a B-spline of degree k with learnable control coefficients $c_i \in \mathbb{R}$, and m_L, m_R represent the strictly positive analytical derivatives of $S(x)$ at the respective grid boundaries.

The basis functions $B_{i,k}(x)$ are defined via the numerically stable Cox-de Boor recursion formula(de Boor, 2001, Ch. IX, Prop. i). For degree $k = 0$:

$$B_{i,0}(x) = \begin{cases} 1 & \text{if } t_i \le x < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

And for degree $k > 0$, they are constructed recursively:

$$B_{i,k}(x) = \frac{x - t_i}{t_{i+k} - t_i} B_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(x) \tag{6}$$

Proposition 1 (B-Spline Derivative). *As a direct consequence of the Cox-de Boor recursion, the first derivative of the B-spline $S(x)$ of degree k is another B-spline of degree $k-1$: (de Boor, 2001, Ch. X, Prop. viii)*

$$S'(x) = \sum_i k \frac{c_i - c_{i-1}}{t_{i+k} - t_i} B_{i,k-1}(x) \tag{7}$$

Theorem 1 (Global Scalar Diffeomorphism). *B.1 Let $S(x)$ be a B-spline of degree $k \ge 1$ over a clamped knot vector T with simple interior knots, with boundary linear extrapolations matching S in value and first derivative at t_k and t_{m-k}. If the sequence of learnable coefficients c is strictly monotonically increasing ($c_i > c_{i-1}$ for all i), then $f_\theta(x)$ is a globally bijective, C^1-smooth diffeomorphism from $\mathbb{R}_{>0}$ to $\mathbb{R}$. Within the open interior (t_k, t_{m-k}) of the knot grid, f_θ is additionally C^{k-1}-smooth (such as C^2 for the cubic splines, $k = 3$, used throughout).*

To apply this constraint directly during continuous gradient descent, we use a cumulative *softplus* reparameterization: $c_i = c_{i-1} + \text{softplus}(w_i) + \epsilon$. While classical metrics enforce rigid deformations, a fully learnable metric must possess the theoretical capacity to model any valid non-linear metric.

Theorem 2 (Universal Approximation on Compact Spectra). *B.2 The constrained SPM mapping $f_\theta(x)$ is dense in the space of strictly increasing C^1 diffeomorphisms from the positive real line to the unconstrained real line, denoted $\text{Diff}^1_+(\mathbb{R}_{>0}, \mathbb{R})$, over any compact spectral interval. That is, for any target geometry $g \in \text{Diff}^1_+(\mathbb{R}_{>0}, \mathbb{R})$, any compact data interval $[a, b] \subset \mathbb{R}_{>0}$, and any error tolerance $\epsilon > 0$, there exists a finite knot vector T encompassing $\log([a, b])$ and a strictly increasing sequence of control coefficients $c_i > c_{i-1}$ such that both $\sup_{x \in [a,b]} |f_\theta(x) - g(x)| < \epsilon$ and $\sup_{x \in [a,b]} |f'_\theta(x) - g'(x)| < \epsilon$.*

A direct mathematical consequence of this dense approximation capacity is that SPM unifies prior metric literature into a single framework. Proofs are presented in the Appendix.

Corollary 1 (Universal Subsumption of Analytical Geometries). *B.3 Let $\mathcal{H}_{SPM}$ denote the hypothesis space of Riemannian metrics induced by the SPM framework. By virtue of Theorem 2, $\mathcal{H}_{SPM}$ theoretically subsumes all orientation-preserving, topologically valid parametric and non-parametric scalar-diffeomorphism-based pullback metrics currently established in the literature.*

Proposition 2 (Topological Safety and Geometric Expressivity of the SPM Generator). *B.4 Let $f_\theta(x)$ be the SPM scalar generator using a strictly increasing control polygon. The mapping $f_\theta(x)$ is a global diffeomorphism immune to Runge's phenomenon, with a strictly positive first derivative $f'_\theta(x)$. Although strictly monotonic, the second derivative $f''_\theta(x)$ is topologically unconstrained. This allows it to form dynamic inflection points for Localized Spectral modelling*

3.2 INSTANTIATION I: SPECTRAL-SPM (S-SPM) AND EIGENSPACE OPTIMIZATION

Applying a scalar diffeomorphism to the eigenvalues of an SPD matrix creates a global diffeomorphism on the matrix space itself. Permutation invariance is important when handling unstructured physiological data with arbitrary sensor ordering. We define the S-SPM mapping $\phi_{S-SPM} : \mathcal{S}^n_{++} \to \mathcal{S}^n$ using the eigendecomposition $S = U\Sigma U^\top$:

$$\phi_{S-SPM}(S) = U\text{diag}(f_\theta(\lambda_1), \ldots, f_\theta(\lambda_n))U^\top \tag{8}$$

Theorem 3 (Diffeomorphism and Congruence Invariance of S-SPM). *B.5 The mapping $\phi_{S-SPM} : \mathcal{S}^n_{++} \to \mathcal{S}^n$ is a global diffeomorphism, and its isometric pullback induces a valid Riemannian metric on $\mathcal{S}^n_{++}$ that is invariant under orthogonal congruence: $g_{RSR^\top}(RVR^\top, RWR^\top) = g_S(V, W)$ for all $R \in O(n)$.*

Integrating SPM into end-to-end deep learning architectures requires an analytically exact and numerically stable backward pass. Because computing the exact Lipschitz constant of a deep network is NP-hard (Virmaux & Scaman, 2018), network stability relies on ensuring that the operator norm of the local gradient at every layer is strictly bounded. We first state the analytical gradient for our mapping, followed by a mechanism to bound its Lipschitz constant. Note that we distinguish two classes of statements below. *Theorems* and *Propositions* establish continuous mathematical guarantees under exact real arithmetic. *Computational Safeguard Protocols* are discrete engineering mechanisms ensuring these guarantees hold stably under finite-precision GPU arithmetic; they do not alter or substitute for the underlying continuous geometry.

Proposition 3 (Analytical Structural Gradient). *B.6 Let $X = \phi_{SPM}(S)$ be the forward mapping. Following the standard adjoint calculus for spectral functions (Ionescu et al., 2015) and its unified closed-form derivation for SPD manifolds (Bhatia, 1997), the backpropagation gradient with respect to the input SPD matrix S is analytically defined as:*

$$\nabla_S \mathcal{L} = U \left[K \odot (U^\top (\nabla_X \mathcal{L}) U) \right] U^\top \tag{9}$$

where $\mathcal{L}$ is the loss function, $\odot$ is the Hadamard product, and K is the Daleckĭĭ-Kreĭn divided difference matrix. Our monotonic spline f_θ constitutes a Primary Matrix Function (Bhatia, 1997), so K evaluates using only the eigenvalue magnitudes:

$$K_{ij} = \begin{cases} \frac{f_\theta(\lambda_i) - f_\theta(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j \\ f'_\theta(\lambda_i) & \text{if } \lambda_i = \lambda_j \end{cases} \tag{10}$$

The orthogonal matrices U hold an operator norm of 1. The maximum absolute value in matrix K bottlenecks the Lipschitz constant of this backward pass. When spectral singularities occur ($\lambda_i = \lambda_j$), evaluating K_{ij} becomes unstable. We bound the operator norm using an asymmetric spectral perturbation.

Computational Safeguard Protocol (Gradient Bounding via Asymmetric Spectral Perturbation). *While continuous theory smoothly resolves identical eigenvalues via L'Hôpital's rule, discrete GPU eigensolvers fault on $0/0$ division. To safeguard hardware execution without altering the continuous theoretical geometry, let $\Sigma = diag(\lambda_1, \ldots, \lambda_n)$ represent the sorted eigenvalue matrix. Before evaluating the Daleckĭĭ-Kreĭn gradient matrix K, we perturb the eigenvalues in 64-bit precision:*

$$\tilde{\Sigma} = \Sigma_{f64} + \text{diag}(\vec{\epsilon}) \tag{11}$$

where $\vec{\epsilon} \in \mathbb{R}^n$ is an increasing vector such that $\epsilon_i = i \cdot \delta$ with a small machine constant $\delta > 0$. This asymmetric perturbation acts as a discrete engineering safeguard to strictly bound the Lipschitz constant of the backward pass in practice.

3.3 INSTANTIATION II: CHOLESKY-SPM (C-SPM) VIA PRODUCT GEOMETRY

For hierarchically ordered data, such as kinematic skeletons and radar arrays, the $\mathcal{O}(n^3)$ cost of eigendecomposition poses a computational bottleneck. C-SPM circumvents this by leveraging the intrinsic product geometry of the Cholesky manifold $\mathcal{L}^n_{++}$. Following recent advancements (Lin, 2019; Chen et al., 2026), the Cholesky manifold can be decoupled into a product space:

$\mathcal{L}^n_{++} \cong \mathcal{SL}^n \times (\mathbb{R}_{>0})^n$, where $\mathcal{SL}^n$ is the flat Euclidean space of strictly lower triangular matrices, and $(\mathbb{R}_{>0})^n$ represents the strictly positive diagonals.

We define a global diffeomorphism $\phi_{\mathcal{L}}$ that assigns the identity mapping to the Euclidean strictly lower triangular elements ($\lfloor \cdot \rfloor$), while applying our scalar diffeomorphism f_θ to the non-Euclidean diagonal elements:

$$\phi_{\mathcal{L}}(L) = \lfloor L \rfloor + \text{diag}\Big(f_\theta(L_{11}), \ldots, f_\theta(L_{nn})\Big) \tag{12}$$

The C-SPM metric on the target SPD manifold $\mathcal{S}^n_{++}$ is obtained via a composite pullback through the Cholesky decomposition $\text{Chol} : \mathcal{S}^n_{++} \to \mathcal{L}^n_{++}$. The composite mapping to the flat space $\mathcal{L}^n$ is therefore:

$$\phi_{C-SPM}(S) = \phi_{\mathcal{L}}(\text{Chol}(S)) \tag{13}$$

Theorem 4 (Diffeomorphism of C-SPM). *B.8 The composite mapping $\phi_{C-SPM} : \mathcal{S}^n_{++} \to \mathcal{L}^n$ is a global diffeomorphism, and its pullback induces a valid Riemannian metric on $\mathcal{S}^n_{++}$.*

Remark: C-SPM avoids the Daleckĭĭ-Kreĭn structural gradient described in Proposition 3 by only working in the Cholesky product space. The backward pass depends entirely on the stable differentials of the Cholesky factorization, which gives it better native numerical stability and a speedup in algorithms.

3.4 Geometric and topological properties of the SPM manifold

By establishing the Spline-Pullback Metric (SPM) through a global diffeomorphism to the ambient Euclidean space, the resultant Riemannian manifold acquires several advantageous topological characteristics that alleviate traditional computational challenges in SPD manifold learning.

Theorem 5 (Global Flatness and Zero Sectional Curvature). *B.9 The Riemannian manifold $(\mathcal{S}^n_{++}, g_{SPM})$ induced by either the S-SPM or C-SPM pullback has constantly zero sectional curvature everywhere.*

Theorem 6 (Existence, Uniqueness, and Closed-Form of the Fréchet Mean). *B.10 For any finite set of SPD matrices $\{P_1, \ldots, P_N\} \in \mathcal{S}^n_{++}$ with positive weights w_i such that $\sum w_i = 1$, the SPM Fréchet Mean is globally unique and admits an exact closed-form expression.*

This eliminates the need for iterative Karcher flow approximations required by metrics with non-zero curvature.

Theorem 7 (Path-Independent Parallel Transport). *B.11 Let $A, B \in \mathcal{S}^n_{++}$ and $V \in \mathcal{T}_A\mathcal{S}^n_{++}$. The parallel transport of the tangent vector V from A to B along the SPM geodesic is path-independent and evaluated in closed form.*

Corollary 2 (Algebraic Structures). *Because ϕ_{S-SPM} and ϕ_{C-SPM} act as global isometries to flat vector spaces, the resulting metric space $\{\mathcal{S}^n_{++}, g_{SPM}\}$ inherits the algebraic structure of the ambient Euclidean space, forming an Abelian Lie group under the transported operation $S_1 \oplus_\phi S_2 := \phi^{-1}(\phi(S_1) + \phi(S_2))$, and a Hilbert space over $\mathbb{R}$.*(Arsigny et al., 2006)

Remark (Bi-invariance). It is with respect to $\oplus_\phi$, not orthogonal congruence, that g_{SPM} is bi-invariant in the classical Lie-group sense: since ϕ is a group isomorphism onto $(\mathbb{R}^n, +)$ by construction, translation under $\oplus_\phi$ corresponds to Euclidean translation under ϕ, which preserves the flat Frobenius metric.

4 Applications to deep Riemannian networks

The closed-form and globally stable nature of SPM enables the robust construction of deep neural network architectures directly on the SPD manifold. We integrate SPM into two fundamental deep learning building blocks: Multinomial Logistic Regression and Residual Blocks.

Riemannian Multinomial Logistic Regression (MLR): Euclidean MLR acts as a hyperplane classifier (Lebanon & Lafferty, 2004), which generalizes to Riemannian manifolds via point-to-hyperplane distances: $p(y = k|X) \propto \exp(\langle A_k, \text{Log}_{P_k}(X)\rangle_{P_k})$, where P_k is the class-specific manifold centroid and A_k is the tangent normal vector(Chen et al., 2024b).

Theorem 8 (SPM-MLR Classifier). *B.12 Because the SPM framework operates as a global isometric pullback to a flat Euclidean space, the Riemannian inner product $\langle A_k, Log_{P_k}(X)\rangle_{P_k}$ algebraically simplifies to a standard Euclidean Frobenius inner product on the mapped matrices.*

This formulation eliminates the need to compute complex tangent space projections or inverse differentials during the final classification stage, maximizing computational throughput and stability.

Riemannian Residual Blocks: Deep Euclidean residual connections (Katsman et al., 2023) are generalized via the Riemannian Exponential map to ensure outputs remain on the SPD manifold: $X^{(i)} = \text{Exp}_{X^{(i-1)}}(\mathcal{F}_\omega(X^{(i-1)}))$. Because SPM is geodesically complete, the network can process large residual tangent updates without violating topological constraints, unlike geodesically incomplete metrics.

5 Numerical Experiments

Before evaluating our proposed framework on real-world SPD manifold datasets, we conduct two controlled numerical experiments. These experiments are designed to empirically validate the topological safety bounds of Proposition 2 and to demonstrate the capacity limitations of classical metrics when faced with structured spectral noise.

5.1 1D Synthetic Function Approximation

We optimize the spline to approximate three different target functions, demonstrating its universal diffeomorphism approximation capabilities, as illustrated in Figure 1.

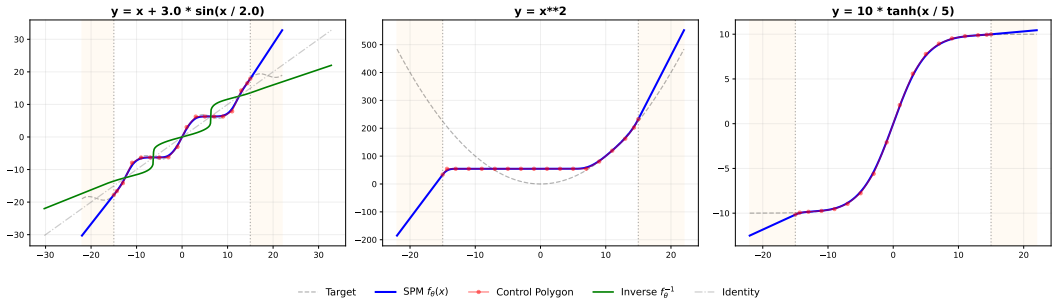

Figure 1: Empirical validation of SPM properties. **(Left)**Unconstrained second-order differences ($\Delta^2 c_i$) enable the metric to closely monitor inflection points while maintaining monotonicity with the smooth inverse precluding singular gradients. **(Center)** The SPM survives an adversarial, non-monotonic gradient by refusing to break bijectivity by maintaining a positive slope throughout. **(Right)** SPM, utilizing its dynamic inflection capabilities, can cap noisy outliers, unachievable by fixed algebraic metrics.

5.2 Adversarial Classification

To illustrate the capacity limitations of fixed geometries, we created a synthetic binary classification dataset comprising 1000 4×4 SPD matrices partitioned into alternating signal and noise spectral bands. As quantitatively detailed and visualized in Appendix D, all non-adaptive classical baselines suffered from severe underfitting (max 66.00% accuracy) due to their fixed derivatives. Conversely, S-SPM achieved 100.00% accuracy by leveraging Localized Spectral Modelling to dynamically expand signal bands and compress noise bands—a sequence of dynamic inflections mathematically impossible under classical metric priors.

6 Experiments

To assess the empirical efficacy of SPM, we use 3 widely used datasets, **HDM05(Müller et al., 2007)**, **FPHA(Garcia-Hernando et al., 2018)**, and **Radar(Brooks et al., 2019)**, each with 3 different

models, **Direct Linear Probe**, **SPDNet (SPD-MLR)**, and **Deep RResNet:**, across 7 metrics. The standard train-test split was used for FPHA evaluation, while 5-fold cross-validation was used for the rest. This set of datasets and models was chosen for a holistic evaluation and fair comparison with the literature due to its popular benchmarking usage in the literature (Chen et al., 2026; 2024d). Further implementation details are presented in the appendix.

Table 1: Classification accuracy on the HDM05 dataset. All models utilize 5-fold stratified cross-validation. Best results are highlighted in bold.

Metric	Linear	SPD MLR	RResNet
AIRM	0.5408 ± 0.0413	0.3876 ± 0.0205	0.5980 ± 0.0244
LE	0.5437 ± 0.0441	0.6594 ± 0.0179	0.6734 ± 0.0227
LC	0.5005 ± 0.0260	0.6066 ± 0.0302	0.6676 ± 0.0209
θ-PCM ($\theta = 0.5$)	0.4962 ± 0.0334	0.6330 ± 0.0168	0.6580 ± 0.0362
ALEM (Learn)	0.5855 ± 0.0568	0.6719 ± 0.0110	0.6777 ± 0.0195
C-SPM (Ours)	0.5831 ± 0.0195	0.6292 ± 0.0306	0.6498 ± 0.0188
S-SPM (Ours)	$\mathbf{0.6032 \pm 0.0341}$	$\mathbf{0.6763 \pm 0.0272}$	$\mathbf{0.6801 \pm 0.0124}$

Table 2: Classification accuracy on the FPHA dataset. Evaluated on the standard predefined train-test split.

Metric	Linear	SPD MLR	RResNet
AIRM	0.6226	0.6174	0.7461
LE	0.8400	0.8557	0.8487
LC	0.6957	0.7426	0.7739
θ-PCM	0.6817 ($\theta = 0.75$)	0.7357 ($\theta = 0.75$)	0.7600 ($\theta = 0.5$)
ALEM (Learn)	0.8539	0.8852	0.8487
C-SPM (Ours)	0.6887	0.7391	0.7791
S-SPM (Ours)	**0.8574**	**0.9043**	**0.8557**

Table 3: Classification accuracy on the Radar Signal Processing Dataset (5-fold CV).

Metric	Linear Probe	SPD MLR	RResNet
AIRM	0.9627 ± 0.0108	0.9627 ± 0.0100	0.9087 ± 0.0215
LE	0.9630 ± 0.0087	0.9600 ± 0.0129	0.9267 ± 0.0130
LC	0.9677 ± 0.0105	0.9617 ± 0.0115	0.9303 ± 0.0128
θ-PCM	0.9660 ± 0.0112 ($\theta = 0.5$)	0.9593 ± 0.0140 ($\theta = 0.5$)	0.9183 ± 0.0192 ($\theta = -1.0$)
ALEM (Learnable)	0.9650 ± 0.0064	0.9580 ± 0.0090	0.9210 ± 0.0248
C-SPM (Ours)	$\mathbf{0.9687 \pm 0.0090}$	0.9617 ± 0.0078	0.9240 ± 0.0177
S-SPM (Ours)	0.9633 ± 0.0075	$\mathbf{0.9657 \pm 0.0076}$	$\mathbf{0.9323 \pm 0.0106}$

As demonstrated in Tables 1, 2, and 3, the proposed SPM framework achieves superior accuracy relative to all the ablated established baselines across all 9 architectural configurations. Predefined (LE/LCM) and scalar parametric metrics (θ-PCM) rigidly induce information erasure during tangent space projection, failing to adapt when intermediate layers transform the manifold. Conversely, SPM securely maps these unregularized features. On the complex FPHA manifold, S-SPM (0.9043) provides a significant improvement over PCM (0.7357, Bootstrap $p < 0.05$) in the SPDNet architecture, while also drastically outpacing baselines on uncompressed Linear Probes (like 0.8574 vs. PCM's 0.6817).

Finally, on the hierarchical Radar dataset, C-SPM achieves the highest linear probe accuracy (0.9687) and maintains highly competitive scores in deep RResNets. When strictly compared against other Cholesky-based metrics, C-SPM unveils a performance advantage while maintaining superior computational efficiency and numerical stability.

Unlike rigid metrics, SPM does not have any manually tunable hyperparameter that directly affects the curve's shape. Regarding architectural hyperparameters (spline degree and grid size), SPM is exceptionally stable. This is expected because these hyperparameters dictate the curve's resolution rather than its global shape, which is learned. We demonstrate this via an ablation study on the FPHA dataset. As shown in the appendix Table 6, accuracy and training time remained consistent across 9 different configurations, tightly clustered around the reported result

6.1 Large-Scale Benchmarks and Statistical Significance

To evaluate scalability to high-dimensional and large-scale data, we expanded our benchmarks to the modern DAIR AI-NLP dataset (Saravia et al., 2018) ($N = 20,000$ text covariance matrices) and the AFEW dataset (Dhall et al., 2018) (400×400 vision covariance matrices). We employed a strict metric-isolation protocol using uncompressed architectures to avoid confounding variables from heavy backbones. Paired predictions were evaluated using McNemar's test with continuity correction, adjusted via the Benjamini-Hochberg FDR procedure, alongside 95% CIs from 10,000 bootstraps. As detailed in Appendix F, S-SPM achieved **statistically significant** superior performance relative to established baselines under these controlled setups. Furthermore, on the high dimensional 400×400 AFEW matrices, C-SPM reduced the per-epoch training time from 18.2s (AIRM) to just 4.3s, with the dynamic B-spline parameterization introducing a negligible overhead of ~ 20 milliseconds per epoch compared to rigid analytical baselines.

7 Discussion and conclusion

The transition to SPM mirrors the paradigm shift from classical feature engineering to deep representation learning. While parametric metrics are fundamentally limited by their rigid mathematical priors, despite expensive grid searches, SPM, utilizing constrained B-splines, establishes a globally complete Euclidean pullback space and treats the Riemannian metric tensor itself as an end-to-end learnable representation, yielding superior performance across a suite of experiments. As a parameter-efficient universal diffeomorphism approximator, SPM inherits the high expressivity of deep learning, opening three exciting frontiers in geometric research:

Riemannian Metric Explainable AI (RMXAI): While predefined metrics do not provide insight into the underlying data distribution, visualizing and interpreting the learned SPM spline offer direct insight into the dataset's spectral structure. We refer readers to Appendix G for an RMXAI analysis of SPM and its stability.

Riemannian Metric Initialization (RMI): Like neural networks, SPM benefits from different initialization strategies. Future research in RMI will explore initialization distributions tailored to the SPM manifold, akin to Glorot(Glorot & Bengio, 2010) or He (He et al., 2015) initialization, to ensure faster and more stable metric convergence.

Riemannian Metric Regularization (RMR): Though our experiments do not suffer from any pronounced overfitting, future applications of SPM in complex and data-scarce environments may lead to overfitting concerns. Work in RMR will find regularization techniques, beyond standard Euclidean weight decay, which would ensure the integrity of the topological and diffeomorphic rules of the manifold. A simple idea, for example, is to regularize the second derivative of the SPM curve. An analysis of the framework's limitations is provided in Appendix K.

AI use statement

In this submission, generative AI tools were utilized to refine the manuscript's grammar, improve overall readability, and polish language presentation. We confirm that no AI tools were employed to formulate the primary mathematical claims, develop the core theoretical models, write the end-to-end software implementations, or generate synthetic data. The authors have rigorously vetted all AI-assisted work, text refinement, and structural formatting. We assume full responsibility for the integrity, original theoretical contributions, and entirety of the paper's content.

Ethics statement

Our evaluations were conducted only on standard, open-source benchmark datasets, none of which include sensitive or personally identifiable information. We do not anticipate any adverse societal consequences or ethical concerns resulting from the methodological improvements introduced in this study.

Reproducibility statement

To ensure reproducibility, all mathematical assertions are grounded in explicit geometric axioms, and their corresponding proofs are thoroughly detailed in Appendix B. Complete documentation of our experimental protocols, encompassing network configurations, dataset preprocessing methods, and essential numerical safeguards, is provided in the Appendix. The source code will be made publicly available following the acceptance of this manuscript.

References

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.

Rajendra Bhatia. *Matrix Analysis*. Springer, 1997. doi: 10.1007/978-1-4612-0653-8.

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the bures–wasserstein distance between positive definite matrices. *Expositiones mathematicae*, 37(2):165–191, 2019.

Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for spd neural networks. *Advances in neural information processing systems*, 32, 2019.

Rudrasis Chakraborty, Jose Bouza, Jonathan H Manton, and Baba C Vemuri. Manifoldnet: A deep neural network for manifold-valued data with applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(2):799–810, 2020.

Ziheng Chen, Yue Song, Gaowen Liu, Ramana Rao Kompella, Xiao-Jun Wu, and Nicu Sebe. Riemannian multinomial logistics regression for spd neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17086–17096, 2024a.

Ziheng Chen, Yue Song, Rui Wang, Xiao-Jun Wu, and Nicu Sebe. Rmlr: Extending multinomial logistic regression into general geometries. *Advances in Neural Information Processing Systems*, 37:59716–59754, 2024b.

Ziheng Chen, Yue Song, Xiaojun Wu, and Nicu Sebe. Gyrogroup batch normalization. In *The Thirteenth International Conference on Learning Representations*, 2024c.

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, 2024d.

Ziheng Chen, Yue Song, Xiaojun Wu, and Nicu Sebe. Fast and stable riemannian metrics on SPD manifolds via cholesky product geometry. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=5S8ruWKe8l.

Calin Cruceru, Gary Bécigneul, and Octavian-Eugen Ganea. Computationally tractable riemannian manifolds for graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 7133–7141, 2021.

Carl de Boor. *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, revised edition, 2001. ISBN 0-387-95366-3.

Abhinav Dhall, Amanjot Kaur, Roland Goecke, and Tom Gedeon. Emotiw 2018: Audio-video, student engagement and group-level affect prediction. In *Proceedings of the 20th ACM international conference on multimodal interaction*, pp. 653–656, 2018.

Ian L Dryden, Alexey Koloydenko, and Diwei Zhou. Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, pp. 1102–1123, 2009.

Guillermo Garcia-Hernando, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim. First-person hand action benchmark with rgb-d videos and 3d hand pose annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 409–419, 2018.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings, 2010.

H Gonska, M Wozniczka, and F Zeilfelder. A note on approximation properties of derivatives of schoenberg splines. *Mathematica (Cluj)*, 52(1):15–29, 2010.

Andi Han, Bamdev Mishra, Pratik Jawanpuria, and Junbin Gao. Learning with symmetric positive definite matrices via generalized bures-wasserstein geometry. In Frank Nielsen and Frédéric Barbaresco (eds.), *Geometric Science of Information*, pp. 405–415, Cham, 2023. Springer Nature Switzerland.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Geonho Hwang. Minimum width for deep, narrow MLP: A diffeomorphism approach. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=S2fswKCN9T.

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.

Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.

Isay Katsman, Eric Chen, Sidhanth Holalkere, Anna Asch, Aaron Lou, Ser Nam Lim, and Christopher M De Sa. Riemannian residual neural networks. *Advances in Neural Information Processing Systems*, 36:63502–63514, 2023.

Guy Lebanon and John Lafferty. Hyperplane margin classifiers on the multinomial manifold. In *Proceedings of the twenty-first international conference on Machine learning*, pp. 66, 2004.

John M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, New York, 2nd edition, 2012. doi: 10.1007/978-1-4419-9982-5.

John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer, 2018.

Shanglin Li, Motoaki Kawanabe, and Reinmar J Kobler. Spdim: Source-free unsupervised conditional and label shift adaptation in eeg. *arXiv preprint arXiv:2411.07249*, 2024.

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

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov–arnold networks. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Ozo7qJ5vZi.

Martin J Marsden. An identity for spline functions with applications to variation-diminishing spline approximation. *Journal of Approximation Theory*, 3(1):7–49, 1970. ISSN 0021-9045. doi: https://doi.org/10.1016/0021-9045(70)90058-4. URL https://www.sciencedirect.com/science/article/pii/0021904570900584.

Meinard Müller, Tido Röder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. Mocap database hdm05. *Institut für Informatik II, Universität Bonn*, 2(7):1–34, 2007.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A riemannian framework for tensor computing. *International Journal of computer vision*, 66(1):41–66, 2006.

Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 2002.

Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. CARER: Contextualized affect representations for emotion recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3687–3697, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1404. URL https://www.aclweb.org/anthology/D18-1404.

Takeshi Teshima, Isao Ishikawa, Koichi Tojo, Kenta Oono, Masahiro Ikeda, and Masashi Sugiyama. Coupling-based invertible neural networks are universal diffeomorphism approximators. *Advances in Neural Information Processing Systems*, 33:3362–3373, 2020.

Yann Thanwerdas and Xavier Pennec. The geometry of mixed-euclidean metrics on symmetric positive definite matrices. *Differential Geometry and its Applications*, 81:101867, 2022.

Aladin Virmaux and Kevin Scaman. Lipschitz regularity of deep neural networks: analysis and efficient estimation. *Advances in neural information processing systems*, 31, 2018.

A Appendix: preliminaries

Table 4: Summary of principal notation.

Symbol	Meaning
$\mathcal{S}^n$, $\mathcal{S}^n_{++}$	Symmetric matrices; SPD cone
$\mathcal{L}^n$, $\mathcal{L}^n_{++}$	Lower-triangular matrices; positive-diagonal submanifold
$\phi, \phi_{S-SPM}, \phi_{C-SPM}$	Global diffeomorphism; spectral/Cholesky instantiations
g_ϕ, g_{SPM}	Pullback Euclidean metric induced by ϕ
f_θ	SPM scalar generator (spline in log-domain)
$S(y)$, $B_{i,k}(y)$	B-spline of degree k; its basis functions
$c_i, T = \{t_0, \dots, t_m\}$	Learnable control coefficients; clamped knot vector
m_L, m_R	Boundary linear-extrapolation slopes
K_{ij}	Daleckĭĭ-Kreĭn divided-difference matrix
U, Σ	Eigenvectors, eigenvalues in $S = U\Sigma U^\top$
L	Cholesky factor, $S = LL^\top$
$\oplus_\phi$	Transported operation, $S_1 \oplus_\phi S_2 = \phi^{-1}(\phi(S_1) + \phi(S_2))$

Riemannian Calculus and Operators

Endowing the topological space $\mathcal{S}^n_{++}$ with a Riemannian metric g allows for the generalization of Euclidean calculus to the manifold. For any points $P, Q \in \mathcal{S}^n_{++}$ and a tangent vector $V \in \mathcal{T}_P\mathcal{S}^n_{++}$, the fundamental Riemannian operators are defined as follows:

Geodesics and Distance: A geodesic $\gamma(t)$ is the locally length-minimizing curve connecting two points on the manifold, acting as the non-Euclidean generalization of a straight line. The geodesic distance $d_g(P, Q)$ is the integrated length of this curve under the metric g.

Exponential and Logarithmic Maps: To traverse the manifold, one must map between the curved space and its flat tangent approximations. The Riemannian Exponential map, $\text{Exp}_P(V) : \mathcal{T}_P\mathcal{S}^n_{++} \to \mathcal{S}^n_{++}$, projects a tangent vector back onto the manifold such that the curve $\gamma(t) = \text{Exp}_P(tV)$ is a geodesic with initial velocity V. Its inverse, the Riemannian Logarithm map, $\text{Log}_P(Q) : \mathcal{S}^n_{++} \to \mathcal{T}_P\mathcal{S}^n_{++}$, calculates the exact tangent vector (initial velocity) required to reach Q from P.

The Fréchet Mean: Because arithmetic averaging falls outside the SPD cone, the center of mass for a set of points $\{P_i\}_{i=1}^N$ with weights w_i must be computed geometrically. The Riemannian Fréchet Mean generalizes the Euclidean average by minimizing the sum of squared geodesic distances(Karcher, 1977):

$$\mu_F = \arg\min_{M \in \mathcal{S}^n_{++}} \sum_{i=1}^{N} w_i d_g^2(M, P_i) \tag{14}$$

Under metrics with non-zero sectional curvature (e.g., AIRM), computing the Fréchet mean requires iterative, computationally expensive gradient descent (Karcher flow). Conversely, under the globally flat geometry of a Pullback Euclidean Metric, the Fréchet mean and all associated Riemannian maps admit mathematically exact, closed-form Euclidean solutions within the pullback space.

Splines in Geometric Computing. While B-splines have a rich history in computational anatomy (like LDDMM (Beg et al., 2005)) and image registration (like Free-Form Deformations(Rueckert et al., 2002)), these classical frameworks utilize splines to parameterize *spatial* coordinate grids to align volumetric data, and recently, the benefits of their Universal Approximation Properties in deep learning are gaining traction, as evidenced by their utilization in Kolmogorov-Arnold Networks (KANs) (Liu et al., 2025). Furthermore, in SPD literature, splines are occasionally used for temporal trajectory interpolation *on top of* a fixed, predefined Riemannian metric. However, for the first time, instead of using splines to traverse a manifold spatially or temporally, the SPM framework uses learnable, strictly monotonic B-splines to construct the spectral Riemannian metric tensor itself, reducing complex multi-dimensional Jacobian constraints to a computationally trivial 1D monotonicity constraint.

B Proofs of theoretical claims

B.1 Proof of theorem 1

Proof. A smooth, globally bijective map with a smooth inverse forms a global diffeomorphism(Lee, 2012, Ch. 2, p. 39). Our scalar generator is defined as a composition: $f_\theta(x) = S(\log(x))$, where $\log : \mathbb{R}_{>0} \to \mathbb{R}$ is the standard natural logarithm, and $S : \mathbb{R} \to \mathbb{R}$ is the monotonic B-spline with linear boundary extrapolation. The natural logarithm is a standard global C^∞ diffeomorphism. Since composing two diffeomorphisms yields another diffeomorphism (Lee, 2012, Ch. 2, Prop. 2.15), we can prove $f_\theta(x)$ meets the necessary topological requirements. We verify the specific criteria below:

1. Global Bijection (One-to-One and Onto): Let $y = \log(x)$ be the input to the spline. Inside the knot vector $t_k \le y \le t_{m-k}$, the B-spline basis functions possess strictly local support and are globally non-negative ($B_{i,k-1}(y) \ge 0$) (de Boor, 2001, Ch. IX, Prop. ii). The partition of unity property ($\sum B_{i,k-1}(y) = 1$) (de Boor, 2001, Ch. IX, Prop. iv) guarantees the strict positivity (> 0) of at least one basis function at any given y. The knot increments are strictly positive. Hence, according to Proposition 1, the sign of the spline derivative $S'(y)$ depends only on the coefficient differences $c_i - c_{i-1}$. Under our architectural constraint $c_i > c_{i-1}$, it mathematically follows that $S'(y) > 0$. Outside the knot vector, the boundary slopes m_L and m_R match the strictly positive derivatives at the grid edges, ensuring $S'(y) > 0$ globally on $\mathbb{R}$. We evaluate the total derivative of the composite function with respect to the original eigenvalue $x \in \mathbb{R}_{>0}$ via the chain rule, which yields $f'_\theta(x) = S'(\log x) \cdot \frac{1}{x}$. Because $x > 0$ and $S'(y) > 0$, the gradient $f'_\theta(x) > 0$ is strictly bounded above zero, rendering the function strictly monotonically increasing and thus injective (one-to-one). Regarding surjectivity, the logarithmic mapping pushes the manifold boundary ($x \to 0^+$) to $y \to -\infty$. Because the linear extrapolation extends to infinity on both tails with strictly positive

slopes, $S(y) \to \pm\infty$ as $y \to \pm\infty$. The range covers the entire unconstrained real line $(-\infty, \infty)$, rendering it surjective (onto). Thus, $f_\theta(x)$ is a global bijection.

2. Smooth Forward Map (C^1 Continuity): Standard B-splines of degree $k \geq 3$ are at least C^2 continuous inside their knot grid. We clamp the terminal knots of the grid with multiplicity k to transition smoothly to the unconstrained domain. Following standard spline behavior for repeated knots(de Boor, 2001, Ch. IX), this multi-knot insertion intentionally reduces continuity at the edges, forcing the curve to interpolate the boundary control points directly. Consequently, the boundary derivatives are defined by the first and last segments of the control polygon, which are adopted by the linear extrapolation, at the boundaries t_k and t_{m-k}, to ensure a globally smooth forward mapping and a C^1 continuity at the transition points, by matching the tangent slopes m_L and m_R. Importantly, within the active knot grid, a cubic ($k = 3$) B-spline is C^2 smooth, fully supporting classical sectional curvature and parallel transport operations. The linear extrapolation tails, which restrict global smoothness to C^1, function strictly as a discrete computational fallback for extreme spectral outliers. Theoretically, by extending the grid boundaries $t_0 \to -\infty$ and $t_m \to \infty$, the linear tails vanish, and C^2 smoothness holds globally on $\mathbb{R}_{>0}$. Higher order differentiability can easily be obtained by increasing k.

3. Smooth Inverse Map: The mapping requires a smooth inverse to be a diffeomorphism. By the Inverse Function Theorem (Lee, 2012, Appendix C, Thm. C.34), if a continuously differentiable mapping has an invertible total derivative (Jacobian) at every point, a C^1 smooth inverse is guaranteed to exist. For a scalar function $f_\theta : \mathbb{R}_{>0} \to \mathbb{R}$, the Jacobian is simply a 1×1 matrix $[f'_\theta(x)]$. The monotonic parameterization and the chain rule keep $f'_\theta(x) > 0$ globally, rendering the determinant non-zero and satisfying the invertibility condition. Consequently, $f_\theta(x)$ is a smooth bijection with a smooth inverse, forming a true global diffeomorphism (Lee, 2012, Ch. 2, p. 39).

Conclusion: Because $f_\theta(x)$ is the composition of the global diffeomorphism $\log(x)$ and the global C^1 diffeomorphism $S(y)$, it mathematically constitutes a global C^1 diffeomorphism from $\mathbb{R}_{>0}$ to $\mathbb{R}$. ■

B.2 Proof of theorem 2

Proof. Let $g \in \mathrm{Diff}^1_+(\mathbb{R})$ be an arbitrary target strictly monotonic C^1 diffeomorphism, and let $[a, b]$ be the compact interval covering the dataset's logarithmic spectrum. The Schoenberg variation-diminishing property of B-splines allows Schoenberg's spline operator to uniformly approximate $g(x)$ on $[a, b]$ (de Boor, 2001, Ch. XI, Eq. 33). This yields a spline $S_k(x) = \sum c_i B_{i,k}(x)$ over a finite knot vector T covering $[a, b]$.

Let $y = \log(x)$ denote the change of variables mapping the domain to $\mathbb{R}$, so that $f_\theta(x) = S(y)$ and, by the chain rule, $f'_\theta(x) = S'(y) \cdot \frac{1}{x}$. If the control coefficients are chosen as $c_i = g(e^{\xi_i})$, where ξ_i are the Greville abscissae (knot averages (de Boor, 2001, Ch. IX, Eq. 39; Ch. XI, Eq. 33)) in the y-domain, the spline uniformly converges to $g(x)$ as the knot spacing $h \to 0$. Since $g(x)$ strictly increases, any $\xi_i > \xi_{i-1}$ means the sampled control points satisfy $g(\xi_i) > g(\xi_{i-1})$. As established by de Boor (2001, Ch. XI), this shape-preserving operator maps monotone functions to monotone functions, which yields our architectural constraint $c_i > c_{i-1}$. (Note: In the discrete neural instantiation, the parameter ϵ in $c_i = c_{i-1} + \mathrm{softplus}(w_i) + \epsilon$ acts as a strictly positive lower bound. For true asymptotic universal approximation as $h \to 0$, it is required that $\epsilon \to 0$ such that $\epsilon < \min_i(g(\xi_i) - g(\xi_{i-1}))$.)

We use cubic B-splines (order $= 4$), and the target $g(x)$ is C^1 smooth. Under Schoenberg's operator, the uniform convergence applies simultaneously to the first derivative (Marsden, 1970; Gonska et al., 2010): $\sup_{x\in[a,b]} |S'_k(x) - g'(x)| \to 0$ as $h \to 0$.

The linear extrapolation of $f_\theta(x)$ maintains global topological bijectivity outside T. Inside $[a, b]$, the spline holds dense C^1 approximation capacity. The function f_θ and its derivative f'_θ entirely parameterize the SPM pullback metric tensor and its Daleckĭĭ-Kreĭn structural gradient. This simultaneous C^1 convergence ensures SPM universally approximates the data-driven Riemannian metric tensor for a given dataset. ■

B.3 PROOF OF COROLLARY 1

Proof. Let $y = \log(x)$ be the input to the parameterized spline $S(y)$. Because standard backpropagation relies on the first derivative to compute the Daleckĭĭ-Kreĭn gradient, the C^1 convergence guaranteed by Theorem 2 provides the necessary and sufficient topological smoothness to functionally subsume these target metrics for deep learning optimization. We review the functional mapping $S(y)$ needed for SPM to recover existing geometries:

1. Logarithmic Metrics (LE, LC): These use the standard natural logarithm $f(x) = \log(x)$. Equating this to the SPM composition gives $S(\log x) = \log x$, simplifying to $S(y) = y$. Hence, SPM can easily replicate these by learning the identity function across the control polygon.

2. Power-Cholesky Metric (PCM): Parametric power metrics rely on fractional powers $f(x) = x^\theta$ for $\theta > 0$. The SPM composition becomes $S(\log x) = x^\theta$. Substituting $x = e^y$, the required spline mapping is $S(y) = e^{\theta y}$. The exponential function $e^{\theta y}$ is a strictly increasing C^∞ diffeomorphism. Theorem 2 ensures the monotonic B-spline can uniformly approximate the PCM geometry over any compact spectral domain.

An analysis of ALEM is not presented due to its diffeomorphic limitations.

Conclusion: As a universal approximator for diffeomorphisms, the Spline-Pullback Metric acts as a superset of existing analytical metrics. ■

B.4 PROOF OF PROPOSITION 2

Proof. First, we evaluate topological safety. Learning a non-linear metric mapping via high-degree polynomial regression usually triggers Runge's phenomenon. This generates macro-oscillations at the boundaries, forcing the first derivative to cross zero ($f'_\theta(x) \leq 0$) and breaking the bijection. The SPM generator avoids this by applying Schoenberg's Variation-Diminishing operator to cubic B-splines (de Boor, 2001, Ch. XI, Cor. 28). The parameterization requires a monotonic control polygon ($c_i > c_{i-1}$). The variation-diminishing property then stops the continuous curve $f_\theta(x)$ from oscillating. We clamp the boundary knots to interpolate the terminal control points without overshoot. This yields a global diffeomorphism safe from Runge's phenomenon.

Next, we establish geometric expressivity. Let $\Delta c_i = c_i - c_{i-1}$ denote the first-order finite difference of the control points. To ensure monotonicity, we parameterize the control polygon via a cumulative sum of strictly positive steps: $c_i = c_0 + \sum_{j=1}^{i} \Delta c_j$. The steps themselves are parameterized as unconstrained neural network weights processed through a strictly positive activation, $\Delta c_j = \text{softplus}(w_j) + \epsilon$. According to Proposition 1, enforcing $\Delta c_i > 0$ bounds the first derivative strictly above zero ($f'_\theta(x) > 0$).

However, the second-order finite difference, $\Delta^2 c_i = \Delta c_{i+1} - \Delta c_i$, remains unconstrained ($\Delta^2 c_i \in \mathbb{R}$), as the underlying weights w_j can be optimized to any real value. Because the control points of the first derivative spline curve are proportional to Δc_i, the control polygon of this derivative curve, though strictly above the x-axis, is free to undergo sign changes with respect to any line $y = \lambda, \lambda > 0$, giving freedom to the underlying spline curve to do so (de Boor, 2001, Ch. XI, Cor. 28). This allows the second spline derivative $f''_\theta(x)$ to cross zero and dynamically form inflection points. Unlike classical metrics such as the Log-Euclidean ($f''(x) = -1/x^2$) or Power-Cholesky, which impose a fixed global concavity or convexity, the unconstrained $\Delta^2 c_i$ in SPM enables Localized Spectral modelling. The metric can learn "S-shaped" deformations that compress noise and expand specific eigenvalue subspaces. This flexibility is impossible with rigid algebraic priors ■

B.5 PROOF OF THEOREM 3

*Proof.*According to Theorem 1, the scalar generator $f_\theta(x)$ constitutes a C^1 global diffeomorphism from $\mathbb{R}_{>0}$ to $\mathbb{R}$. We elevate this scalar mapping to the matrix manifold via the calculus of Primary Matrix Functions, defining the forward mapping $\phi_{S-SPM}(S) = U f_\theta(\Sigma) U^\top$ (Bhatia, 1997, Ch. V.1).

1. Matrix Bijection: Because f_θ is a scalar bijection, its inverse $f_\theta^{-1} : \mathbb{R} \to \mathbb{R}_{>0}$ uniquely exists. By the same calculus, this induces a well-defined inverse matrix function $\phi^{-1}(Y) = U f_\theta^{-1}(\Sigma_Y) U^\top$. Applying this inverse yields the identity mappings $\phi^{-1}(\phi(X)) = X$ and $\phi(\phi^{-1}(Y)) = Y$. The first

identity ensures injectivity (as $\phi(X_1) = \phi(X_2) \implies X_1 = X_2$). The second ensures surjectivity (as every Y has a valid pre-image), without repeated eigenvalues posing any challenge, as observed in rank-based metrics. This two-sided inverse guarantees that the matrix mapping ϕ_{S-SPM} is a strict global bijection between the SPD manifold $\mathcal{S}^n_{++}$ and the space of symmetric matrices $\mathcal{S}^n$.

2. Matrix Smoothness: The Daleckĭĭ-Kreĭn theorem (Bhatia, 1997, Thm. V.3.3) states that applying a C^1 smooth scalar function to a symmetric matrix creates a Primary Matrix Function that is continuously Fréchet differentiable. Both f_θ and f_θ^{-1} are C^1 smooth scalars. As a result, the forward matrix mapping ϕ and its inverse ϕ^{-1} are globally C^1 smooth. The matrix mapping inherits the scalar generator's strict bijection and smooth invertibility. This meets all criteria for a global diffeomorphism and induces a valid bi-invariant Riemannian metric.

3. Congruence Invariance: Because the forward mapping $\phi_{S-SPM}(S)$ is defined via the spectral decomposition, it constitutes a Primary Matrix Function (Bhatia, 1997, Ch. V). By definition, such functions commute with orthogonal congruence transformations: $\phi(RSR^\top) = R\phi(S)R^\top$ for any orthogonal R. Differentiating this identity, the pushforward satisfies $\mathcal{D}\phi_{RSR^\top}(RVR^\top) = R(\mathcal{D}\phi_S(V))R^\top$. Substituting into the pullback Frobenius inner product (Eq. 1) and using the unitary invariance of the Frobenius norm ($\langle AXB, AYB\rangle_F = \langle X, Y\rangle_F$ for orthogonal A, B):

$$g_{RSR^\top}(RVR^\top, RWR^\top) = \langle R(\mathcal{D}\phi_S(V))R^\top, R(\mathcal{D}\phi_S(W))R^\top\rangle_F = \langle \mathcal{D}\phi_S(V), \mathcal{D}\phi_S(W)\rangle_F = g_S(V, W) \tag{15}$$

the orthogonal matrices R cancel, establishing that g_{SPM} is invariant under orthogonal congruence. This is distinct from bi-invariance in the Lie-group sense, which requires a group operation on $\mathcal{S}^n_{++}$ not induced by congruence alone. As Corollary 2 shows, ϕ_{S-SPM} transports the ambient additive structure onto $\mathcal{S}^n_{++}$ via $S_1 \oplus_\phi S_2 := \phi^{-1}(\phi(S_1)+\phi(S_2))$; it is with respect to this specific transported Abelian operation that g_{SPM} is bi-invariant. ■

B.6 Proof of proposition 3

Proof. This closed-form gradient is a direct consequence of applying the Fréchet differential of symmetric matrix functions (Bhatia, 1997, Thm. V.3.3) to the Frobenius inner product chain rule $d\mathcal{L} = \langle \nabla_X \mathcal{L}, dX\rangle$. By substituting the analytical B-spline basis evaluations for f_θ and f'_θ into the divided difference matrix K, we obtain the structural gradient required for Riemannian backpropagation. ■

B.7 Justification of Computational Safeguard Protocol 1

Proof. Let the perturbed eigenvalues be $\tilde{\lambda}_i = \lambda_i + \epsilon_i$. Because the original eigenvalues are sorted (generally intrinsically by computational eigenvalue solvers) ($\lambda_{i+1} - \lambda_i \geq 0$), and $\vec{\epsilon}$ enforces a strict minimal growth $\epsilon_{i+1} - \epsilon_i = \delta > 0$, we mathematically guarantee that for all $i < j$:

$$\tilde{\lambda}_j - \tilde{\lambda}_i = (\lambda_j - \lambda_i) + (\epsilon_j - \epsilon_i) \geq \delta > 0 \tag{16}$$

This establishes a minimal spectral gap $|\tilde{\lambda}_i - \tilde{\lambda}_j| \geq \delta$ for all $i \neq j$. Applying this minimum gap to the Daleckĭĭ-Kreĭn matrix K, the off-diagonal elements are strictly bounded:

$$\max_{i \neq j} |K_{ij}| \leq \frac{|f_\theta(\tilde{\lambda}_i) - f_\theta(\tilde{\lambda}_j)|}{\delta} \tag{17}$$

The B-spline output f_θ is continuously differentiable and bounded smoothly by its control points, meaning the numerator remains finite. The denominator is strictly positive and bounded below by δ. Together, this permanently bounds the gradient norm. Capping the Lipschitz constant of the backward pass this way protects the Riemannian optimization from eigenvalue degeneracy and $0/0$ gradient explosions. As $\delta \to 0$, the perturbed off-diagonal elements in the matrix naturally converge to the true analytical derivative $f'_\theta(\lambda)$ through the finite-difference limit. ■

B.8 Proof of theorem 4

Proof. We establish this hierarchically through the product space. The mapping $\phi_{\mathcal{L}}$ acts as the identity on the Euclidean subspace $\mathcal{SL}^n$ and as the scalar diffeomorphism f_θ (established in Theorem 1) on the diagonal subspace $(\mathbb{R}_{>0})^n$. Because both subspace mappings are C^1 diffeomorphisms, their direct product $\phi_{\mathcal{L}}$ is a global diffeomorphism from $\mathcal{L}^n_{++}$ to $\mathcal{L}^n$. Furthermore, the standard Cholesky

factorization Chol : $\mathcal{S}^n_{++} \to \mathcal{L}^n_{++}$ is a well-established global diffeomorphism (Lin, 2019). Because the composition of two diffeomorphisms is itself a diffeomorphism (Lee, 2012, Ch. 2, Prop. 2.15), $\phi_{C-SPM} = \phi_{\mathcal{L}} \circ \text{Chol}$ forms a global bijection from $\mathcal{S}^n_{++}$ to the Euclidean space $\mathcal{L}^n$ with a smooth inverse, establishing a valid pullback geometry. ■

B.9 Proof of theorem 5

Proof. Let $\phi : \mathcal{S}^n_{++} \to \mathcal{E}$ be the SPM global diffeomorphism mapping the SPD manifold to the Euclidean vector space $\mathcal{E}$ (where $\mathcal{E} = \mathcal{S}^n$ for S-SPM, and $\mathcal{E} = \mathcal{L}^n$ for C-SPM). The target space $\mathcal{E}$ is an unconstrained flat vector space because our scalar generator f_θ is globally surjective onto $\mathbb{R}$. The Christoffel symbols of the Levi-Civita connection vanish in the transformed coordinates because g_{SPM} is defined as the isometric pullback of the standard Euclidean inner product $\langle \cdot, \cdot \rangle_F$ through ϕ. As a result, the Riemannian curvature tensor becomes zero globally, rendering the manifold globally flat (analogous to the isometric preservation of zero sectional curvature established in Log-Euclidean and Log-Cholesky Spaces (Arsigny et al., 2006; Lin, 2019) while upgrading the isometry to a learnable parameterization. ■

B.10 Proof of theorem 6

Proof. The Fréchet Mean on a Riemannian manifold is defined as the minimizer of the sum of squared geodesic distances: $P^* = \arg\min_{P \in \mathcal{S}^n_{++}} \sum w_i d^2_{SPM}(P, P_i)$. Because the manifold $(\mathcal{S}^n_{++}, g_{SPM})$ has zero sectional curvature (Theorem 5) and is simply connected, the Fréchet variance function is strictly convex(Karcher, 1977). Therefore, the unique global minimum exists and maps identically to the weighted Euclidean mean in the pullback space (Pennec et al., 2006):

$$\text{WFM}_{SPM}(\{P_i\}, \{w_i\}) = \phi^{-1}\left(\sum_{i=1}^{N} w_i \phi(P_i)\right) \tag{18}$$

■

B.11 Proof of theorem 7

Proof. In a manifold with zero sectional curvature, parallel transport between any two points is independent of the curve connecting them(Lee, 2018). Under the SPM isometry, parallel transport equates to projecting the tangent vector V to the flat Euclidean space at A, translating it to B (which acts as the identity operation in a flat vector space), and projecting it back to the tangent space at B:

$$\text{PT}_{A \to B}(V) = \mathcal{D}\phi_B^{-1}\Big(\mathcal{D}\phi_A(V)\Big) \tag{19}$$

where $\mathcal{D}\phi$ and $\mathcal{D}\phi^{-1}$ are the forward and inverse pushforward differentials of the SPM mapping. ■

B.12 Proof of theorem 8

Proof. Let $\tilde{A}_k$ represent the unconstrained Euclidean weights of the classifier corresponding to the tangent vector A_k, and let $\phi \in \{\phi_{S-SPM}, \phi_{C-SPM}\}$. By substituting the closed-form SPM Riemannian Logarithm into the MLR probability function, the Riemannian geometry collapses into a flat inner product. Given an input $X \in \mathcal{S}^n_{++}$, the SPM-MLR is evaluated as:

$$p(y = k|X) \propto \exp\Big(\langle \tilde{A}_k, \phi(X) - \phi(P_k) \rangle_F\Big) \tag{20}$$

■

C Riemannian and algebraic operators on the SPM manifold

The Spline-Pullback Metric framework functions as a global diffeomorphism, mapping the curved manifold $\mathcal{S}^n_{++}$ to flat Euclidean spaces, the space of symmetric matrices $\mathcal{S}^n$ for S-SPM, and lower triangular matrices $\mathcal{L}^n$ for C-SPM. Consequently, all fundamental Riemannian and algebraic operators can be derived in closed form through the pullback mechanism.

C.1 Differentials and tangent projections

For an SPD matrix $P \in \mathcal{S}_{++}^n$, let $P = U\Sigma U^\top$ be its eigendecomposition and $P = LL^\top$ be its Cholesky factorization. The differential map projecting a tangent vector $V \in \mathcal{T}_P\mathcal{S}_{++}^n$ to the respective representation spaces depends on the chosen SPM instantiation:

1. **Spectral Differential (S-SPM):** $\mathcal{D}_{eig}(V) = U[K \odot (U^\top V U)]U^\top$, where K is the Daleckĭĭ-Kreĭn matrix derived in Proposition 2.
2. **Cholesky Differential (C-SPM):** $\mathcal{D}_{chol}(V) = \lfloor L(L^{-1}VL^{-\top})\rfloor + \frac{1}{2}\text{diag}(L^{-1}VL^{-\top})L$, adhering to standard Cholesky pushforward differentials (Lin, 2019).

C.2 Closed-form Riemannian operators

Let $\phi \in \{\phi_{S-SPM}, \phi_{C-SPM}\}$ denote the chosen global diffeomorphism, and let $\mathcal{D}_\phi$ denote its respective differential map. For any $P, Q \in \mathcal{S}_{++}^n$ and tangent vector $V \in \mathcal{T}_P\mathcal{S}_{++}^n$, the globally complete Riemannian operators are:

$$\text{Geodesic Path } (t \in [0,1]): \quad \gamma_{SPM}(P,Q;t) = \phi^{-1}\Big((1-t)\phi(P) + t\phi(Q)\Big) \tag{21}$$

$$\text{Riemannian Logarithm:} \quad \text{Log}_P(Q) = \mathcal{D}_\phi^{-1}\Big(\phi(Q) - \phi(P)\Big) \tag{22}$$

$$\text{Riemannian Exponential:} \quad \text{Exp}_P(V) = \phi^{-1}\Big(\phi(P) + \mathcal{D}_\phi(V)\Big) \tag{23}$$

$$\text{Geodesic Distance:} \quad d_{SPM}^2(P,Q) = \|\phi(P) - \phi(Q)\|_F^2 \tag{24}$$

D Extended Adversarial Classification Results

The defined monotonic derivatives of fixed geometries limit their ability to learn. To illustrate this, we create a synthetic binary classification dataset comprising 1000 4×4 SPD matrices. The eigenspectrum is partitioned into four alternating bands: low-frequency noise $\lambda \in [0.1, 2.0]$, low-frequency signal, significant high-frequency noise $\lambda \in [12.0, 25.0]$, and high-frequency signal. The orthogonal matrices U are chosen at random, which means that the linear probe can only use the learned spectral distances.

As detailed in Table 5, all non-adaptive classical baselines suffer from severe underfitting. Because their derivatives are fixed (such as $f'(\lambda) = 1/\lambda$ for LE), they face a trade-off: they must either over-amplify the low-frequency noise or over-amplify the high-frequency noise, blinding the classifier to the discriminative signal bands.

Table 5: Classification accuracy on the alternating eigenvalue bands dataset. SPM successfully attains perfect performance, whereas rigid classical metrics lack the geometric capacity to separate the bands.

Metric Family	Geometry	Train Acc.	Test Acc.
Rigid Baselines	Log-Euclidean (LE)	65.50%	66.00%
	Log-Cholesky (LC)	63.50%	62.90%
	Power-Cholesky (PCM, $\theta = 0.5$)	64.50%	63.90%
Ours	C-SPM	71.25%	69.90%
	S-SPM	**100.00%**	**100.00%**

On the contrary, our proposed SPM achieves 100% accuracy in both train and test sets, signalling the lack of overfitting. As visualized in Figure 2, SPM learns a staircase-function-like geometry to expand and minimize signal and noise, respectively. Moreover, the performance gap between SPM (100%) and C-SPM (69.90%) is attributable to the fact that the Cholesky decomposition mixes the eigenvalues with the randomized rotation matrix U. Hence, the success of C-SPM should be gauged by fairly comparing it with LC, over which it attains a prominent performance gain.

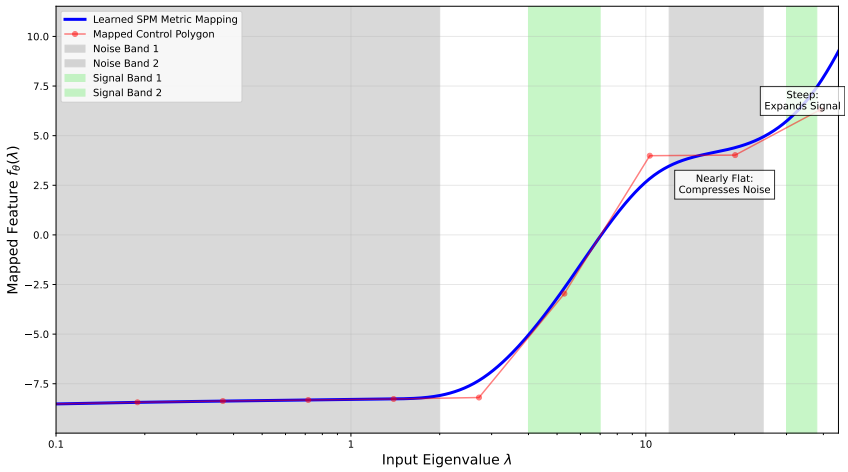

Figure 2: S-SPM expressiviness and universal approximation capacity mapped over a logarithmic eigenspectrum. To achieve perfect class separability, the learned s-SPM performs *Localized Spectral Modelling*. It drives its derivative toward zero to horizontally compress the overlapping noise bands (gray) while simultaneously ramping vertically to stretch the discriminative signal bands (green). This sequence of dynamic inflections is mathematically impossible under classical metric priors.

E Hyperparameter Stability Ablation

Unlike metrics where hyperparameters dictate the global topological shape (e.g., fractional powers), SPM's hyperparameters (Grid Size and Spline Order) merely control curve resolution. To demonstrate SPM's stability, we ablated the Linear Probe model on the FPHA dataset. As detailed in Table 6, accuracy and runtime remain remarkably consistent across 9 diverse configurations.

Table 6: SPM Hyperparameter Ablation (FPHA Linear Probe, NVIDIA T4 GPU).

Grid Size	Spline Order	Accuracy	Time/Epoch (s)
5	2	0.8661	1.33
5	3	0.8487	1.33
5	4	0.8520	1.35
10	2	0.8713	1.32
10	**3**	**0.8574**	**1.33**
10	4	0.8522	1.33
20	2	0.8617	1.32
20	3	0.8548	1.34
20	4	0.8678	1.35

F Extended Large-Scale Benchmark Results

As detailed in Tables 7 and 8, S-SPM achieved statistically significant superior performance relative to established baselines under these controlled setups. We used a Linear Probe (AI-NLP) and SPDNet (AFEW) to ensure the raw metric performance is not masked by heavy base models.

G Riemannian Metric Explainable AI (RMXAI)

A central requirement for Riemannian Metric Explainable AI (RMXAI) is that the learned geometric deformations represent stable and generalizable data characteristics. In order to empirically verify the stability of the proposed method, the final learned Spectral-SPM (S-SPM) diffeomorphisms were recorded in all 5 independent folds of the Stratified Cross-Validation on the Radar dataset.

Table 7: AI-NLP DAIR Dataset ($N = 20,000$). Evaluated via McNemar's tests with FDR adjustment.

Metric	Test Acc	Time/Ep (s)	Peak VRAM (MB)	95% CI of Δ vs S-SPM	FDR p-value	Significant
S-SPM (Ours)	**0.8192**	1.73	2012.65	*Reference*	-	-
ALEM	0.8072	1.19	2012.65	$[+0.11\%, +2.31\%]$	0.0360	Yes
LE	0.7872	1.08	2019.51	$[+1.94\%, +4.44\%]$	< 0.001	Yes
AIRM	0.7772	2.91	2026.38	$[+2.81\%, +5.56\%]$	< 0.001	Yes
LC	0.7653	0.71	116.38	$[+3.97\%, +6.75\%]$	< 0.001	Yes
PCM	0.7644	0.73	116.38	$[+4.08\%, +6.86\%]$	< 0.001	Yes
C-SPM (Ours)	0.7503	1.31	116.38	-	-	-

Table 8: AFEW Dataset (400×400 matrices). Evaluated via McNemar's tests with FDR adjustment.

Metric	Test Acc	Time/Ep (s)	Peak VRAM (MB)	95% CI of Δ vs S-SPM	FDR p-value	Significant
S-SPM (Ours)	**0.3046**	7.92	2048.69	*Reference*	-	-
C-SPM (Ours)	0.2668	4.34	2509.72	-	-	-
PCM ($\theta = 0.75$)	0.2372	4.12	2509.72	$[+1.08\%, +12.40\%]$	0.0291	Yes
AIRM	0.2372	18.29	2510.33	$[+1.35\%, +12.40\%]$	0.0291	Yes
LE	0.2264	7.75	2048.69	$[+2.16\%, +13.48\%]$	0.0140	Yes
LC	0.2183	4.11	2048.69	$[+3.23\%, +14.02\%]$	0.0078	Yes
ALEM	0.2022	7.72	2048.69	$[+4.31\%, +16.17\%]$	0.0065	Yes

Table 9: Computational Runtime Comparison (FPHA Linear Probe, NVIDIA T4 GPU). The dynamic B-spline parameterization introduces a negligible overhead of ~ 20 milliseconds per epoch compared to rigid analytical baselines, while bypassing the severe bottleneck of AIRM.

Metric Architecture	Time/Epoch (s)
Log-Euclidean (LE)	1.30
Log-Cholesky (LC)	1.30
ALEM (Learnable)	1.30
PCM (Learnable)	1.31
S-SPM (Ours)	**1.32**
C-SPM (Ours)	**1.32**
Global AIRM	1.59

As illustrated in Figure 3, despite being exposed to different data splits, the monotonic B-splines converged to consistent topological shapes, learning to compress noisy logarithmic eigenvalues. This demonstrates a stable geometric grouping across the folds, indicating that SPM optimization is stable and effectively captures the inherent spectral properties of the dataset.

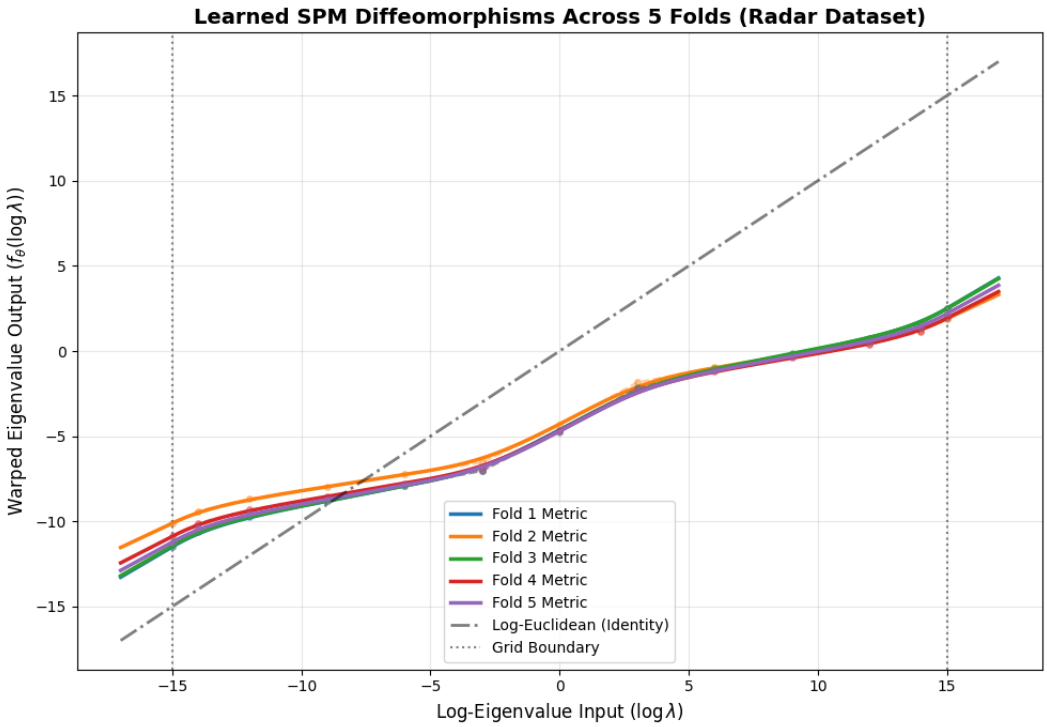

Figure 3: Learned Spectral-SPM diffeomorphisms across 5-fold cross-validation on the Radar signal processing dataset. The consistent convergence to a non-linear topology across independent folds highlights the stability of the SPM optimization and its ability to act as an interpretable diagnostic tool for spectral noise.

H Network Architectures and Implementation

H.1 Optimization and Training Protocols

Across all datasets and architectures, we maintained a uniform training protocol to prevent hyperparameter overfitting. To eliminate variance from stochasticity, a universal random seed (`seed=42`) was enforced across all hardware and data-loading workers. The PyTorch library(Paszke et al., 2019) was used for the implementations. All experiments were conducted on NVIDIA GPU T4 hardware. The full source code and complete training scripts will be publicly released on GitHub upon the acceptance of this manuscript. Comprehensive hyperparameter details and architectural configurations are provided below to ensure theoretical and algorithmic reproducibility:

- **Optimizer:** Riemannian Adam (`geoopt.optim.RiemannianAdam`)
- **Hyperparameters:** Learning rate $lr = 10^{-3}$, weight decay $= 10^{-4}$, and batch size $= 32$.
- **Gradient Norm:** For stable convergence, a maximum gradient norm of 2.0 was implemented.
- **Early Stopping:** Models were trained for a maximum of 100 epochs. We utilized an early stopping criterion with a patience of 15 epochs, monitored via the accuracy on a 20% internal validation split isolated from the training fold. The weights from the best-performing validation epoch were restored for final test-set inference.

H.2 NUMERICAL SAFEGUARDS

Implementation Detail: While Computational Safeguard Protocol 1 establishes theoretical Lipschitz bounds via direct spectral perturbation, standard automatic differentiation frameworks compute the Fréchet derivative internally and fault on exact degeneracies before the gradient matrix can be manually bounded. We therefore implement a 2-tier safeguard: (i) an input-matrix diagonal perturbation ($X \to X + \text{diag}(\vec{\delta})$, $\|\vec{\delta}\|_2 = 2 \times 10^{-8}$) applied before the eigensolver, almost always sufficient to break degeneracy; (ii) if a gap $|\lambda_i - \lambda_j| < 10^{-7}$ persists after decomposition, the sorted-eigenvalue perturbation of Protocol 1 is applied directly. Because tier (i)'s perturbation has operator norm 2×10^{-8}, Weyl's Eigenvalue Perturbation Inequality guarantees every resulting eigenvalue deviates from its unperturbed value by no more than 2×10^{-8}.

As noted in the main text, full-rank kinematic SPD matrices (e.g., HDM05) are severely rank-deficient and frequently precipitate `LinAlgError` crashes during standard 32-bit eigendecomposition or Cholesky factorization. All models in our suite (including our stabilized implementations of the baselines) were equipped with the following numerical armor:

- **Symmetry Enforcement & 64-Bit Upcasting:** Prior to any decomposition, matrices were strictly symmetrized ($X = \frac{1}{2}(X + X^\top)$) and upcast to `Float64` precision.
- **Eigenvalue Jitter (Tie-Breaking):** To prevent degenerate eigenvalue collapse, a uniformly spaced infinitesimal tie-breaker (`torch.linspace`$(10^{-8}, 2 \cdot 10^{-8}, n)$) was added to the diagonal prior to decomposition.
- **Cholesky Fallback Mechanism:** For Cholesky-based metrics, we implemented a dual-stage factorization. The primary attempt utilized a standard $10^{-4}I$ jitter. If the PyTorch backend threw a `LinAlgError`, the module dynamically intercepted the exception, fell back to an eigendecomposition, hard-clamped all eigenvalues to a minimum of 10^{-4}, reconstructed the matrix, and successfully performed the Cholesky factorization on the sanitized matrix.

H.3 SPLINE-PULLBACK METRIC (SPM) PARAMETERS

The SPM mapping was parameterized using a Cox-de Boor B-spline evaluated on a clamped knot vector.

- **Grid Configuration:** Grid Size $= 10$, Spline Order $= 3$ (cubic), Domain Range $= [-15.0, 15.0]$. To ensure a fair and controlled comparison, we fix this spline configuration across all models and datasets for both S-SPM and C-SPM.
- **Monotonicity Constraint:** The control points c_i were dynamically generated via cumulative summation of softplus-activated weights, ensuring C^1 monotonic bijectivity.
- **Base-Metric Initialization:** For experiments utilizing the identity initialization, the raw weights of the spline were initialized such that the resulting curve approximated $y = x$. The first control point was locked to the grid minimum (-15.0), and subsequent step sizes were initialized to $\ln(\exp(h) - 1.0)$, where h is the grid spacing. The choice of spline initialization, either identity-based or random, is determined through inner-fold validation and a priori hypotheses about how model depth interacts with the geometric prior of the metric. This approach avoids data leakage while keeping the selection process consistent.

H.4 DATASETS

HDM05(Müller et al., 2007): A 117-class kinematic action recognition dataset (93×93 SPD matrices). Evaluated using 5-Fold Stratified Cross-Validation.
FPHA(Garcia-Hernando et al., 2018): First-Person Hand Action dataset, containing irregular kinematic tracking data (63×63 SPD matrices). Evaluated using the standard predefined train-test split.
Radar(Brooks et al., 2019): A signal processing covariance dataset (20×20 SPD matrices). Evaluated using 5-Fold Stratified CV.

H.5 ARCHITECTURE CONFIGURATIONS

In all configurations, the output of the metric mapping layer was vectorized, passed through a `BatchNorm1d` layer to stabilize Euclidean convergence, and fed into a terminal Linear classifier.

1. Direct Linear Probe: In order to gauge the intrinsic learning capability of a metric, it is imperative to isolate it from the compensatory effect of highly parameterized deep learning layers and classifiers. Hence, in this experiment, we project the uncompressed SPD matrices directly into the tangent space using respective metric before feeding them into a single linear layer. High accuracy here proves the metric itself has separated the classes.

- **Flow:** Input $\rightarrow$ Metric Projection $\rightarrow$ Flatten $\rightarrow$ BatchNorm1d $\rightarrow$ Linear Classifier
- **Global AIRM Baseline:** Because AIRM lacks a standard direct projection, the Linear Probe for the AIRM baseline whitened the data using the Global Fréchet Mean of the training fold, $M_{global}^{-1/2} X M_{global}^{-1/2}$, before projecting to the tangent space via the matrix logarithm.

2. SPDNet (SPD-MLR): We adopt the standard SPDNet building blocks (Huang & Van Gool, 2017): Bilinear Mapping (BiMap) and Riemannian Eigenvalue Rectification (ReEig). Following recent advancements in generalized Riemannian classifiers (Chen et al., 2024b), the final classification layer of this backbone is implemented as an SPD Multinomial Logistic Regression (SPD MLR) to directly evaluate the performance of each metric mapping. To isolate the effect of spatial filtering from information bottlenecking, this architecture preserves the full-rank manifold topology without dimensionality reduction. We denote the architecture geometries by an array $[d_{in}, d_{out}]$.

- **HDM05:** $[93, 93]$
- **FPHA:** $[63, 63]$
- **Radar:** $[20, 20]$
- **Architecture:** Input $\rightarrow$ BiMap $\rightarrow$ ReEig $\rightarrow$ Metric Projection $\rightarrow$ Flatten $\rightarrow$ BatchNorm1d $\rightarrow$ Linear Classifier

3 Deep RResNet Following Katsman et al. (2023), the Riemannian Residual Network integrates residual connections within the manifold constraints. To test the metrics in bottlenecked environments, this architecture applies a dimensionality-reducing BiMap prior to the residual blocks.

- **HDM05:** $[93, 30]$
- **FPHA:** $[63, 33]$
- **Radar:** $[20, 8]$
- **Architecture:** Input $\rightarrow$ BiMap $\rightarrow$ ReEig $\rightarrow$ RResNet Block $\times$ 2 $\rightarrow$ Metric Projection $\rightarrow$ Flatten $\rightarrow$ BatchNorm1d $\rightarrow$ Linear Classifier
- **AIRM Baseline Exception:** For the Deep RResNet configuration, the AIRM baseline utilized native AIRM RResNet blocks followed by a standard Log-Euclidean projection at the classification head, adhering to standard Riemannian ResNet practices, mirroring the official implementation.

Moreover, the θ-PCM baseline (Chen et al., 2026) applies a fractional power to the diagonals of the Cholesky factor, scaled by an isometry constant. To grant the baseline maximum competitive advantage, we adopted the optimal θ values reported by Chen et al. for each dataset.

However, because the original implementation relies on heavily dimension-reduced manifolds (like $[93 \rightarrow 30]$), applying these optimal θ values to the uncompressed, full-rank Linear Probe and SPDNet architectures frequently resulted in numerical divergence (NaNs) or `LinAlgError` exceptions due to the evaluation of fractional powers on microscopic eigenvalues. Our selection protocol was as follows: evaluate the author's optimal θ; if the forward pass collapsed, iteratively evaluate the next-best reported θ until stability was achieved.

- **Radar:** Linear Probe & SPDNet ($\theta = 0.5$), RResNet ($\theta = -1.0$).
- **HDM05:** All architectures ($\theta = 0.5$).
- **FPHA:** Linear Probe & SPDNet ($\theta = 0.75$), RResNet ($\theta = 0.5$).

I THEORETICAL VULNERABILITIES OF RANK-DEPENDENT MATRIX MAPPINGS

Recent works, such as the Adaptive Log-Euclidean Metrics (ALEM) Chen et al. (2024d), propose learning parameterized metrics by applying distinct scalar functions to eigenvalues based on their sorted indices. For an SPD matrix $S = U\Sigma U^\top$, ALEM defines a rank-dependent matrix logarithm as:

$$\phi_{ALEM}(S) = U\text{diag}(\log_{a_1}(\sigma_1), \ldots, \log_{a_n}(\sigma_n))U^\top \tag{25}$$

where a_i are independently learned base parameters for the i-th sorted eigenvalue. We demonstrate mathematically that this mapping is ill-defined (one-to-many) at spectral singularities and breaks standard backpropagation paradigms.

I.1 PROOF OF ILL-DEFINEDNESS VIA EIGENSPACE NON-UNIQUENESS

Consider a simple 2×2 SPD matrix S with a repeated eigenvalue (a spectral singularity):

$$S = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2I \tag{26}$$

Because $\sigma_1 = \sigma_2 = 2$, the eigenspace spans all of $\mathbb{R}^2$. The eigenvector matrix U is not unique; any 2×2 orthogonal matrix is a valid basis. Consider two valid, solver-generated eigendecompositions $S = U_1\Sigma U_1^\top$ and $S = U_2\Sigma U_2^\top$, where:

$$U_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad U_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \tag{27}$$

Both U_1 and U_2 perfectly reconstruct S. Now, assume the ALEM network has learned distinct parameters for the first and second eigenvalue ranks, such that $\log_{a_1}(2) = 1$ and $\log_{a_2}(2) = 2$. The transformed eigenvalue matrix is thus:

$$\bar{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \tag{28}$$

Reconstructing the output matrix using the first basis U_1 yields:

$$Y_1 = U_1\bar{\Sigma}U_1^\top = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \tag{29}$$

However, reconstructing the output matrix using the second valid basis U_2 yields:

$$Y_2 = U_2\bar{\Sigma}U_2^\top = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 1.5 \end{bmatrix} \tag{30}$$

We observe that $Y_1 \neq Y_2$. The same input matrix S has been mapped to two entirely different output matrices. Because the transformation breaks the isotropic symmetry of the repeating eigenvalues, the arbitrary rotation chosen by the computational solver does not cancel. Therefore, ALEM is a one-to-many mapping, failing the fundamental definition of a mathematical function.

In contrast to rank-dependent approaches, our proposed method (SPM) avoids these failures by strictly adhering to the calculus of Primary Matrix Functions. SPM applies a universally rank-invariant, strictly monotonic scalar spline $f_\theta : \mathbb{R}_{>0} \to \mathbb{R}$ to all eigenvalues equally, independent of their sorted index.

I.1.1 WELL-DEFINEDNESS AT SPECTRAL SINGULARITIES

Revisiting the problematic $S = 2I$ case with our proposed SPM, the eigenvalues are identically $\sigma_1 = \sigma_2 = 2$. Because our function is rank-invariant, it evaluates to a single scalar $c = f_\theta(2)$ for both dimensions. The transformed eigenvalue matrix becomes a scalar multiple of the identity:

$$\bar{\Sigma}_{SPM} = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix} = cI \tag{31}$$

When we reconstruct the output matrix using the two different, solver-generated eigenvector bases U_1 and U_2, we obtain:

$$Y_1 = U_1(cI)U_1^\top = c(U_1U_1^\top) = cI \tag{32}$$

$$Y_2 = U_2(cI)U_2^\top = c(U_2U_2^\top) = cI \tag{33}$$

Because $UU^\top = I$ is a universal axiom for any orthogonal matrix, the arbitrary rotations mathematically cancel themselves. We achieve $Y_1 = Y_2 = cI$.

I.2 Topological constraints of Cholesky Power Metric (PCM)

Recent methods have attempted to avert the computational cost of spectral decompositions by utilizing the Cholesky manifold, notably the Power-Cholesky Metric (PCM) Chen et al. (2026). PCM defines a product geometry where the strictly lower triangular elements of the Cholesky factor L remain Euclidean, while a power function x^θ is applied to the positive diagonal elements.

While this avoids eigenvector ambiguity, it introduces a severe topological limitation: a lack of global surjectivity onto the unconstrained tangent space.

To allow for unconstrained network optimization, a valid Riemannian mapping must project the SPD manifold diffeomorphically onto the entirety of the unconstrained Euclidean space $\mathbb{R}$. The matrix logarithm achieves this because the scalar logarithm surjectively maps $(0, \infty) \to (-\infty, \infty)$.

In contrast, PCM relies on the mapping $\phi_\theta(L_{ii}) = L_{ii}^\theta$. For any $L_{ii} > 0$ and $\theta > 0$, the image is strictly $(0, \infty)$. Consequently, the Riemannian logarithm for PCM on the diagonal is structurally bounded. As defined by the authors, for diagonal elements $K_{ii}, L_{ii} > 0$:

$$\mathrm{Log}_L(K)_{ii} = \frac{1}{\theta} L_{ii} \left[\left(\frac{K_{ii}}{L_{ii}} \right)^\theta - 1 \right] \tag{34}$$

As $K_{ii} \to 0$, the tangent vector is strictly bounded from below by $-\frac{1}{\theta} L_{ii}$. It is mathematically impossible for the PCM logarithm to map to tangent vectors that are more negative than this bound. Because the mapping is not surjective onto $\mathbb{R}$, PCM does not constitute a true global diffeomorphism to an unconstrained space, forcing its geodesics to be only locally defined.

Because the tangent space is bounded, unconstrained updates may project the diagonal elements outside the valid manifold domain. To prevent computational failure, the authors propose a "numeric trick" (Remark 3.9 in Chen et al. (2026)) utilizing a hard clamp: $d_i \leftarrow \max(d_i, \epsilon)$.

The $\max$ function is a non-injective (many-to-one) operator; any pre-image value $x \leq \epsilon$ is mapped to the same output. Thus, bijectivity is lost. The gradient of this clamping operation evaluates to zero for all $x < \epsilon$, limiting backpropagation. Thus, though PCM is a valid geodesically incomplete metric, its utility in unconstrained optimization, where unconstrained tangent-space operations must be mapped back to the manifold, is highly limited.

I.2.1 Resolution via surjective spline generators (SPM)

Our proposed method avoids these topological failures. SPM, being a vaild diffeomorphism, achieves an unconstrained tangent space, inducing geodesic completeness. The network can output any arbitrary real number in the tangent space without requiring non-injective clamping operations establishing a mathematically sound framework for SPD matrix learning.

J Validation against strictly diffeomorphic targets and C^0 baselines

To directly validate Theorem 2 under adversarial conditions, we evaluated SPM against strictly diffeomorphic targets rather than simple quadratics. As promised by our proofs, the constrained spline demonstrates excellent approximation capacity:

Table 10: SPM Approximation of Valid Diffeomorphisms.

Target Diffeomorphism	**Function MSE**	**Max Value Error (L_∞)**	**Max Deriv. Error (L_∞)**
$y = x + 0.8\sin(x)$	0.0001	0.0291	0.0720
$y = x + 5\tanh(x/2)$	0.0036	0.4898	0.8639
$y = 3\sinh(x/3)$	0.1287	2.1698	4.5616

To demonstrate why simpler C^0 monotone maps (such as Piecewise Linear Maps, PLMs) are unfit for metric learning, we conducted an eigenvalue optimization experiment targeting an eigenvalue of 6.0:

- **PLM Baseline:** Faced catastrophic gradient explosion at the C^0 "kink", resulting in a high loss spike (1.27×10^7) and step-gradients. It actively diverged after 700 epochs, plunging the eigenvalue to $-28,891.71$ (Final Loss: 8.80×10^5).
- **SPM (Ours):** Maintained smooth gradients via its C^2 interior regularity and successfully converged the eigenvalue to 6.13 (Final Loss: 0.208) in under 20 epochs.

This highlights the necessity of our mathematically constrained C^2 spline formulation over naïve monotonically constrained networks.

K LIMITATIONS AND FUTURE DIRECTIONS

While the Spline-Pullback Metric (SPM) eliminates the limitations of rigid mathematical priors, its high expressivity introduces specific limitations that invite future research. First, the expressivity of the diffeomorphism relies on the predefined B-spline grid resolution, spline order, and spectral boundaries. We mitigate this sensitivity in our current framework by fixing the grid bounds for all the experiments without using the dataset for further grid search or tuning to avoid optimization bias, which is glaringly prevalent in the literature. Furthermore, SPM's bounded linear extrapolation acts as a safety net to process any out-of-distribution spectral singularities without requiring exhaustive hyperparameter grid searches. Future work, however, will explore automated, data-driven knot placement, like utilizing the eigenspectrum information of the train and validation sets of the specific fold. Additionally, non-linear metric tensors may be sensitive to their initial geometric state. We currently resolve this concern of early-epoch instability by introducing specific Base-Metric (Identity) and Random initializations for stable convergence. Future research in Riemannian Metric Initialization (RMI), as stated before in the Discussion and Conclusion section, will expand upon this by deriving initialization distributions tailored specifically to the SPM manifold. Also, the computational learning requirement of the SPM, as opposed to predefined metrics, might appear as a computational limitation; however, this concern is more than offset by the time and computational need to manually choose and test the optimal metric for the specific dataset, which SPM learns automatically. The single-layer, parameter-efficient C-SPM further mitigates computational concerns by averting expensive eigendecompositions.

Secondly, as a universal diffeomorphism approximator, SPM possesses the theoretical capacity to overfit in data-scarce domains. While our empirical evaluations did not suffer from any pronounced overfitting, this limitation motivates the development of Riemannian Metric Regularization (RMR). Rather than relying solely on standard Euclidean weight decay, future RMR techniques will aim to explicitly ensure the integrity of the topological and diffeomorphic rules of the manifold. Finally, since anonymized datasets, provided by the original authors, were utilized in this work, no privacy or ethical issues arise.