NESTED INDUCTIVE BIAS FRAMEWORK FOR CURVATURE-ALIGNED SPD MANIFOLD LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

In Geometric Deep Learning, inductive biases serve two primary functions: enforcing manifold constraints and embedding relational priors. Currently, representation learning on SPD manifolds frequently relies on pullback Euclidean metrics to satisfy the former. While computationally efficient in avoiding domain boundary violations, these metrics induce a flat geometry that may fail to capture the intrinsic relational priors of datasets. While metrics such as the Poincaré metric are widely utilized to induce domain-aligned relational priors, generalizing them from standard vector representations to the SPD manifold has remained a challenge. To bridge this gap, we introduce a Nested Inductive Bias framework that utilizes a multistage diffeomorphic composition to formally pull back non-Euclidean target geometries onto the SPD manifold. Moreover, we identify that mapping unbounded Euclidean representations onto manifolds with finite injectivity radii induces a "Norm Mismatch," leading to degraded performance and diffeomorphic failure. We resolve this by introducing the Rational Conformal Metric (RCM), which preserves bijectivity as well as provides a strong regularization against outliers. Empirical evaluations on kinematic and signal processing benchmarks, together with synthetic experiments, demonstrate that deep manifold networks experience degradation in class separability due to curvature/norm mismatch between the metric and the intrinsic data distribution. Finally, our curvature-aligned Riemannian classifiers resolve the computational intractability of classical non-flat SPD metrics (like AIRM), achieving a 95% reduction in computational overhead.

1 INTRODUCTION

Symmetric Positive Definite (SPD) matrices have emerged as fundamental mathematical representations across different domains, enabling significant progress in medical image analysis (Chakraborty et al., 2020), neural signal decoding (Kobler et al., 2022; Li et al., 2024), kinematic modeling (Huang & Van Gool, 2017), and radio-frequency target identification (Brooks et al., 2019). Geometrically, these matrices reside on a strictly positive, convex conical manifold. Applying standard Euclidean computations directly to these representations violates the manifold boundary. To operate within these manifold boundaries, the field of Riemannian deep learning relies on metric tensors, such as the Affine-Invariant Riemannian Metric (AIRM) (Pennec et al., 2006) and Pullback Euclidean metrics like the Log-Euclidean Metric (LEM) (Arsigny et al., 2006). These metrics have enabled the adaptation of dense neural network components, such as residual connections (Katsman et al., 2023), normalization layers (Brooks et al., 2019), and universal parameterized metrics (Das et al., 2026), for manifold-valued data. Furthermore, the theoretical formalization of *deformed metrics* (Thanwerdas & Pennec, 2022) has demonstrated that valid Riemannian structures can be systematically generated via diffeomorphic pullbacks.

However, this classical paradigm reveals a gap in Geometric Deep Learning: the disconnect between enforcing *domain constraints* and embedding *relational priors*. For vector data (such as natural language), researchers often inject relational priors, such as Hyperbolic geometry (Ganea et al., 2018) and the unified κ-stereographic model (Skopek et al., 2020), to exploit hierarchical or spherical structures. Conversely, for structured data like SPD matrices, the manifold is often treated purely as a physical constraint. In recent literature, pullback Euclidean metrics, like the Log-Euclidean Metric, have gained popularity because their logarithmic maps safely project SPD matrices into a Euclidean tangent space for unconstrained deep learning operations (Arsigny et al., 2006; Huang &

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.

Van Gool, 2017; Yger et al., 2017). This prevents the violation of the SPD constraint, but effectively stops at a "Level 1" inductive bias. We argue that satisfying this domain constraint is necessary, but geometrically insufficient. Real-world SPD datasets frequently possess secondary relational biases, such as the kinematic hierarchies inherent in human action data. Forcing these tree-like distributions into the flat geometry of LEM may ignore their latent data geometry, leading to degraded model performance. Besides, the popular, non-positive-curvature-inducing AIRM may not be suitable for non-hierarchical datasets and may incur substantial computational overhead.

Recently, Bdeir et al.(Bdeir et al., 2025) embedded EEG covariance matrices into the Lorentz manifold to leverage negative curvature for improved classification. However, they define only a heuristic feedforward unidirectional projection pipeline, without any diffeomorphic pullback analyses or guarantees. Also, because they operate in unbounded Lorentz coordinates, they rely on *Tanh* distance rescaling to prevent numerical overflow, which may introduce anisotropic angular distortion. Due to the lack of a formal diffeomorphic structure, these methods often violate the SPD domain constraints, precluding the Lorentz-to-SPD inverse mappings that are vital for tasks such as Riemannian Fréchet mean, latent interpolation, and generative modeling.

To bridge these limitations, we introduce the framework of **Nested Inductive Biases**. We postulate that optimal representation learning on matrix manifolds requires simultaneously satisfying the physical constraints of the data and its latent relational structure. We achieve this via a multi-stage diffeomorphic pullback. First, SPD matrices are mapped to the Log-Euclidean tangent space at the identity, thereby satisfying the domain constraint. Second, we identify this intermediate Euclidean vector space with the origin-centered tangent space of a secondary target manifold (such as the Poincaré ball or the Projected Hypersphere). By applying the Riemannian exponential map of the target, we map these coordinates onto the target manifold. Composing these mappings allows us to pull the target metrics back onto the SPD manifold, endowing it with computationally efficient Hyperbolic and Spherical inductive biases for the first time. Because this composition is diffeomorphic, it provides valid inverse mappings for intrinsic matrix operations like the Riemannian Fréchet mean. This also enables the utilization of the native RMLR (Chen et al., 2024a) for classification without relying on tangent-space Euclidean classification. Furthermore, we introduce the Rational Conformal Metric, a novel bounded geometry designed to provide regularization against outliers instead of relying on heuristic approaches, and also act as an intermediate diffeomorphism.

While prior works construct pullback metrics primarily to endow the SPD manifold with the flat Euclidean metric(Chen et al., 2024b), our Nested Inductive Bias framework diverges in proposing a nested diffeomorphic pullback to inject non-zero constant curvature into the SPD manifold.

The main contributions of our work are summarized as follows:

- We formalize the theory of Nested Inductive Biases, defining a multi-stage diffeomorphic pullback mechanism that bridges matrix manifold constraints with non-Euclidean relational priors traditionally limited to vector embeddings.
- We derive computationally efficient Riemannian Multinomial Logistic Regression (RMLR) layers for the Nested Hyperbolic Metric (NHM) and the RCM-stabilized Nested Spherical Metric (NSM), and introduce the geodesically incomplete Rational Conformal Metric (RCM) as a dual-purpose construction: an intermediate diffeomorphism and an independent bounded feature map for outlier regularization.
- We empirically validate our framework across kinematic and signal processing datasets using Linear Probes, SPDNet, and deep RResNet architectures. We demonstrate that aligning the geometric prior with the intrinsic data structures yields superior performance, corroborated through controlled synthetic geometric ablations.

2 Preliminaries: The κ-Stereographic Model

To unify target geometries of varying constant sectional curvatures K, we adopt the generalized κ-stereographic model (Skopek et al., 2020), where $\kappa = K$ and $c = |K|$. This transitions between the Poincaré ball ($\kappa < 0$), Euclidean space ($\kappa = 0$), and the Projected Hypersphere ($\kappa > 0$). The manifold is defined as $\mathcal{M}_\kappa^n = \{z \in \mathbb{R}^n : -\kappa\|z\|^2 < 1\}$.

The geometry is defined by the conformal factor $\lambda_z^\kappa = \frac{2}{1+\kappa\|z\|^2}$, yielding the metric tensor $g_z^\kappa = (\lambda_z^\kappa)^2\mathbf{I}$. For operations mapping between the origin-centered tangent space $\mathcal{T}_\mathbf{0}\mathcal{M}_\kappa^n$ and the manifold, the exponential and logarithmic maps are defined using generalized trigonometric functions:

$$\exp_\mathbf{0}^\kappa(v) = \tan_\kappa\left(\sqrt{|\kappa|}\|v\|\right)\frac{v}{\sqrt{|\kappa|}\|v\|} \tag{1}$$

$$\log_\mathbf{0}^\kappa(z) = \tan_\kappa^{-1}\left(\sqrt{|\kappa|}\|z\|\right)\frac{z}{\sqrt{|\kappa|}\|z\|} \tag{2}$$

where $\tan_\kappa$ and $\tan_\kappa^{-1}$ correspond to $\tanh$ and arctanh for $\kappa < 0$, and $\tan$ and $\arctan$ for $\kappa > 0$.

3 The Nested Inductive Biases Framework

An open and convex cone within the ambient space of symmetric matrices, the Symmetric Positive Definite (SPD) matrix manifold, $\mathcal{S}_{++}^d$, possesses no inherent geometric structure or curvature. A geometry is only realized when the manifold is endowed with a Riemannian metric tensor. Classical formulations typically endow $\mathcal{S}_{++}^d$ with the Affine-Invariant Riemannian Metric (AIRM), which induces non-positive sectional curvature but requires computationally intensive $\mathcal{O}(d^3)$ eigendecompositions and iterative approximations for basic operations. In recent years, pullback Euclidean metrics such as the Log-Euclidean Metric (LEM) have gained popularity, offering computational efficiency but inducing flat ($K = 0$) curvature, thereby precluding the ability to model curved relational structures.

To resolve this dichotomy, we introduce the framework of Nested Inductive Biases. By establishing a multi-stage diffeomorphic composition, we execute an isometric pullback of target non-Euclidean geometries, specifically the Poincaré ball $\mathbb{P}_c^n$ and the Hypersphere $\mathbb{D}_c^n$, directly onto the SPD cone. This endows the SPD manifold with constant non-zero sectional curvature while providing the computational efficiency akin to pullback Euclidean metrics. Furthermore, unlike applying the same metric to different types of datasets, we show the importance of dataset-aligned metric use.

3.1 The Two-Stage Diffeomorphic Composition

We construct a composite mapping $\Phi : \mathcal{S}_{++}^d \to \mathcal{M}_{target}$ (where $\mathcal{M}_{target}$ represents a manifold possessing constant sectional curvature $K \neq 0$). This is formulated via $\Phi = \phi_2 \circ \phi_1$.

Proposition 1 *Let $\phi_1 : \mathcal{S}_{++}^d \to \mathbb{R}^n$ (where $n = d(d+1)/2$) be defined by mapping an SPD matrix $\mathbf{S}$ to the tangent space at the identity $\mathcal{T}_\mathbf{I}\mathcal{S}_{++}^d$ via the principal matrix logarithm, followed by the isometric half-vectorization operator $\mathrm{vech}_2(\cdot)$, which scales strictly upper-triangular entries by $\sqrt{2}$ relative to the diagonal:*

$$\mathbf{v} = \phi_1(\mathbf{S}) = \mathrm{vech}_2(\mathrm{Log}_\mathbf{I}(\mathbf{S})), \quad [\mathrm{vech}_2(M)]_{ii} = M_{ii}, \ [\mathrm{vech}_2(M)]_{ij} = \sqrt{2}\,M_{ij}\ (i<j) \tag{3}$$

The mapping ϕ_1 is a global diffeomorphism, and moreover an isometry from $(\mathcal{S}_{++}^d, g^{LE})$ to $(\mathbb{R}^n, \delta)$: $\|\mathrm{vech}_2(M)\|_2 = \|M\|_F$ for all $M \in \mathbb{S}^d$. Proofs are presented in the Appendix.

Proposition 2 *Let $\phi_2 : \mathbb{R}^n \to \mathcal{M}_{target}$ be defined by mapping the Euclidean vector $\mathbf{v}$ onto the target manifold via the Riemannian exponential map at the origin:*

$$\mathbf{z} = \phi_2(\mathbf{v}) = \exp_\mathbf{0}^{\mathcal{M}_{target}}(\mathbf{v}) \tag{4}$$

Let $c = |K|$ denote the absolute constant curvature of the target manifold. For the Poincaré ball $\mathbb{P}_c^n$, the exponential map is a global diffeomorphism. For the Projected Hypersphere $\mathbb{D}_c^n$, ϕ_2 is a diffeomorphism within the Riemannian injectivity radius of $\pi/\sqrt{c}$, equivalent to the Euclidean norm, $\|\mathbf{v}\|_2 < \pi/(2\sqrt{c})$.

Theorem 1 (Nested Diffeomorphism). *The nested mapping $\Phi(\mathbf{S}) = \phi_2(\phi_1(\mathbf{S}))$ constitutes a diffeomorphism between the SPD manifold and the target geometry $\mathcal{M}_{target}$ (globally for $K \leq 0$, and within the injectivity radius $\pi/\sqrt{c}$ for $K > 0$).*

Theorem 2 *Let g_{target} denote the Riemannian metric tensor of $\mathcal{M}_{target}$. We define the nested Riemannian metric g_{nested} on $\mathcal{S}_{++}^d$ as the pullback of g_{target} by Φ:*

$$g_{nested} = \Phi^* g_{target} \tag{5}$$

For any $\mathbf{S} \in \mathcal{S}_{++}^d$ and tangent vectors $\mathbf{V}, \mathbf{W} \in \mathcal{T}_{\mathbf{S}}\mathcal{S}_{++}^d$, the inner product evaluates to:

$$(g_{nested})_{\mathbf{S}}(\mathbf{V}, \mathbf{W}) = (g_{target})_{\Phi(\mathbf{S})} \left(\Phi_{*,\mathbf{S}}(\mathbf{V}), \Phi_{*,\mathbf{S}}(\mathbf{W})\right) \tag{6}$$

where $\Phi_{*,\mathbf{S}}$ is the pushforward differential of Φ. Under this metric, Φ acts as an isometry between $(\mathcal{S}_{++}^d, g_{nested})$ and $(\mathcal{M}_{target}, g_{target})$, ensuring that the SPD manifold strictly inherits the constant sectional curvature of the target space.

Proposition 3 (Congruence Equivariance, Induced Invariance, and Limit Consistency). *The composite map Φ is equivariant under orthogonal congruence ($\mathbf{S} \mapsto \mathbf{U}\mathbf{S}\mathbf{U}^\top$ for $\mathbf{U} \in O(d)$) and matrix inversion. Consequently, the nested metric g_{nested} is invariant under pullback along both actions. Furthermore, as the target curvature $\kappa \to 0$, $g_{nested} \to 4g_{LE}$. Thus, up to a constant scaling factor, the Log-Euclidean Metric is recovered as the flat boundary case of our nested framework.*

3.2 Nested Riemannian Multinomial Logistic Regression

Because the SPD manifold equipped with g_{nested} now possesses constant non-zero curvature, linear classification cannot be performed via standard Euclidean hyperplanes. We formulate the classification boundary directly within the target geometry using generalized Riemannian Multinomial Logistic Regression (RMLR) (Chen et al., 2024a).

Definition 1 (Nested Riemannian Hyperplane). *For a target class k, let $\mathbf{P}_k \in \mathcal{S}_{++}^d$ denote the learned class anchor, mapped to the target manifold as $p_k = \Phi(\mathbf{P}_k) \in \mathcal{M}_{target}$. Let $\tilde{a}_k \in \mathcal{T}_{\mathbf{0}}\mathcal{M}_{target} \setminus \{\mathbf{0}\}$ be optimized as a Euclidean parameter at the origin, and parallel transported to the anchor as $a_k = \mathrm{PT}_{\mathbf{0} \to p_k}(\tilde{a}_k)$. Using the Riemannian Logarithmic map, the decision boundary is defined by:*

$$\tilde{H}_{a_k, p_k} = \left\{\mathbf{S} \in \mathcal{S}_{++}^d : \langle \mathrm{Log}_{p_k}(\Phi(\mathbf{S})), a_k \rangle_{p_k} = 0\right\} \tag{7}$$

Theorem 3 (Nested RMLR Classification). *Following the generalized RMLR framework, the multinomial probability for class k evaluates directly via the Riemannian inner product within the tangent space of the class centroid:*

$$p(y = k|\mathbf{S}) \propto \exp\left(\langle \mathrm{Log}_{p_k}(\Phi(\mathbf{S})), a_k \rangle_{p_k}\right) \tag{8}$$

where $a_k \in \mathcal{T}_{\mathbf{0}}\mathcal{M}_{target} \setminus \{\mathbf{0}\}$ is optimized as a Euclidean parameter within the origin tangent space $\mathcal{T}_{\mathbf{0}}\mathcal{M}_{target}$ and parallel transported to p_k.

Remark 1. *While our framework establishes a global diffeomorphism to the punctured hypersphere, we identify a theoretical vulnerability of the established Generalized RMLR (Chen et al., 2024a) for spherical classification in its Log_{p_k} map when its input approaches the injectivity radius. In Appendix B, we formally diagnose this vulnerability, offer a topological boundary solution ($\alpha = 4\sqrt{c}/\pi$), and empirically demonstrate why cross-entropy optimization naturally averts this in practice (0.00% violation rate).*

4 The Rational Conformal Metric (RCM)

We show that mapping unbounded Euclidean data onto manifolds with finite tangent-space injectivity radii introduces two forms of metric misalignment: **Curvature Mismatch** and **Norm Mismatch**.

The Projected Hypersphere $\mathbb{D}_c^n$ corresponds to the stereographic chart of the sphere with one antipodal point removed; its exponential map at the origin loses injectivity beyond the radius $\pi/(2\sqrt{c})$. If a framework maps vectors that exceed this norm limit, it not only leads to algorithmic instability; it also represents a failure of the global bijection on that dataset. We quantified this unconstrained Norm Mismatch: on deep architectures (such as RResNet), $> 90\%$ of the vectors to exceed the Euclidean injectivity domain ($\|\mathbf{v}\|_2 \geq \pi/(2\sqrt{c})$). Consequently, the standard spherical $\exp_{\mathbf{0}}$ map led to gradient explosions and diffeomorphism failure. Although this limitation is noted in the literature, prior work (Skopek et al., 2020) sidestepped solving this failure directly. We resolve it formally, ensuring that the nested spherical projection remains universally diffeomorphic for any SPD manifold regardless of the dataset's variance.

Also, while Hyperbolic space is geodesically complete and Spherical space's exponential map is defined for tangent vectors of arbitrary Euclidean magnitude, neither underlying construction bounds

the magnitude of the tangent vectors it accepts as input. In real-world sensor domains (such as Radar signal processing), amplitude anomalies may manifest as outliers within the base Log-Euclidean tangent space.

To resolve the Norm Mismatch without relying on anisotropic heuristic approaches, together with bounding outliers, we propose a novel metric: the Rational Conformal Metric (RCM). The metric is designed to provide C^∞ smoothness and provide asymptotic gradient stability, averting vanishing gradients of purely quadratic metrics as the norm becomes large.

Definition 2 (Rational Conformal Metric). *We define the Rational Conformal Metric on the Euclidean vector space $\mathbb{R}^n$ via a radially symmetric conformal transformation. Let the ϵ-smoothed radius $\tilde{r}_\epsilon(\mathbf{x}) = \sqrt{\|\mathbf{x}\|^2 + \epsilon^2} - \epsilon$, where $0 < \epsilon < \frac{1}{2\alpha}$. For a scale parameter $\alpha > 0$, the conformal factor $\lambda(\mathbf{x})$ that scales infinitesimal lengths is defined as:*

$$\lambda(\mathbf{x}) = \frac{1 - \alpha\epsilon + \frac{\alpha\epsilon^2}{\sqrt{\|\mathbf{x}\|^2+\epsilon^2}}}{(1 + \alpha\tilde{r}_\epsilon(\mathbf{x}))^2} \tag{9}$$

The Riemannian metric tensor is defined as $g^{RCM}_{\mathbf{x}} = \lambda(\mathbf{x})^2\mathbf{I}$, yielding:

$$g^{RCM}_{\mathbf{x}} = \left(\frac{1 - \alpha\epsilon + \frac{\alpha\epsilon^2}{\sqrt{\|\mathbf{x}\|^2+\epsilon^2}}}{(1 + \alpha(\sqrt{\|\mathbf{x}\|^2 + \epsilon^2} - \epsilon))^2} \right)^2 \mathbf{I} \tag{10}$$

where $\mathbf{I}$ is the Euclidean identity matrix. This establishes a C^∞ smooth metric tensor everywhere on $\mathbb{R}^n$. Moreover, since $\epsilon < 1/(2\alpha)$ implies $\alpha\epsilon < 1/2$, the numerator satisfies $1 - \alpha\epsilon + \alpha\epsilon^2/\sqrt{\|\mathbf{x}\|^2 + \epsilon^2} > 1 - \alpha\epsilon > 1/2 > 0$ for all $\mathbf{x}$, so $\lambda(\mathbf{x}) > 0$ everywhere and g^{RCM} is a positive-definite Riemannian metric.

Unlike standard spaces of constant curvature, the Rational Conformal Metric yields a geodesically incomplete manifold. A geodesic ray extending to infinite Euclidean distance possesses a finite Riemannian length. We leverage this deliberate incompleteness to formulate a bounded Tangent Space Mapping (TSM) framework for outlier regularization, as well as setting up an intermediate diffeomorphism.

Theorem 4. *Under the Rational Conformal Metric, $\mathrm{Log}_{\mathbf{0}}$ map acts as a C^∞ global diffeomorphism from $\mathbb{R}^n$ onto the open Euclidean ball $\mathcal{B}_{1/\alpha}$. Hence, the TSM ensures a unique and bounded representation for any $\mathbf{x} \in \mathbb{R}^n$ and restricts the domain of the inverse exponential map $\mathrm{Exp}_{\mathbf{0}}$ to the open ball.*

Remark 2 (Computational and Numerical Efficiency). *Beyond its topological utility for outlier compression, the Rational Conformal Metric offers a practical computational advantage for deep learning. Standard constant-curvature manifolds (such as the κ-stereographic models) rely heavily on transcendental functions (*$\tan$, $\arctan$, $\tanh$*) and square roots to compute their exponential and logarithmic maps. In contrast, the RCM's $\mathrm{Exp}_{\mathbf{0}}$ and $\mathrm{Log}_{\mathbf{0}}$ mappings evaluate via basic rational polynomials, eliminating the numerical instabilities commonly associated with trigonometric back-propagation in automatic differentiation frameworks. While methods such as element-wise* $\tanh$ *could theoretically constrain the tangent space, they compress vectors toward a hypercube, inducing anisotropic angular distortion. RCM acts as a conformal scaling without relying on unstable transcendental functions.*

4.1 Stabilizing Positive Curvature via RCM Bounding

Instability and divergence in optimizing models on high-dimensional spherical spaces have been a documented challenge in the literature. (Skopek et al., 2020) observed that spherical space model optimization remains significantly more difficult than hyperbolic or Euclidean spaces. They cited a lack of a more thorough investigation and "side-stepped" the problem in their paper. Further, they conjectured vectors approaching the antipodal South Pole, besides floating-point precision issues, as the problem's cause.

Although correct in identifying the issue, our analyses and experiments demonstrate that the injectivity radius violation may play a more important role in causing it. In frameworks where unbounded

tangent space vectors are directly mapped to the spherical manifold via the exp0 map, vectors with large norms may approach or overstep the injectivity radius. When the Euclidean norm $\|\mathbf{v}\|_2$ approaches the injectivity boundary of $\pi/(2\sqrt{c})$, the exp0 mapping approaches the tangent-function asymptote, leading to gradient explosion. If the radius is overstepped, the mapping wraps around, breaking the global bijection and the inverse logarithmic mappings ($\log_{\mathbf{0}}$). This can explain the instabilities and divergence in models like Mixed-Curvature VAEs (Skopek et al., 2020), reliant on the inverse mappings for operations like evaluating the Wrapped Normal probability density function. Furthermore, reliance on heuristic gradient clipping or coordinate clamping induces zero-gradient boundaries and diffeomorphic violations.

We resolve this by utilizing a specific instantiation of the Rational Conformal Metric to unify our framework. We employ the RCM logarithmic map as an intermediate diffeomorphism prior to projecting onto the spherical manifold, creating a 3-Level inductive bias by aligning the tangent spaces of the RCM and the NSM, where RCM supplies a canonical, bounded, everywhere-differentiable coordinate chart whose image coincides with the target manifold's injectivity domain.

Theorem 5 (3-Stage Diffeomorphism) *Let the RCM scale parameter be fixed to $\alpha = 2\sqrt{c}/\pi$. The mapping $\Phi_3 = \exp_{\mathbf{0}}^{\mathbb{D}_c^n} \circ Log_{\mathbf{0}}^{RCM} \circ \phi_1$ establishes a global diffeomorphism from the SPD manifold $\mathcal{S}_{++}^d$ onto the (punctured) Projected Hypersphere $\mathbb{D}_c^n$.*

One might hypothesize that simply decreasing the curvature parameter c (or treating it as a learnable parameter) could expand the injectivity domain enough to avoid wrap-around. However, this fails due to two reasons. Topologically, because the Log-Euclidean space $\mathbb{R}^n$ is unbounded, for any finite curvature $c > 0$, there will always exist SPD matrices whose representations exceed the finite injectivity radius $\pi/(2\sqrt{c})$, breaking the global bijection. Empirically, cross-entropy optimization can push deep representations outward until this expanded boundary is breached regardless. By uniquely decoupling the spatial bound from the manifold curvature, the RCM mapping establishes a global diffeomorphism that preserves both the target geometry and gradient stability.

5 Experimental Methodology

Our framework enables the imposition of custom geometric priors by composing the Log-Euclidean mapping with κ-stereographic exponential maps. We denote these composite Riemannian metrics on the SPD manifold as **NHM** (Nested Hyperbolic Metric, $K < 0$) and **NSM** (Nested Spherical Metric, $K > 0$). This allows the subsequent Riemannian Multinomial Logistic Regression (RMLR) layer (Chen et al., 2024a) to compute decision boundaries that simultaneously respect the symmetric positive definiteness of the data and the intrinsic relational curvature of the dataset.

Datasets and Architectures: We evaluate our framework on three standard benchmarks representing varied intrinsic geometries: First-Person Hand Action (FPHA) (Garcia-Hernando et al., 2018), HDM05 (Müller et al., 2007), and Radar Target Classification (Brooks et al., 2019). To isolate metric inductive bias from network capacity, we utilize three architectures: a shallow Linear Probe, SPDNet (Huang & Van Gool, 2017), and a deep Riemannian ResNet (RResNet) (Katsman et al., 2023). We benchmark NHM and NSM-RCM against the Log-Euclidean Metric (LEM), unconstrained NSM, and the Affine-Invariant Riemannian Metric (AIRM). We measure the intrinsic geometry of our datasets via the discrete Ollivier-Ricci Curvature (ORC). Complete details regarding dataset preprocessing, network configurations, hyperparameter optimization, and necessary numerical safeguards for rank-deficient matrices are provided in Appendix C.

Synthetic Geometric Control: To validate the classification bounds of NSM and NHM, we synthesize a covariance dataset localized to a compact, positively curved submanifold within $\mathcal{S}^{20}$. By modulating a radial embedding parameter, we shift the intrinsic dataset geometry between macro-scale positive curvature and noise-dominated negative curvature. The generation procedure is detailed in Appendix C.

Adversarial Outlier Regularization: To empirically validate the Rational Conformal Metric's utility as a bounded Euclidean feature map, we generate linearly separable core classes in a 2D Euclidean feature space and inject strategic spatial anomalies (such as cross-boundary outliers, heavy-tailed

noise). We ablate the RCM against standard L_2 normalization and $\tanh$ squashing. The full anomaly generation protocols are defined in Appendix C.

6 RESULTS

A fundamental premise of our work is that metric curvature must align with data geometry. The ORC analyses validated our empirical neural network results: HDM05 exhibited the strongest hyperbolic geometry (ORC = -0.1021), followed closely by FPHA (ORC = -0.0886), while Radar exhibited much softer negative curvature (ORC = -0.0439). This aligns well with the modalities of the datasets. FPHA and HDM05 represent 3D skeletal tracking data, in which human kinematic articulation forms a branching physical hierarchy, whereas the Radar dataset comprises statistical signal covariances, demonstrating a less tree-like structure.

Consequently, although the NHM dominates NSM in the majority of cases (all 3 datasets show negative curvature), the average relative performance drop of the RCM-bounded NSM compared to the NHM was greater on HDM05 and FPHA than on Radar. Importantly, across our benchmark datasets, we observe a **strong linear correlation** ($r \approx 0.993$) between the magnitude of the dataset's negative intrinsic curvature and the dataset's average performance gain of NHM relative to NSM (RCM).

We also observe that while on the nearly flat Radar dataset, the flat LEM performs relatively well, its performance falls behind both NSM and RCM-rescued NSM on the highly hierarchical HDM05 dataset. This may be attributed to the geometric distortion of using flat metrics or single tangent space Euclidean classification documented in the literature (Jaquier et al., 2024) and highlights the need to use fast, curved pullback metrics on datasets with high curvature. This also highlights the effectiveness of the RCM intervention, which "saves" the competitiveness of the native NSM, falling behind only the NHM on the hierarchical dataset.

6.1 MAIN BENCHMARK RESULTS

Table 1: Native Riemannian Classification (RMLR) on the FPHA Dataset.

Architecture	LEM	AIRM	NHM	NSM (Unconstr.)	NSM (RCM)
Linear Probe	0.8539	0.8417	**0.8557**	0.1443	0.8400
SPDNet	0.8191	0.8243	**0.8504**	0.1374	0.8383
RResNet	0.7652	0.7930	**0.8643**	0.0435	0.8243

Table 2: Native Riemannian Classification (RMLR) on the Radar Dataset.

Arch.	LEM	AIRM	NHM	NSM (Unconstr.)	NSM (RCM)
Linear	0.9600 ± 0.006	0.9513 ± 0.005	0.9633 ± 0.009	0.8163 ± 0.019	$\mathbf{0.9643 \pm 0.008}$
SPDNet	0.9613 ± 0.008	0.9520 ± 0.008	$\mathbf{0.9633 \pm 0.009}$	0.7993 ± 0.029	0.9600 ± 0.006
RResNet	0.9590 ± 0.012	0.9507 ± 0.006	$\mathbf{0.9607 \pm 0.011}$	0.5893 ± 0.121	0.9497 ± 0.007

As detailed in Tables 1 and 2, the native RMLR results corroborate our intrinsic dataset curvature findings. The nested Hyperbolic metric (NHM) shows superior performance on the intrinsically negatively curved datasets. Most notably, in the deep FPHA RResNet architecture, NHM achieves a prominent +9.9% accuracy gain over the baseline Log-Euclidean RMLR and a +7.1% gain over the standard AIRM benchmark. Conversely, the Spherical metric (NSM) exhibits substantial margin collapse, demonstrating the need for metric alignment to preserve separability.

Notably, a striking result is the substantial failure of the unconstrained original spherical metric. This demonstrates the necessary requirement of metric and norm alignment to preserve separability in the NSM, a problem formerly "side-stepped" in literature. Moreover, the RCM intervention achieves significant success in "saving" the NSM, restoring it to a competitive state. Importantly, even after the RCM intervention, the NSM-RCM continues to underperform the NHM in a majority of the cases. Because the Norm Mismatch is eliminated, this remaining performance gap is attributable to

Curvature Mismatch, supporting our hypothesis that enforcing positive curvature on intrinsically negatively curved or hierarchical data degrades linear separability.

To further validate the framework, we evaluated the highly multi-class (117 classes), high-dimensional (93×93) HDM05 dataset. As shown in Table 3, the Hyperbolic prior continues to dominate flat and spherical metrics.

Table 3: Classification Accuracy on the HDM05 Dataset. Bold values indicate the highest accuracy.

Arch.	LEM	Tangent AIRM	NHM	NSM (Unconstr.)	NSM (RCM)
Linear	0.4231 ± 0.027	0.3084 ± 0.006	$\mathbf{0.6153 \pm 0.016}$	0.1273 ± 0.016	0.5855 ± 0.014
SPDNet	0.5927 ± 0.019	0.3031 ± 0.013	$\mathbf{0.6729 \pm 0.007}$	0.1273 ± 0.017	0.6186 ± 0.017
RResNet	0.5514 ± 0.016	0.4002 ± 0.037	$\mathbf{0.6983 \pm 0.012}$	0.0456 ± 0.017	0.6864 ± 0.025

6.2 Computational Intractability of Native AIRM

While the AIRM is a classical metric that imparts non-positive curvature, its iterative time complexity makes it computationally prohibitive for highly multi-class datasets. Benchmarking on HDM05 (Table 4) revealed that Native AIRM RMLR requires over 750 seconds per epoch to compute pairwise matrix logarithms. In stark contrast, our Nested Inductive Bias framework evaluates the identical network in roughly 38 seconds per epoch. This shows that our framework induces non-zero curvature to the SPD manifold while maintaining the speed of flat pullback Euclidean metrics, achieving a 95% reduction in computational overhead (nearly a $20\times$ acceleration) compared to the AIRM.

Table 4: Computational Intractability Benchmark (HDM05, NVIDIA T4 GPU, Batch = 32).

Metric	Latency (ms / batch)	Time (sec / epoch)
Log-Euclidean (Flat)	910.40	38.24
Tangent Space AIRM	911.09	38.27
NHM (Hyperbolic)	911.63	38.29
NSM (RCM)	912.02	38.30
Native AIRM RMLR	**17873.77**	**750.70**

6.3 Validation via Synthetic Geometric Control Results

To ensure the performance degradation of the NSM on real-world data was a consequence of intrinsic geometric mismatch rather than algorithmic instability, we evaluated the RResNet architecture on our controlled synthetic dataset (Section C.2). By modulating the radial parameter r, we shift the data distribution between positive and negative intrinsic curvatures.

Table 5: Classification Test Set Accuracy on Synthetic Geometric Control Dataset.

Radial Scale	Intrinsic Curvature (ORC)	NHM	NSM
Macroscopic ($r = 1.5$)	$+0.010$	0.9950	$\mathbf{1.0000}$
Microscopic ($r = 0.25$)	-0.213	$\mathbf{0.8000}$	0.6875

As detailed in Table 5, the empirical results validate the volume-growth mechanics established in our experimental setup. At the macroscopic scale ($r = 1.5$, positive ORC), NSM natively matches the spherical geometry to achieve perfect accuracy (1.0000), while NHM's radial distance distortion slightly degrades separability (0.9950).

However, at the microscopic scale ($r = 0.25$, negative ORC), the sub-Euclidean volume growth of NSM fails to match the negatively curved noise-dominated core, degrading performance (0.6875). NHM aligns with the negative curvature; its super-Euclidean expansion disperses the overlapping representations, pulling the classes apart and restoring accuracy to 0.8000. This symmetrical ablation demonstrates that imposing Riemannian priors is only effective when explicitly aligned with the dataset's intrinsic geometry.

6.4 RCM AS A REGULARIZER: OUTLIER RESULTS

Table 6: Classification Accuracy under Adversarial Spatial Outliers. The unbounded Log-Euclidean baseline sacrifices core separability to accommodate extreme anomalies, while the proposed Rational Conformal Metric (RCM) mathematically bounds their gradient contribution, restoring perfect core accuracy.

	Baseline (Log-Euclidean)		**Proposed (RCM)**	
Adversarial Experiment	**Total Acc.**	**Core Acc.**	**Total Acc.**	**Core Acc.**
Cross-Boundary Outlier Injection	22.5%	17.0%	83.3%	$\mathbf{100.0}\%$
Directional Heavy-Tailed Noise	75.0%	78.0%	$\mathbf{92.3}\%$	$\mathbf{100.0}\%$
Symmetric Extreme Outliers	75.9%	73.5%	90.9%	$\mathbf{100.0}\%$

As illustrated in Figure 2 (Appendix D.2) and quantitatively detailed in Table 6, the unbounded Log-Euclidean baseline is highly susceptible to unconstrained spatial anomalies. Because outliers incur large cross-entropy penalties, the linear optimizer systematically shifts the decision hyperplanes to compensate. This global hyperplane shift degrades the separability of the core data distributions, dropping core accuracy to 22.5% under cross-boundary outlier injection. However, instead of just using the RCM as an intermediate diffeomorphism provider, we can use it as a regularizer. Mapping the identical feature spaces through the Rational conformal metric transformation asymptotically confines all representations within an open Euclidean ball of radius $1/\alpha$. By bounding the spatial magnitude of the outliers, their gradient contribution during backpropagation is limited. The linear classifier is subsequently able to optimize the hyperplanes relative to the dense core distributions, stabilizing the decision boundaries and restoring core classification accuracy to near-perfect levels across all adversarial variants.

Ablation: To isolate the structural advantages of this specific radial compression, we ablated the Rational mapping against standard Euclidean normalization techniques (L_2 normalization and $\tanh$ squashing). As detailed in Appendix D, competing methods either collapse intrinsic data magnitude into rigid 1D arcs or introduce anisotropic box distortion. The proposed RCM provides the favorable geometric compromise, enforcing mathematical bounds to limit outlier impact while perfectly preserving the relative spatial variance of the core distributions, positioning it as a robust geometric prior for mitigating unbounded outliers in deep learning architectures

7 CONCLUSION

In this work, we introduced a nested inductive bias framework for Symmetric Positive Definite (SPD) manifolds, bridging the theoretical gap between native matrix geometries and established vector-space geometric priors. By utilizing nested diffeomorphisms, we pulled back Hyperbolic and Spherical geometries to construct curvature-aligned Riemannian classifiers. Our empirical evaluations and synthetic ablations demonstrate that deep manifold networks experience degradation in class separability unless the metric curvature and norm align with the data distribution. Furthermore, for standard vectorized architectures, our proposed Rational Conformal Metric (RCM) imposes a radial bound on the representation space, enabling a 3-stage diffeomorphism and providing a strong robustness against extreme spatial anomalies. A direction for future research is to extend this nested inductive bias pullback framework to other manifolds beyond the SPD manifold.

AI USE STATEMENT

In this work, generative AI tools were used for editing the research paper to improve readability, grammar, and language presentation. We have not used generative AI tools for developing core theoretical models, formulating mathematical claims, autonomously generating synthetic datasets, or implementing end-to-end software code. All AI-assisted work, structural formatting, and text refinement were thoroughly reviewed and verified by the authors for correctness. We take full responsibility for the entire content, original theoretical claims, and integrity of this submission, including AI-assisted work.

Ethics statement

This research relies only on standard, publicly available benchmark datasets that do not contain personally identifiable or sensitive information. To the best of our knowledge, we do not foresee any negative societal impacts or ethical concerns arising from the methodological advancements presented in this work.

Reproducibility statement

All theoretical claims are established under explicit geometric assumptions, with complete mathematical proofs provided in Appendix. Comprehensive experimental details, including dataset preprocessing, network architectures, and necessary numerical safeguards, are outlined in Section 5 and Appendix C. The source code will be released upon acceptance.

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.

Ahmad Bdeir, Johannes Burchert, Lars Schmidt-Thieme, and Niels Landwehr. Robust hyperbolic learning with curvature-aware optimization. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=lJ5WCJZfQn.

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, 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, 2024a.

Ziheng Chen, Yue Song, Tianyang Xu, Zhiwu Huang, Xiao-Jun Wu, and Nicu Sebe. Adaptive log-euclidean metrics for spd matrix learning. *IEEE Transactions on Image Processing*, 33:5194–5205, 2024b.

Tushar Das, Subrata Dutta, Sarmistha Neogy, and Koushlendra Kumar Singh. Beyond rigid geometries: The spline-pullback metric for universal diffeomorphic spd representation learning. *arXiv preprint arXiv:2605.04406*, 2026.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in neural information processing systems*, 31, 2018.

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.

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.

Noémie Jaquier, Leonel Rozo, and Tamim Asfour. Unraveling the single tangent space fallacy: An analysis and clarification for applying riemannian geometry in robot learning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 242–249. IEEE, 2024.

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.

Reinmar Kobler, Jun-ichiro Hirayama, Qibin Zhao, and Motoaki Kawanabe. Spd domain-specific batch normalization to crack interpretable unsupervised domain adaptation in eeg. *Advances in Neural Information Processing Systems*, 35:6219–6235, 2022.

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.

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.

Ondrej Skopek, Octavian-Eugen Ganea, and Gary Bécigneul. Mixed-curvature variational autoencoders. In *International Conference on Learning Representations*, 2020. URL `https://openreview.net/forum?id=S1g6xeSKDS`.

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

Florian Yger, Maxime Berar, and Fabien Lotte. Riemannian approaches in brain-computer interfaces: A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1753–1762, 2017. doi: 10.1109/TNSRE.2016.2627016.

A Appendix: Proofs of theoretical claims

A.1 Proof of Proposition 1

Proof. $\mathrm{Log}_{\mathbf{I}} : \mathcal{S}^d_{++} \to \mathbb{S}^d$ is a global C^∞ diffeomorphism. vech_2 is linear and bijective, a rescaled coordinate map with respect to the standard orthonormal basis of $(\mathbb{S}^d, \langle\cdot,\cdot\rangle_F)$. hence a diffeomorphism $\mathbb{S}^d \to \mathbb{R}^n$, with C^∞ smoothness with smooth inverses, being a finite-dimensional linear isomorphism. The isometry follows since $\|M\|_F^2 = \sum_i M_{ii}^2 + 2\sum_{i<j} M_{ij}^2 = \sum_i M_{ii}^2 + \sum_{i<j}(\sqrt{2}M_{ij})^2 = \|\mathrm{vech}_2(M)\|_2^2$. Composition of diffeomorphisms yields a global diffeomorphism $\phi_1 : \mathcal{S}^d_{++} \to \mathbb{R}^n$. ■

A.2 Proof of Proposition 2

Proof. By the Cartan-Hadamard theorem, any simply connected, complete Riemannian manifold with non-positive sectional curvature (such as the Poincaré ball $\mathbb{P}^n_c$) contains no conjugate points. Consequently, its Riemannian exponential map $\exp_{\mathbf{0}}$ is a global diffeomorphism from the tangent space to the manifold. For the Projected Hypersphere $\mathbb{D}^n_c$ ($K > 0$), conjugate points (antipodes) exist at a Riemannian distance of $\pi/\sqrt{c}$. Under the generalized κ-stereographic model, the conformal factor at the origin is $\lambda^\kappa_{\mathbf{0}} = 2$. Therefore, the Riemannian norm of a tangent vector relates to its ambient Euclidean norm via $\|\mathbf{v}\|_{\mathcal{M}} = \lambda^\kappa_{\mathbf{0}}\|\mathbf{v}\|_2 = 2\|\mathbf{v}\|_2$. For the exponential map to avoid the cut locus and remain a diffeomorphism, the Riemannian norm must satisfy $\|\mathbf{v}\|_{\mathcal{M}} < \pi/\sqrt{c}$. Substituting the Euclidean conversion yields the required ambient bound for the tangent space vectors: $\|\mathbf{v}\|_2 < \pi/(2\sqrt{c})$. Furthermore, restricted to this domain, $\exp_{\mathbf{0}}$ is surjective onto $\mathbb{D}^n_c = \mathbb{R}^n$, because the function $\tan(\sqrt{c}t)/\sqrt{c}$ sweeps $[0,\infty)$ as t sweeps $[0, \pi/(2\sqrt{c}))$. ■

A.3 Proof of Theorem 1

Proof. By the Inverse Function Theorem, the composition of two C^∞ diffeomorphisms yields a C^∞ diffeomorphism. Since ϕ_1 is globally diffeomorphic and ϕ_2 is diffeomorphic within its respective bounds, their composition $\Phi = \phi_2 \circ \phi_1$ provides a smooth bijection with a smooth inverse

$\Phi^{-1} = \phi_1^{-1} \circ \phi_2^{-1}$. For the Poincaré ball this holds globally on all of $\mathcal{S}^d_{++}$; for the Projected Hypersphere it holds on the open subset $\Phi^{-1}(\mathcal{B}_{\pi/(2\sqrt{c})}) \subset \mathcal{S}^d_{++}$ corresponding to the injectivity radius. ■

A.4 Proof of Theorem 2

Proof. By definition of the pullback operation, equipping a manifold with a pullback metric via a diffeomorphism preserves the first fundamental form, rendering the mapping an isometry. By the Fundamental Theorem of Riemannian Geometry, the metric tensor uniquely determines the Levi-Civita connection and the associated Riemann curvature tensor. Because local isometries preserve the metric tensor, the sectional curvature of the SPD manifold under g_{nested} is equal to the target manifold's curvature at all points. ■

A.5 Proof of Proposition 3

Congruence and Inversion. Under orthogonal congruence $\mathbf{S} \mapsto \mathbf{U}\mathbf{S}\mathbf{U}^\top$, the matrix logarithm satisfies $\log(\mathbf{U}\mathbf{S}\mathbf{U}^\top) = \mathbf{U}\log(\mathbf{S})\mathbf{U}^\top$. Thus, the mapping ϕ_1 is *equivariant* under orthogonal action, not invariant. However, because the target metrics in the κ-stereographic model are radially symmetric (invariant under origin-centered rotations), this induced orthogonal action preserves the inner product, rendering the pulled-back metric g_{nested} invariant under orthogonal congruence. Furthermore, ϕ_1 maps matrix inversion to vector negation ($\mathbf{S}^{-1} \mapsto -\mathbf{v}$). Because radially symmetric target metrics are invariant to negation, inversion invariance is also preserved.

Limit Consistency. As $\kappa \to 0$, the conformal factor $\lambda_{\mathbf{0}}^\kappa = 2/(1 + \kappa\|\mathbf{z}\|^2)$ approaches 2 pointwise. Consequently, the target metric $g_z^\kappa = (\lambda_z^\kappa)^2\mathbf{I} \to 4\mathbf{I}$. Therefore, the pulled-back metric evaluates to $g_{nested} \to 4g_{LE}$. Up to a fixed global scale factor of 4 (inherited from standard κ-stereographic conformal normalization), the Log-Euclidean Metric is recovered as the flat boundary case of the nested framework. ■

A.6 Proof of Theorem 3

Proof. Let $\mathbf{z} = \Phi(\mathbf{S}) \in \mathcal{M}_{target}$. The inner product evaluates to $\langle \mathrm{Log}_{p_k}(z), a_k\rangle_{p_k} = \|\mathrm{Log}_{p_k}(z)\|_{p_k}\|a_k\|_{p_k}\cos(\theta)$. As established by Chen et al. (2024a), in manifolds with constant curvature, this inner product directly encodes a monotonically increasing function of the signed orthogonal distance to the decision boundary $\tilde{H}_{a_k,p_k}$. Therefore, it serves as a valid logit for cross-entropy optimization. ■

A.7 Proof of Theorem 4

Proof: Under the RCM $g_{\mathbf{x}}^{RCM} = \lambda(\|\mathbf{x}\|)^2\mathbf{I}$, the geodesic distance from the origin along a radial path $\gamma(t) = t\frac{\mathbf{x}}{\|\mathbf{x}\|}$ is given by the integral of the conformal factor. Let $r = \|\mathbf{x}\|$. The distance is:

$$d_{RCM}(\mathbf{0}, \mathbf{x}) = \int_0^{\|\mathbf{x}\|} \lambda(r)\, dr \tag{11}$$

To evaluate this, let:

$$u(r) = 1 + \alpha(\sqrt{r^2 + \epsilon^2} - \epsilon) = 1 - \alpha\epsilon + \alpha\sqrt{r^2 + \epsilon^2} \tag{12}$$

The derivative of $u(r)$ with respect to r is:

$$u'(r) = \frac{\alpha r}{\sqrt{r^2 + \epsilon^2}} \tag{13}$$

$$u(r) - ru'(r) = (1 - \alpha\epsilon + \alpha\sqrt{r^2 + \epsilon^2}) - r\left(\frac{\alpha r}{\sqrt{r^2 + \epsilon^2}}\right) \tag{14}$$

$$= 1 - \alpha\epsilon + \frac{\alpha(r^2 + \epsilon^2) - \alpha r^2}{\sqrt{r^2 + \epsilon^2}} \tag{15}$$

$$= 1 - \alpha\epsilon + \frac{\alpha\epsilon^2}{\sqrt{r^2 + \epsilon^2}} \tag{16}$$

The conformal factor can be rewritten as:

$$\lambda(r) = \frac{u(r) - ru'(r)}{u(r)^2} \tag{17}$$

Substituting the $\lambda(r)$ into the distance integral:

$$\int \frac{u(r) - ru'(r)}{u(r)^2}\, dr = \int \frac{1}{u(r)}\, dr - \int ru(r)^{-2}u'(r)\, dr \tag{18}$$

Applying integration by parts to the second term:

$$\int ru(r)^{-2}u'(r)\, dr = r\left(-\frac{1}{u(r)}\right) - \int \left(-\frac{1}{u(r)}\right)\, dr \tag{19}$$

$$= -\frac{r}{u(r)} + \int \frac{1}{u(r)}\, dr \tag{20}$$

Substituting back:

$$\int \lambda(r)\, dr = \int \frac{1}{u(r)}\, dr - \left(-\frac{r}{u(r)} + \int \frac{1}{u(r)}\, dr\right) = \frac{r}{u(r)} \tag{21}$$

Evaluating the definite integral from 0 to $\|\mathbf{x}\|$:

$$d_{RCM}(\mathbf{0}, \mathbf{x}) = \left[\frac{r}{1 + \alpha(\sqrt{r^2 + \epsilon^2} - \epsilon)}\right]_0^{\|\mathbf{x}\|} = \frac{\|\mathbf{x}\|}{1 + \alpha(\sqrt{\|\mathbf{x}\|^2 + \epsilon^2} - \epsilon)} \tag{22}$$

Finally:

$$\mathbf{v} = \mathrm{Log}_{\mathbf{0}}(\mathbf{x}) = d_{RCM}(\mathbf{0}, \mathbf{x})\frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{\mathbf{x}}{1 + \alpha(\sqrt{\|\mathbf{x}\|^2 + \epsilon^2} - \epsilon)} \tag{23}$$

$$\|\mathbf{v}\| = \frac{\|\mathbf{x}\|}{1 + \alpha(\sqrt{\|\mathbf{x}\|^2 + \epsilon^2} - \epsilon)} \tag{24}$$

To determine the asymptotic bound, we evaluate the limit as $\|\mathbf{x}\| \to \infty$:

$$\lim_{\|\mathbf{x}\| \to \infty} \|\mathbf{v}\| = \frac{1}{\alpha} \tag{25}$$

As the radial function is monotonically increasing, since $\frac{d}{dr}(r/u(r)) = \lambda(r) > 0$ by the positive-definiteness shown above, we have the inequality $\|\mathbf{v}\| < \frac{1}{\alpha}$ for all finite $\mathbf{x} \in \mathbb{R}^n$. Therefore, $\mathrm{Log}_{\mathbf{0}}$ injectively maps the unbounded space $\mathbb{R}^n$ into the open Euclidean ball $\mathcal{B}_{1/\alpha}$.

The squared Euclidean norm $\|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2$ is a smooth polynomial in $\mathbb{R}^n$. The radicand satisfies $\|\mathbf{x}\|^2 + \epsilon^2 \geq \epsilon^2 > 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and the square root operation is infinitely differentiable on this domain. The composition, products, and quotients of positive C^∞ smooth functions yield a C^∞ smooth mapping globally. Hence, $\mathrm{Log}_{\mathbf{0}}$ is a C^∞ global diffeomorphism

To prove that $\mathrm{Log}_{\mathbf{0}}$ is a bijective global diffeomorphism onto $\mathcal{B}_{1/\alpha}$, we construct its inverse, the Riemannian Exponential map $\mathrm{Exp}_{\mathbf{0}}(\mathbf{v})$.

Given a tangent vector $\mathbf{v} \in \mathcal{B}_{1/\alpha}$, let $v = \|\mathbf{v}\|$. We uniquely solve for the unbounded Euclidean norm $r = \|\mathbf{x}\|$:

$$v = \frac{r}{1 - \alpha\epsilon + \alpha\sqrt{r^2 + \epsilon^2}} \tag{26}$$

$$\alpha v\sqrt{r^2 + \epsilon^2} = r - v(1 - \alpha\epsilon) \tag{27}$$

$$(1 - \alpha^2 v^2)r^2 - 2v(1 - \alpha\epsilon)r + v^2(1 - 2\alpha\epsilon) = 0 \tag{28}$$

Applying the quadratic formula yields two roots $r_\pm = \dfrac{v(1-\alpha\epsilon) \pm \alpha v\sqrt{\epsilon^2 + v^2(1-2\alpha\epsilon)}}{1-\alpha^2 v^2}$. To select between them, we use the pre-squaring constraint $\alpha v\sqrt{r^2+\epsilon^2} = r - v(1-\alpha\epsilon) \geq 0$, i.e. $r \geq v(1-\alpha\epsilon)$. Direct substitution gives $r_+ - v(1-\alpha\epsilon) = \frac{\alpha v(\sqrt{\epsilon^2+v^2(1-2\alpha\epsilon)}+\alpha v^2(1-\alpha\epsilon))}{1-\alpha^2v^2} \geq 0$, satisfying the constraint, while $r_- - v(1-\alpha\epsilon) = \frac{\alpha v(\alpha v^2(1-\alpha\epsilon)-\sqrt{\epsilon^2+v^2(1-2\alpha\epsilon)})}{1-\alpha^2v^2} < 0$ for sufficiently small v, violating it. We therefore retain $r = r_+$:

$$r = \frac{v(1-\alpha\epsilon) + \alpha v\sqrt{\epsilon^2 + v^2(1-2\alpha\epsilon)}}{1-\alpha^2 v^2} \tag{29}$$

Finally:

$$\mathbf{x} = \mathrm{Exp}_{\mathbf{0}}(\mathbf{v}) = r\frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{30}$$

This inverse mapping is uniquely defined, real, and continuous if and only if the denominator is positive: $1-\alpha^2v^2 > 0$, which enforces the domain constraint $\|\mathbf{v}\| < \frac{1}{\alpha}$. Furthermore, the radicand $\epsilon^2 + v^2(1-2\alpha\epsilon)$ is positive for all v given the definition's numerical constraint $\epsilon < \frac{1}{2\alpha}$.

Since the inverse $\mathrm{Exp}_{\mathbf{0}}$ is defined and C^∞ smooth everywhere on the open ball $\mathcal{B}_{1/\alpha}$, the mapping is a C^∞ global diffeomorphism. ■

A.8 Proof of Theorem 5

Proof. Let $\mathbf{x} = \phi_1(\mathbf{S}) \in \mathbb{R}^n$ represent the unbounded Log-Euclidean vector representation of an SPD matrix. By Theorem 4, the mapping $\mathbf{v} = \mathrm{Log}_{\mathbf{0}}^{RCM}(\mathbf{x})$ acts as a global diffeomorphism from $\mathbb{R}^n$ onto the open Euclidean ball $\mathcal{B}_{1/\alpha}$. By setting the conformal scale parameter to $\alpha = 2\sqrt{c}/\pi$, the codomain of this intermediate mapping is confined to the open ball $\mathcal{B}_{\pi/(2\sqrt{c})}$. By Proposition 2, this bounded domain is the maximum Euclidean injectivity radius within which the spherical exponential map $\exp_{\mathbf{0}}^{\mathbb{D}_c^n}(\mathbf{v})$ remains a valid diffeomorphism, avoiding the $\tan$ asymptote and the cut locus. Because the composition of diffeomorphisms yields a diffeomorphism, Φ_3 is a diffeomorphism.

Because RCM compresses $\mathbb{R}^n$ into this domain, this 3-stage composition enables a global diffeomorphism between the entire SPD cone and the Projected Hypersphere $\mathbb{D}_c^n$. $\mathbb{D}_c^n$ is topologically equivalent to a punctured sphere (excluding the singular antipode). This ensures that all SPD matrices, regardless of spectral magnitude, can be projected to $\mathbb{D}_c^n$ and inverse-mapped, thereby enabling native Riemannian operations on the manifold. ■

B The Topology vs. Geometry Tradeoff in Spherical RMLR

While the Rational Conformal Metric (RCM) establishes a global diffeomorphism from the SPD manifold to the punctured Projected Hypersphere ($\alpha = 2\sqrt{c}/\pi$), evaluating the intrinsic Riemannian logarithm $\mathrm{Log}_{p_k}(z)$ during Generalized RMLR classification introduces a theoretical vulnerability. As defined by (Chen et al., 2024a), intrinsic Riemannian classifiers depend on the Log_{p_k} map. However, on positively curved manifolds, the Log_{p_k} map becomes mathematically undefined at the cut locus (i.e., when $d(p_k, z) \geq \pi/\sqrt{c}$).

This singularity is a topological limitation of applying the RMLR framework to spherical geometry, not a defect of our proposed nested pullback framework. We offer solutions: to mathematically guarantee the avoidance of the cut locus, one could tighten the RCM scale parameter to $\alpha = 4\sqrt{c}/\pi$. This restricts the distance from the origin for any mapped point to $d(\mathbf{0}, \cdot) < \frac{\pi}{2\sqrt{c}}$. By the triangle inequality, the maximum Riemannian distance between any data point z and class anchor p_k satisfies $d(p_k, z) \leq d(p_k, \mathbf{0}) + d(\mathbf{0}, z) < \pi/\sqrt{c}$.

This strictly confines the representation space to an open hemi-hypersphere. While this preserves the local geometry (positive constant curvature), it fundamentally alters the topology, sacrificing half of the manifold's volume, abandoning the canonical structure of the Projected Hypersphere (Skopek et al., 2020), and potentially inviting class crowding.

Our framework deliberately preserves the full topology of the Projected Hypersphere ($\alpha = 2\sqrt{c}/\pi$) to maximize class separability. We hypothesize that cross-entropy optimization empirically penalizes anchor divergence, tethering class anchors p_k to their respective data distributions z. To empirically validate this, we tracked the dynamic Riemannian distances between data and anchors throughout training. Across all folds, the injectivity radius violation rate was exactly 0.00%. The optimizer natively prevented the evaluation of the Log_{p_k} map at the antipodal cut locus, ensuring C^∞ gradient smoothness without requiring topological amputation.

For completeness, we ablated our framework using the hemi-hyperspherical bound ($\alpha = 4\sqrt{c}/\pi$). As expected, because optimization naturally prevents cut-locus violations even under the looser bound, the empirical performance difference between the two topologies was negligible. Formulating a Spherical RMLR that inherently avoids the cut-locus singularity *absolutely*, without restricting the topological volume, potentially via extrinsic ambient coordinate representations, remains a compelling direction for future work.

C Experimental Methodology Extended

While standard RMLR formulations evaluate class probabilities using standard metrics (such as AIRM or LEM), our nested mechanism allows us to impose custom geometric priors. By composing the Log-Euclidean mapping with the κ-stereographic exponential map (as derived before), we induce a family of composite Riemannian metrics on the SPD manifold. We denote these as **NHM** (Nested Hyperbolic Metric, $K < 0$) and **NSM** (Nested Spherical Metric, $K > 0$). This framework allows the subsequent RMLR layer to compute decision boundaries that simultaneously respect the symmetric positive definiteness of the data and the intrinsic relational curvature of the dataset.

C.1 Datasets and Preprocessing

Empirical validation is conducted on three popular datasets to assess representation learning on SPD manifolds:

- **First-Person Hand Action (FPHA) (Garcia-Hernando et al., 2018):** Representing human kinematics, this task utilizes 3D joint trajectories converted into temporal covariance matrices (63×63 dimensions). Performance is reported on the official test set.
- **HDM05:**(Müller et al., 2007) A high-dimensional (93×93 dimensions), highly multi-class (117 classes) 3D skeletal tracking dataset, evaluated via 5-fold cross-validation.
- **Radar Target Classification (Brooks et al., 2019):** A signal processing benchmark consisting of 20×20 covariance matrices captured under varying signal-to-noise conditions. Generalization is evaluated via 5-fold cross-validation.

To isolate metric inductive bias from network capacity, we evaluate three architectures::

1. Linear Probe: Maps inputs directly to the classification space, isolating the linear separability induced solely by the metric mapping.

2. SPDNet: Incorporates sequential Bilinear Mapping (BiMap) and Rectified Eigenvalue (ReEig) layers (Huang & Van Gool, 2017). To prevent confounding factors from information bottlenecking, manifold dimensions are strictly preserved throughout the forward pass ($[63 \to 63]$ for FPHA, $[20 \to 20]$ for Radar, $[93 \to 93]$ for HDM05).

3. Deep RResNet: A deep Riemannian Residual architecture (Katsman et al., 2023) incorporating manifold-constrained skip connections. ($[63 \to 63]$ for FPHA, $[20 \to 20]$ for Radar, $[93 \to 93]$ for HDM05).

All models were implemented in PyTorch (Paszke et al., 2019) and trained on NVIDIA T4 GPUs. Reproducibility is ensured through a global random seed initialization (`seed=42`) across all operations. Network weights are optimized using Riemannian Adam (`geoopt`) configured with a 10^{-3} learning rate, 10^{-4} weight decay, and a batch size of 32. Gradient clipping is capped at an L_2 norm of 2.0. Training proceeds for a maximum of 100 epochs, governed by an early stopping mechanism (15-epoch patience) tied to a 20% validation holdout to prevent overfitting.

Numerical Safeguards: Standard 32-bit eigendecomposition frequently triggers `LinAlgError` exceptions on rank-deficient kinematic covariances. To ensure stable automatic differentiation through our nested mappings, we introduce a numerical failsafe. Input matrices are symmetrized ($\mathbf{X} = \frac{1}{2}(\mathbf{X} + \mathbf{X}^\top)$) and upcast to 64-bit precision. To prevent gradient degeneration during pullback computations, a microscopic, uniformly distributed tie-breaking vector ($\sim 10^{-8}$) is injected into the matrix diagonal prior to decomposition, ensuring isolated eigenvalues and well-conditioned Fréchet derivatives.

Native Riemannian Classification For intrinsic classification, we replace standard Euclidean linear layers with the generalized Riemannian Multinomial Logistic Regression (RMLR) (Chen et al., 2024a), equipped with the target manifold's metric. RMLR constructs geodesic hyperplanes directly on the manifold. We benchmark our proposed metrics, NHM (using Hyperbolic RMLR), and NSM-RCM (using Spherical RMLR, stabilized with RCM), against the popular Log-Euclidean Metric (LEM, flat curvature), standard NSM (without RCM intervention), and the Affine-Invariant Riemannian Metric (AIRM, non-positive curvature, using AIRM RMLR). Due to extreme computational intractability on high-dimensional datasets, Tangent Space AIRM is substituted for Native AIRM RMLR for the HDM05 evaluations.

Furthermore, to measure the intrinsic geometry of our datasets without introducing metric bias, we project the matrices to the Log-Euclidean tangent space and normalize the pairwise distances $d \in [0, 1]$ to prevent Optimal Transport solver divergence. We then construct a k-Nearest Neighbor graph ($k = 5$) and compute the discrete Ollivier-Ricci Curvature (ORC) via random walk diffusion ($\alpha = 0.5$).

C.2 Validation via Synthetic Geometric Control

To validate the classification bounds of RMLR equipped with the Nested Spherical Metric (NSM), we synthesize a covariance dataset localized to a compact, positively curved submanifold. Generating data with explicit global positive curvature in high-dimensional SPD spaces requires structural construction, as standard ambient matrix distributions do not natively guarantee these geometric properties.

We achieve this by embedding a 3-dimensional spherical data distribution directly into the ambient space of symmetric matrices $\mathcal{S}^{20}$. We define three class clusters using spherical coordinates (θ, ϕ) scaled by a radial parameter r. These coordinates are embedded into the top-left 2×2 principal submatrix, padded with low-magnitude isotropic noise ($\mathcal{N}(0, 0.1)$) across the remaining dimensions, and mapped onto the SPD manifold via the matrix exponential.

By modulating the radial parameter r, we dictate the prominence of the spherical manifold structure relative to the ambient noise, synthesizing distributions of opposing intrinsic curvatures:

- **Macroscopic ($r = 1.5$):** The spherical signal dominates the ambient noise ($\mathcal{N}(0, 0.1)$), forming distinct clusters separated by angular great-circle arcs along the curved shell. This empirically induces positive intrinsic curvature (ORC: $+0.010$). The Nested Spherical Metric (NSM) natively matches this geometry.
- **Microscopic ($r = 0.25$):** The spherical signal collapses into the noise floor, producing a dense Gaussian cluster near the origin. This shift induces strongly negative intrinsic curvature (ORC: -0.213). In this negatively curved geometry, NSM's sub-Euclidean volume growth artificially crowds the space and fails to resolve the margins. Conversely, NHM's super-Euclidean volume growth aligns with the negative intrinsic curvature, enhancing class separability and restoring accuracy.

C.3 Geometric Regularization against Adversarial Outliers

To empirically validate the Rational Conformal Metric's use as a bounded Euclidean feature map for outlier regularization , independent of its role as an intermediate chart in Theorem 5, we evaluate its robustness against spatially extreme anomalies. Because the standard Log-Euclidean tangent space is unbounded, large-magnitude outliers incur large penalties under cross-entropy optimization. This forces linear decision boundaries to shift, systematically degrading the separability of the core data distribution.

We define four linearly separable core classes (200 total samples) centered in the four quadrants of a 2D Euclidean feature space. To evaluate geometric robustness, we inject strategic spatial anomalies designed to disrupt linear separability:

1. **Cross-Boundary Outlier Injection:** Anomalies belonging to Class 3 are embedded deep within the spatial domain of Class 0.
2. **Directional Heavy-Tailed Noise:** A high-variance structural noise tail from Class 1 extends linearly across the origin, intersecting opposing class distributions.
3. **Symmetric Extreme Outliers:** Distant, high-density point masses for Classes 0 and 3 are placed at opposite spatial extremes, generating high-magnitude gradient signals that disproportionately dictate the hyperplane optimization.

Under standard optimization (Adam, lr=0.1, 600 epochs to ensure convergence), the unbounded Log-Euclidean baseline misclassifies the dense core clusters to minimize the disproportionate cross-entropy loss incurred by the outliers. Conversely, mapping these identical datasets through the Rational conformal transformation applies a non-linear radial compression. By asymptotically confining all representations within the open ball of radius $1/\alpha$, the spatial magnitude and corresponding gradient contribution of the anomalies are bounded. This geometric regularization isolates the linear optimization to the core distributions, recovering 100% accuracy when using RCM without explicit outlier filtering.

D Extended Outlier Regularization Results

D.1 Ablation Details

To isolate the structural advantages of this specific radial compression, we ablated the Rational mapping against standard Euclidean normalization techniques using the directional heavy-tailed distribution (Figure 1, Appendix D.2).

- L_2 **Normalization ($\mathbf{v}/\|\mathbf{v}\|_2$):** While L_2 projection mathematically bounds outlier gradients and successfully restores core separability (Core Accuracy: 100.0%, Total Accuracy: 91.8%), it eliminates intrinsic data magnitude. This projects the 2D Gaussian distributions onto fixed-radius 1D arcs, discarding the latent variance of the primary data.
- **Tanh ($\tanh(\beta\mathbf{v})$):** Applying an element-wise hyperbolic tangent bounds the vector space but introduces anisotropic distortion by compressing the space toward a hypercube. This non-uniform scaling skews inter-class angles, preventing the recovery of optimal linear margins (Core Accuracy: 92.5%, Total Accuracy: 86.4%).
- **Proposed RCM:** The proposed mapping provides the optimal geometric compromise. It enforces the mathematical bounds necessary to limit outlier impact while preserving the relative spatial variance and angular separation of the core distributions near the origin. This yields superior global performance (Core Accuracy: 100.0%, Total Accuracy: 92.3%) without inducing dimensional reduction or angular distortion.

D.2 Adversarial Outlier Regularization Visualizations

This section provides the decision boundary visualizations for the geometric regularization experiments.

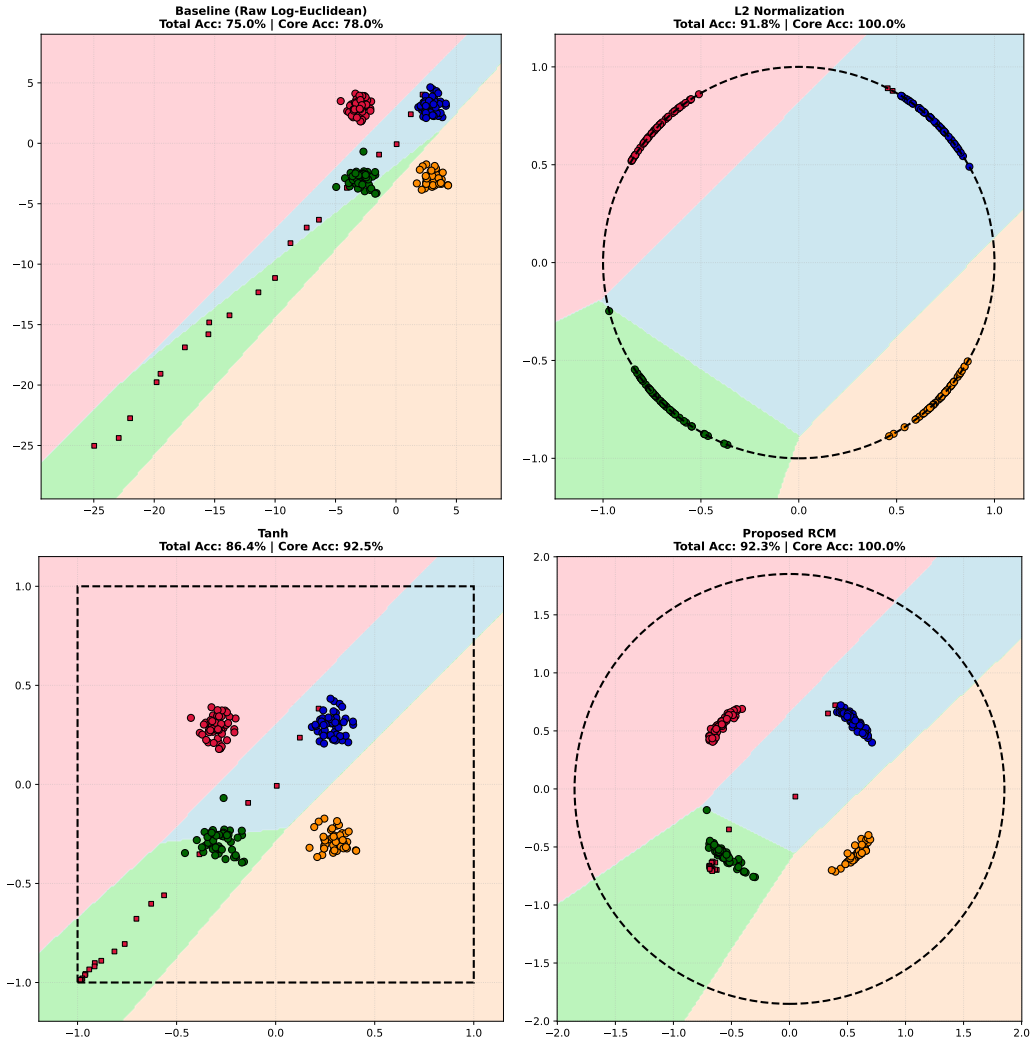

Figure 1: Ablation of geometric regularization techniques on the directional heavy-tailed distribution. While L_2 normalization (top right) restores core accuracy, it collapses the 2D Gaussian distributions into rigid 1D arcs, discarding intrinsic data magnitude, leading to lesser total accuracy than RCM. Tanh squashing (bottom left) bounds the space but introduces anisotropic box distortion, skewing inter-class angles. The proposed Rational Conformal Metric (bottom right) optimally bounds outlier magnitude while preserving the relative spatial variance and angular separation of the primary data.

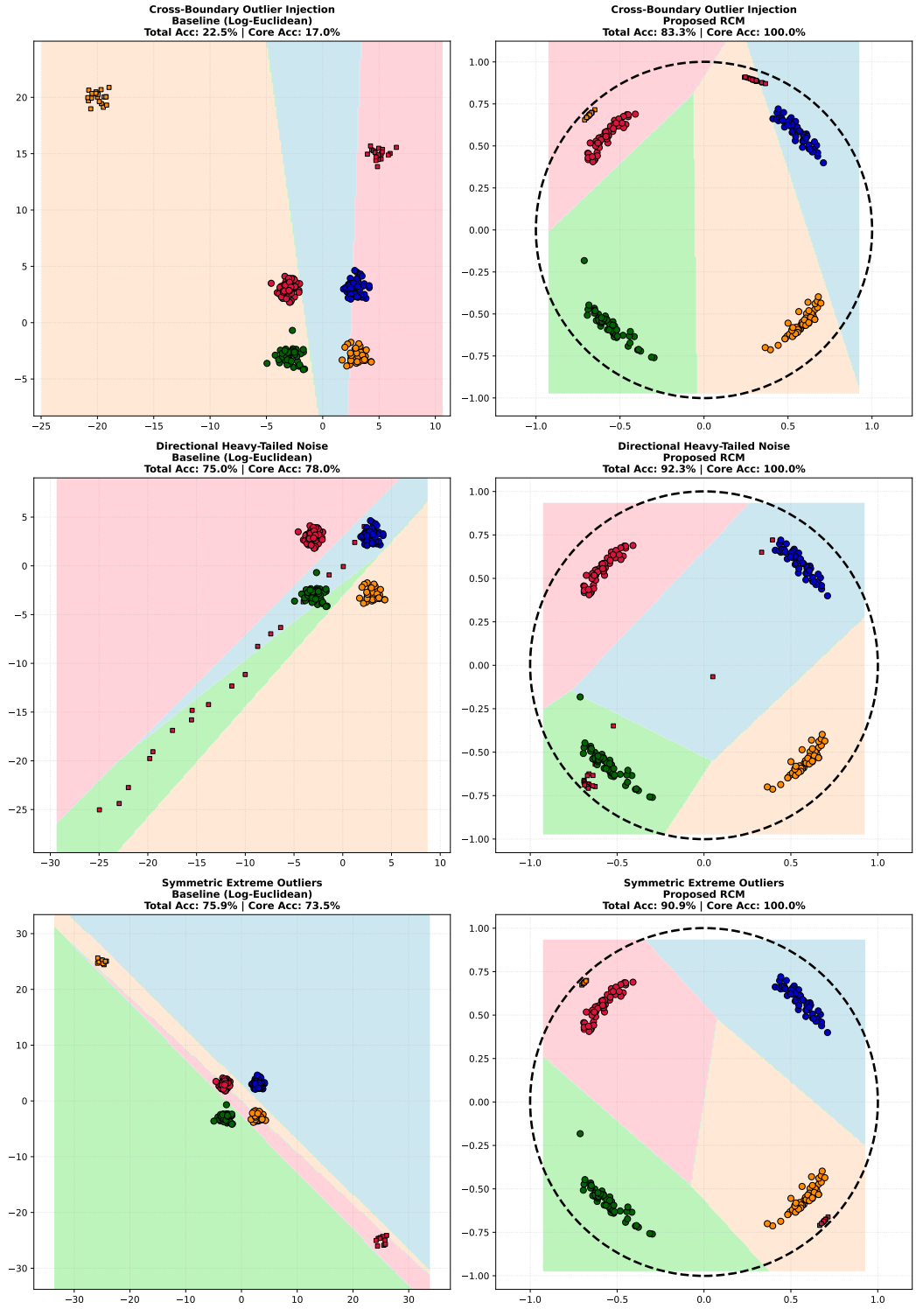

Figure 2: Decision boundaries under adversarial spatial anomalies across three topologies: cross-boundary outlier injection (top), directional heavy-tailed noise (middle), and symmetric extreme outliers (bottom). The unbounded Log-Euclidean baseline (left) systematically shifts hyperplanes to minimize outlier cross-entropy penalties, degrading core separability. The proposed Rational Conformal Metric (right) asymptotically confines outliers within a fixed radial envelope (dashed circle), mathematically bounding their gradient contribution and restoring perfect core accuracy.