RIEMANNIAN SLICED-WASSERSTEIN DISTANCES ON CORRELATION MATRICES

Anonymous authors
Paper under double-blind review

ABSTRACT

Riemannian sliced-Wasserstein distances make distribution comparison on SPD manifolds practical through one-dimensional projections, but covariance-based variants remain sensitive to marginal-variance changes even when the correlation structure is unchanged. We introduce correlation sliced-Wasserstein (CorSW) on full-rank correlation matrices. Its flat branch covers ECM, LECM, OLM, and LSM through their Euclidean isometries. On the product-hyperbolic Cholesky metric, Product-Busemann aggregates all factor coordinates along one global product direction before Wasserstein matching, Product-Geodesic combines exact factor coordinates, whose relation to nearest-product-geodesic projection is characterized. PHCML further joins the row factors by Lorentz direct concatenation for closed-form Log, Busemann, and geodesic slices. We establish metricity for the Log/geodesic routes and for the Busemann routes under the corresponding injectivity conditions. Across nine benchmarks, CorSW performs well, with complementary behavior across its product and HCat routes and low loss-step cost in high dimensions.

1 INTRODUCTION

Symmetric positive-definite (SPD) covariance matrices widely represent second-order statistics in medical imaging (Chakraborty et al., 2018), signal processing (Brooks et al., 2019), and computer vision (Huang & Van Gool, 2017). In brain–computer interfaces (BCIs) and electroencephalogram (EEG) decoding, however, amplitude shifts across sessions and subjects can let high-power channels dominate weaker task-relevant dependencies (Blankertz et al., 2007; Hinss et al., 2021). Correlation matrices normalize marginal variances and thus represent dependence invariant to such scaling, which is useful when amplitude is a nuisance rather than the signal of interest. Yet correlation geometry remains less explored than covariance geometry in deep learning and distribution alignment.

Recent work has established several Riemannian structures for full-rank correlation matrices $\mathrm{Cor}^{+}(d)$. The original SPD quotient construction lacks unique closed forms for several operations (Thanwerdas, 2022), whereas the Euclidean-Cholesky (ECM), Log-Euclidean-Cholesky (LECM), Off-Log (OLM), and Log-Scaled (LSM) metrics are flat through Euclidean isometries (Thanwerdas & Pennec, 2022; 2024b). PHCM instead represents normalized Cholesky rows as separate hyperbolic factors (Thanwerdas & Pennec, 2022). Following product-manifold slicing (Bonet et al., 2025), we retain this canonical product geometry and use one global direction whose block norms set the factor speeds. The resulting exact Product-Busemann coordinate aggregates factor coordinates before Wasserstein matching, retaining cross-factor dependence discarded by factor-wise losses. Complementarily, inspired by recent joint processing of row factors over full-rank correlation matrices (Chen et al., 2026), we introduce PHCML: Lorentz direct concatenation (HCat) (Bdeir et al., 2024) maps them to one constant-negative-curvature hyperboloid, separating product and joint-Lorentz representations.

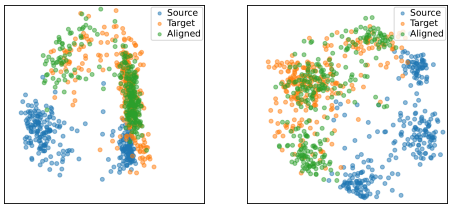


Figure 1: PCA visualization of BCI data alignment with CorSW under LSM and OLM.

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.

Optimal transport (OT) compares probability distributions and supports domain adaptation, but is often computationally demanding (Peyré & Cuturi, 2019). Sliced-Wasserstein (SW) reduces this cost by averaging one-dimensional projections (Bonneel et al., 2015). Bonet et al. (2025) extended SW to Cartan–Hadamard manifolds through geodesic and Busemann slices. On manifolds isometric to Euclidean space, these coordinates coincide up to sign, recover Euclidean SW after the isometry, and yield a metric. We instantiate this flat case through ECM, LECM, OLM, and LSM. Concurrent work by Hu et al. (2026) uses only OLM- and LSM-based losses in a specific EEG architecture and domain-shift protocol.

We propose correlation sliced-Wasserstein (CorSW) across flat, product, and joint-Lorentz representations. The flat route unifies ECM, LECM, OLM, and LSM under one objective. The PHCM route provides Product-Log, Product-Busemann, and Product-Geodesic coordinates. Product-Geodesic combines exact signed geodesic coordinates from the factors according to one global product direction. It generally differs from nearest-point projection onto the product geodesic, and we characterize their relationship and coincidence condition. PHCML uses HCat to support tangent-space Log and intrinsic Busemann and geodesic slicing on one hyperboloid. We evaluate all variants as distribution regularizers and source–target alignment losses, both alone and fused with covariance alignment.

In summary, our main contributions are as follows.

- We formulate CorSW under four pullback-flat metrics and the original PHCM product geometry: the former reduce exactly to Euclidean SW, while the latter retains the row-wise product structure.
- We derive the exact Product-Busemann slice, prove that pre-Wasserstein aggregation retains cross-factor dependence, characterize Product-Geodesic against nearest-point projection, and introduce PHCML-HCat with closed-form Log, Busemann, and geodesic slices.
- Across controlled transport and nine benchmarks spanning SPD classification, EEG transfer, and graph learning, CorSW performs well across tasks, and its product and HCat routes show complementary behavior and low loss-step cost in high dimensions.

2 Preliminaries

2.1 Correlation Geometries

For $\Sigma \in \mathcal{S}_d^{++}(\mathbb{R})$, let $\mathrm{Cor}(\Sigma) = \mathbb{D}(\Sigma)^{-1/2}\Sigma\mathbb{D}(\Sigma)^{-1/2}$, where $\mathbb{D}(\cdot)$ extracts the diagonal. The correlation manifold is $\mathrm{Cor}^+(d) = \{C \in \mathcal{S}_d^{++}(\mathbb{R}) \mid \mathbb{D}(C) = I_d\}$, whose tangent space is $T_C\mathrm{Cor}^+(d) = \mathrm{Hol}(d)$ (Thanwerdas, 2022). Recent work has provided several convenient Riemannian models for $\mathrm{Cor}^+(d)$, each obtained by pulling back a metric from a simpler prototype space. Four of them, ECM, LECM, OLM, and LSM, are flat through global Euclidean coordinates (Thanwerdas & Pennec, 2022; 2024b), whereas PHCM retains a product-hyperbolic structure (Thanwerdas & Pennec, 2022). We briefly summarize the five geometries used in this paper.

ECM. With $L = \mathrm{Chol}(C)$ and $\Theta(C) = \mathbb{D}(L)^{-1}L$, ECM uses $\phi^{EC}(C) = \lfloor\Theta(C)\rfloor$ and pulls back the Euclidean metric from $\mathrm{LT}^0(d)$ (Thanwerdas & Pennec, 2022). It is therefore a flat geometry expressed in normalized Cholesky coordinates.

LECM. LECM applies the matrix logarithm after Θ and uses $\phi^{LEC}(C) = \log\Theta(C)$, again yielding a flat pullback geometry on $\mathrm{LT}^0(d)$ (Thanwerdas & Pennec, 2022). Relative to ECM, it replaces the affine Cholesky chart by its logarithmic counterpart while preserving a Euclidean prototype space.

OLM. OLM uses the off-diagonal logarithmic coordinate $\mathrm{Log}^\circ(C) = \mathrm{off}(\log C)$ and pulls back the Euclidean metric from $\mathrm{Hol}(d)$ (Thanwerdas & Pennec, 2024b). This construction is permutation-invariant and remains flat in the ambient model space.

LSM. LSM uses the row-sum-zero logarithmic coordinate $\mathrm{Log}^\star(C) = \log\left(\mathcal{D}^\star(C)C\mathcal{D}^\star(C)\right)$ and pulls back the Euclidean metric from $\mathrm{Row}_0(d)$ (Thanwerdas & Pennec, 2024b). Like OLM, it is permutation-invariant, but it uses a diagonally rescaled logarithmic representation whose rows sum to zero.

PHCM. PHCM is induced by the normalized Cholesky rows and pulls back a product-hyperbolic metric (Thanwerdas & Pennec, 2022). Unlike the other four geometries, PHCM is not flat. Its factor

planes are negatively curved, while the cross-factor planes are flat. The factorization used below is detailed in Sec. 3.3. This product structure is the basis of our Product-Log and Product-Busemann slices and Product-Geodesic (approx.) surrogate.

2.2 Wasserstein and Sliced Distances

Wasserstein distance. For probability measures $\mu, \nu \in \mathcal{P}_p\left(\mathbb{R}^k\right)$ with finite p-th moments and $p \geq 1$, the p-Wasserstein distance is defined by the optimal transport problem (Villani, 2009; Peyré & Cuturi, 2019)

$$W_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^k \times \mathbb{R}^k} \|x - y\|_2^p \mathrm{d}\pi(x, y), \tag{1}$$

where $\Pi(\mu, \nu)$ is the coupling set. For discrete measures, direct computation costs $\mathcal{O}(n^3 \log n)$ (Peyré & Cuturi, 2019).

When $k = 1$, the Wasserstein distance admits a closed form in terms of quantile functions (Rabin et al., 2011; Bonneel et al., 2015).

$$W_p^p(\mu, \nu) = \int_0^1 \left|F_\mu^{-1}(u) - F_\nu^{-1}(u)\right|^p \mathrm{d}u. \tag{2}$$

For empirical measures, Equation (2) costs $\mathcal{O}(n \log n)$ by sorting.

Sliced-Wasserstein distance. The *Sliced-Wasserstein* (SW) distance exploits the one-dimensional advantage in Equation (2) by averaging Wasserstein distances over random scalar slices (Rabin et al., 2011; Bonneel et al., 2015; Kolouri et al., 2019). Let $s_\theta : \mathcal{M} \to \mathbb{R}$ be a slice coordinate associated with direction $\theta \in \Omega$, and let $(s_\theta)_{\#}\mu$ and $(s_\theta)_{\#}\nu$ denote the corresponding pushforward measures. The SW distance is defined as

$$\mathrm{SW}_p^p(\mu, \nu) = \int_\Omega W_p^p\left((s_\theta)_{\#}\mu, (s_\theta)_{\#}\nu\right) \mathrm{d}\sigma(\theta), \tag{3}$$

where σ is a direction law. With L Monte Carlo directions, the empirical cost is $\mathcal{O}(Ln \log n)$. Bonet et al. (2025) provide the Cartan–Hadamard extensions used below.

3 Slices on Correlation Matrices

Euclidean SW projects $y \in \mathbb{R}^m$ onto the line $G_\theta = \{t\theta \mid t \in \mathbb{R}\}$. For a unit direction θ, the projected point is $\widetilde{P}^\theta(y) = P^\theta(y)\theta$, and its scalar coordinate is the core Euclidean slice

$$\boxed{P^\theta(y) = \langle y, \theta \rangle = \arg\min_{t \in \mathbb{R}} \|t\theta - y\|_2.} \tag{4}$$

On a manifold, the key is to find the counterpart of this inner product: the nearest-point form above leads to geodesic projection.

3.1 Manifold Slices

Let $(\mathcal{M}, g)$ be a Hadamard manifold with origin o, and let $\gamma_v(t) = \exp_o(tv)$ be the complete geodesic in a unit direction $v \in T_o\mathcal{M}$. The *geodesic coordinate* is the parameter of the closest point on γ_v (Bonet et al., 2025, Sec. 3.2):

$$P^v(x) = \arg\min_{t \in \mathbb{R}} d\left(\gamma_v(t), x\right). \tag{5}$$

The Euclidean level sets $\{y \mid \langle y, \theta \rangle = c\}$ are hyperplanes orthogonal to θ. Their Hadamard counterparts are the horospheres defined by the *Busemann coordinate* $B^v(x) = \lim_{t \to \infty}\left(d(x, \gamma_v(t)) - t\right)$ (Bridson & Haefliger, 2013, Definition 8.17). For $\mathcal{M} = \mathbb{R}^m$ and $\gamma_\theta(t) = t\theta$, the two coordinates reduce to $P^\theta(x) = \langle x, \theta \rangle$ and $B^\theta(x) = -\langle x, \theta \rangle$ (Chami et al., 2021; Bonet et al., 2025).

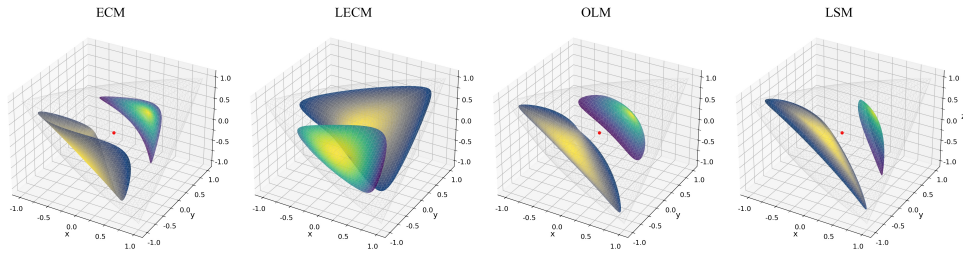


Figure 2: Horospheres under the four flat correlation geometries on $\mathrm{Cor}^+(3)$, embedded in the open elliptope by (c_{21}, c_{31}, c_{32}). Gray marks the boundary, and color encodes the minimum eigenvalue.

3.2 Flat Correlation Slices

Let $\phi : \mathrm{Cor}^+(d) \to V$ be one of the four embeddings above, so that ϕ is an isometry onto the Euclidean space V. It maps geodesics from I_d to straight lines in V and sends $v \in T_{I_d}\mathrm{Cor}^+(d)$ to the model-space direction $d\phi_{I_d}(v)$. Hence the required scalar coordinate is exactly an inner product after embedding.

Corollary 3.1 (Inner-product equivalence)**.** *For any isometry $\phi : \mathrm{Cor}^+(d) \to V$ onto a Euclidean space V and unit $v \in T_{I_d}\mathrm{Cor}^+(d)$,*

$$P^v_\phi(C) = -B^v_\phi(C) = \langle \phi(C) - \phi(I_d), d\phi_{I_d}(v) \rangle_V . \tag{6}$$

Following Bonet et al. (2025, Prop. 7), we use $\Pi^v_\phi := B^v_\phi$. The geodesic convention changes only the sign.

Corollary 3.2. *Let $L = \mathrm{Chol}(C)$, $\Theta(C) = \mathbb{D}(L)^{-1}L$, $X = \log C$, and $\widetilde{X} = \log(\mathcal{D}^\star(C)C\mathcal{D}^\star(C))$. Write $T_v = d\Theta_{I_d}(v)$, $R_v = d(\log\circ\Theta)_{I_d}(v)$, and $V_\star = d\mathrm{Log}^\star_{I_d}(v)$. The corresponding slice coordinates are*

$$\begin{aligned}
\Pi^v_{\mathrm{EC}}(C) &= -\langle \Theta(C) - I_d, T_v \rangle_F , \\
\Pi^v_{\mathrm{LEC}}(C) &= -\langle \log\Theta(C), R_v \rangle_F , \\
\Pi^v_{\mathrm{OLM}}(C) &= -\langle X, v \rangle^{(\alpha,\beta,\gamma)} , \\
\Pi^v_{\mathrm{LSM}}(C) &= -\left\langle \widetilde{X}, V_\star \right\rangle^{(\alpha,\delta,\zeta)} .
\end{aligned} \tag{7}$$

For OLM, (α, β, γ) weight its trace, sum-of-product, and product-of-sums terms. For LSM, (α, δ, ζ) weight its trace, diagonal, and product-of-traces terms.

ECM and LECM use the Frobenius inner product. For both OLM and LSM, we set the parameter triple to $(1, 0, 0)$, yielding the canonical Euclidean inner product. Supplementary Appendix C.3 gives the full inner products and their dimension-dependent admissibility conditions.

Figure 2 visualizes the horospheres $\{C \mid \Pi^v_\phi(C) = t\}$ inside the elliptope for fixed v and varying t. They are the preimages under ϕ of parallel affine hyperplanes in the model space, and their distinct shapes show the effect of each embedding in the original correlation coordinates.

3.3 PHCM Product Identification and Slices

Proposition 4.1 of Chen et al. (2026) identifies the correlation manifold diffeomorphically with a product of Poincaré spaces through the composite map $\Psi_{\mathrm{prod}} \circ \mathrm{Chol}$:

$$C \xmapsto{\mathrm{Chol}} L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ L_{21} & L_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{d1} & L_{d2} & \cdots & L_{dd} \end{pmatrix} \xmapsto{\Psi_{\mathrm{prod}}} \begin{pmatrix} \psi_1(h_1) \\ \vdots \\ \psi_{d-1}(h_{d-1}) \end{pmatrix} \in \mathbb{P}\mathbb{P}^{d-1} := \prod_{k=1}^{d-1} \mathbb{P}^k . \tag{8}$$

Here $h_{i-1} = (L_{i1}, \ldots, L_{ii})^\top \in \mathrm{HS}^{i-1}$ and $\psi_{i-1} = \psi_{\mathrm{HS}^{i-1} \to \mathbb{P}^{i-1}}$ for $2 \le i \le d$. This identification is the common starting point of our two curved constructions. First, following the product-manifold construction of Bonet et al. (2025), we retain the PHCM product and slice all factors with one

global direction, rather than using independent factor-wise losses. Second, we further join the same factors through Lorentz direct concatenation to obtain the PHCML-HCat route in Sec. 3.4. Since the hemisphere, Poincaré, and hyperboloid models give equivalent descriptions of each hyperbolic factor, we use hyperboloid coordinates below for their convenient closed forms. Thus $\mathcal{M}_{\mathrm{PHCM}} = \prod_{i=2}^{d} \mathbb{H}^{i-1}$ is a Cartan–Hadamard product manifold. Let $C = LL^\top$, $L = \mathrm{Chol}(C)$, and define

$$r_i = \frac{L_{i,1:i-1}}{L_{ii}}, \quad \ell_{i,0} = \sqrt{1 + \|r_i\|_2^2}, \quad x_i = (\ell_{i,0}, r_i) \in \mathbb{H}^{i-1}. \tag{9}$$

A global unit direction $u \in S^{D-1}$, where $D = d(d-1)/2$, decomposes as $u = (u_2, \ldots, u_d)$. Set $a_i = \|u_i\|_2$ and $v_i = u_i/a_i$ when $a_i > 0$. Then $\sum_i a_i^2 = 1$, and the product geodesic has speed a_i in factor i.

Corollary 3.3 (PHCM product coordinates)**.** *Let $d_i = \mathrm{asinh}(\|r_i\|_2)$ and let $\Xi_i = d_i r_i/\|r_i\|_2$, with $\Xi_i = 0$ when $r_i = 0$. The Product-Log, Product-Busemann, and Product-Geodesic (approx.) coordinates used in our experiments are*

$$\begin{aligned} s_u^{\text{Prod-Log}}(C) &= \sum_{i=2}^{d} \langle \Xi_i, u_i \rangle, \\ s_u^{\text{Prod-B}}(C) &= \sum_{i:a_i>0} a_i \log\left(\ell_{i,0} - \langle r_i, v_i \rangle\right), \\ s_u^{\text{Prod-G}}(C) &= \sum_{i:a_i>0} a_i \,\mathrm{arctanh}\left(\frac{\langle r_i, v_i \rangle}{\ell_{i,0}}\right). \end{aligned} \tag{10}$$

Prod-Log projects after the product logarithmic map, while Prod-B is the exact Busemann coordinate of the product ray. Product-Geodesic (approx.) combines the exact factor-wise signed coordinates $c_i = \mathrm{arctanh}(\langle r_i, v_i \rangle/\ell_{i,0})$ as $\sum_i a_i c_i$. In contrast, the intrinsic coordinate in Eq. (5) *is*

$$\begin{aligned} t_u^\star(C) &= \arg\min_{t \in \mathbb{R}} \sum_{i:a_i>0} \mathrm{arccosh}^2(m_i \cosh(a_i t - c_i)), \\ m_i &= \sqrt{\ell_{i,0}^2 - \langle r_i, v_i \rangle^2}. \end{aligned} \tag{11}$$

Thus, the experimental Prod-G (approx.) is a distinct closed-form surrogate, not the intrinsic G slice of Eq. (5)*. They coincide when $m_i = 1$ for every active factor, equivalently $r_i \parallel v_i$, because the objective becomes $\sum_i (a_i t - c_i)^2$ and $t_u^\star = \sum_i a_i c_i$. Supplementary Appendix H.8 gives the full objective, stationarity equation, bracketing interval, and proof.*

Product-B proof sketch. For the unit-speed product ray $\gamma_u(t) = (\gamma_{v_2}(a_2 t), \ldots, \gamma_{v_d}(a_d t))$, the factor expansion is $d_{\mathbb{H}}(x_i, \gamma_{v_i}(a_i t)) = a_i t + b_{v_i}(x_i) + o(1)$. Using $\sum_i a_i^2 = 1$ in the product norm gives $d_{\mathrm{Prod}}(x, \gamma_u(t)) = t + \sum_i a_i b_{v_i}(x_i) + o(1)$. Subtracting t proves the Product-B formula. Supplementary Appendix H.8 gives the full derivation and moment assumptions.

3.4 PHCML-HCat Slices

Building on the PHCM factors $x_i = (\ell_{i,0}, r_i)$ from Eq. (9), we further propose PHCML-HCat: Lorentz direct concatenation (HCat) (Bdeir et al., 2024) joins the factors before slicing. Denote the preceding product identification by $\Phi_{\mathrm{PHCM}} := \Psi_{\mathrm{prod}} \circ \mathrm{Chol}$. The resulting pipeline is

$$\begin{aligned} C \xmapsto{\Phi_{\mathrm{PHCM}}} \boldsymbol{x}_{\mathrm{PHCM}} &= (x_2, \ldots, x_d) \\ &\in \mathcal{M}_{\mathrm{PHCM}}, \\ \boldsymbol{x}_{\mathrm{PHCM}} \xmapsto{\mathrm{HCat}} x &= \Psi(C), \\ z &= \mathrm{Concat}(r_2, \ldots, r_d) \in \mathbb{R}^D, \\ x &= \left(\sqrt{1 + \|z\|_2^2}, z\right) \in \mathbb{H}^D_{-1}. \end{aligned} \tag{12}$$

Since the PHCM coordinates $(r_2, \ldots, r_d)$ determine C uniquely, $\Psi : \mathrm{Cor}^+(d) \to \mathbb{H}^D_{-1}$ is a global diffeomorphism, and we equip $\mathrm{Cor}^+(d)$ with the pullback metric $g_\Psi = \Psi^* g_{\mathbb{H}}$. A direction $u \in S^{D-1}$ then slices all rows jointly through the closed-form Lorentz coordinates of Bonet et al. (2025).

Let $o_{\mathbb{H}} = (1, 0)$ and define the tangent coordinate $\Xi(C) = \mathrm{Log}^{\mathbb{H}}_{o_{\mathbb{H}}}(x) = \frac{\operatorname{arccosh}(x_0)}{\sqrt{x_0^2-1}}\bar{x}$, interpreted by continuity at $o_{\mathbb{H}}$. The Euclidean inner product now has three counterparts: an ordinary inner product after the logarithmic map, the Busemann function, and signed geodesic projection:

$$\begin{aligned} s_u^{\mathrm{Log}}(C) &= \langle \Xi(C), u \rangle, && \text{(Log)}, \\ s_u^{\mathrm{B}}(C) &= \log\left(x_0 - \langle \bar{x}, u \rangle\right), && \text{(Busemann)}, \\ s_u^{\mathrm{G}}(C) &= \operatorname{arctanh}\left(\frac{\langle \bar{x}, u \rangle}{x_0}\right), && \text{(Geodesic)}. \end{aligned} \tag{13}$$

The first coordinate is a tangent-space Euclidean slice. The latter two are the Lorentz horospherical and signed-geodesic coordinates of Bonet et al. (2025, Prop. 9), with s_u^{G} defined on the complete geodesic $\gamma_u(\mathbb{R})$. Their difference under negative curvature is explicit. Setting $a = \langle \bar{x}, u \rangle$ and $\bar{x}_\perp = \bar{x} - au$ gives

$$s_u^{\mathrm{B}}(C) + s_u^{\mathrm{G}}(C) = \frac{1}{2}\log(x_0^2 - a^2) = \frac{1}{2}\log\left(1 + \|\bar{x}_\perp\|_2^2\right). \tag{14}$$

Hence Busemann and geodesic coordinates differ only by sign on the slicing geodesic ($\bar{x}_\perp = 0$). Away from it, the Busemann coordinate also captures transverse displacement through the horospheres.

4 Correlation Sliced-Wasserstein Distances

4.1 Riemannian Sliced-Wasserstein

Fix a Riemannian metric g on $\mathrm{Cor}^+(d)$ and the base point I_d. A unit direction $v \in S^g_{I_d}$ defines three scalar slices:

$$s^{\mathrm{Log}}_{g,v}(C) = \langle \mathrm{Log}^g_{I_d}(C), v \rangle_{I_d}, \quad s^{\mathrm{G}}_{g,v}(C) = P^v_g(C), \quad s^{\mathrm{B}}_{g,v}(C) = B^v_g(C). \tag{15}$$

Log takes an inner product after the logarithmic map, whereas G and B use the geodesic projection and Busemann coordinates from Sec. 3.1.

Definition 4.1 (Correlation sliced-Wasserstein distance)**.** Let $p \geq 1$, let $\lambda^g_{I_d}$ be uniform on $S^g_{I_d}$, and let μ, ν be probability measures on $\mathrm{Cor}^+(d)$. For $r \in \{\mathrm{Log}, \mathrm{G}, \mathrm{B}\}$ such that the integral is finite, set $\mu^r_{g,v} = (s^r_{g,v})_{\#}\mu$ and $\nu^r_{g,v} = (s^r_{g,v})_{\#}\nu$, and define

$$\mathrm{CorSW}^p_{p,g\text{-}r}(\mu, \nu) = \int_{S^g_{I_d}} W_p^p(\mu^r_{g,v}, \nu^r_{g,v})\, \mathrm{d}\lambda^g_{I_d}(v). \tag{16}$$

Empirically, the integral is approximated with L directions.

4.2 Flat CorSW

For $\phi \in \{\phi^{EC}, \phi^{LEC}, \mathrm{Log}^\circ, \mathrm{Log}^\star\}$, let g_ϕ be the flat metric under which ϕ is an isometry onto V, and set

$$\mathcal{P}^\phi_p := \left\{ \mu \mid \int \|\phi(C) - \phi(I_d)\|_V^p \mathrm{d}\mu(C) < \infty \right\}. \tag{17}$$

For $p \geq 1$ and $\mu, \nu \in \mathcal{P}^\phi_p$, the three slices satisfy

$$s^{\mathrm{Log}}_{g_\phi,v}(C) = s^{\mathrm{G}}_{g_\phi,v}(C) = -s^{\mathrm{B}}_{g_\phi,v}(C) = \langle \phi(C) - \phi(I_d), d\phi_{I_d}(v) \rangle_V. \tag{18}$$

This follows from Cor. 3.1 and the closed-form logarithmic map $\mathrm{Log}^{g_\phi}_{I_d}(C) = (d\phi_{I_d})^{-1}(\phi(C) - \phi(I_d))$. Since reflection leaves one-dimensional Wasserstein distance unchanged and $d\phi_{I_d}$ maps the tangent-sphere law to the uniform sphere in V, Bonet et al. (2025, Lem. 25) implies that the Log, G, and B variants coincide. We therefore write

$$\mathrm{CorSW}_{p,\phi}(\mu, \nu) = \mathrm{SW}_p(\phi_{\#}\mu, \phi_{\#}\nu). \tag{19}$$

Corollary 4.2 (Flat metricity)**.** *For every $\phi \in \{\phi^{EC}, \phi^{LEC}, \mathrm{Log}^\circ, \mathrm{Log}^\star\}$ and $p \geq 1$, $\mathrm{CorSW}_{p,\phi}$ is a finite metric on $\mathcal{P}^\phi_p$ (Bonet et al., 2025, Prop. 26).*

4.3 PRODUCT AND HCAT BRANCHES

Starting from the identification in Eq. (8), we construct two curved branches: PHCM-Prod retains the product and uses Eq. (10), whereas PHCML-HCat joins the factors on one hyperboloid and uses Eq. (13).

For PHCML, $g_\Psi = \Psi^* g_\mathbb{H}$ and $d\Psi_{I_d}$ identifies the unit tangent sphere with $S^{D-1} \subset T_{o_\mathbb{H}}\mathbb{H}^D_{-1}$. Let $\sigma = \sigma_{D-1}$ be uniform on this sphere. Since $\|\Xi(C)\|_2 = d_\mathbb{H}(o_\mathbb{H}, \Psi(C))$, define

$$\mathcal{P}_p^\Psi := \left\{ \mu \mid \int \|\Xi(C)\|_2^p \mathrm{d}\mu(C) < \infty \right\}. \tag{20}$$

For $p \geq 1$, $\mu, \nu \in \mathcal{P}_p^\Psi$, and $r \in \{\mathrm{Log}, \mathrm{G}, \mathrm{B}\}$, set $\mu_u^r = (s_u^r)_\# \mu$, $\nu_u^r = (s_u^r)_\# \nu$, and $\mathcal{C}_r = \mathrm{CorSW}_{p,\text{PHCML-}r}(\mu, \nu)$. Unlike the flat case, the three coordinates in Eq. (13) give distinct objectives: These are the PHCML-Log, PHCML-G, and PHCML-B sliced objectives defined by Eq. (16); their explicit integrals appear in Supplementary Eq. 139. The PHCM-Prod objectives integrate the one-dimensional Wasserstein distances induced by $s_u^{\text{Prod-Log}}$, $s_u^{\text{Prod-B}}$, or the approximate $s_u^{\text{Prod-G}}$ over the same global direction law. The last is denoted Prod-G (approx.), not $\mathrm{CorSW}_{p,g\text{-G}}$ from Def. 4.1. Crucially, each s_u is aggregated into one scalar per sample before matching. For example, with two equally weighted factors, source coordinates $(0,1),(1,0)$ and target coordinates $(0,0),(1,1)$ have identical factor marginals and hence zero factor-wise loss, but aggregated scores $\{1/\sqrt{2}, 1/\sqrt{2}\}$ and $\{0, \sqrt{2}\}$ have $W_2^2 = 1/2$. Thus aggregation retains cross-factor dependence. Supplementary Appendix H.8 proves the general shared-matching expression.

Proposition 4.3 (PHCML properties). *Let $p \geq 1$ and $\mu, \nu \in \mathcal{P}_p^\Psi$. Write $M := \Psi_\# \mu$ and $N := \Psi_\# \nu$.*

1. *PHCML-Log is a finite metric on $\mathcal{P}_p^\Psi$, with the exact reduction $\mathcal{C}_{\mathrm{Log}} = \mathrm{SW}_p(\Xi_\# \mu, \Xi_\# \nu)$.*
2. *PHCML-G is a finite metric on $\mathcal{P}_p^\Psi$. It equals $\mathrm{GHSW}_p(M, N)$, satisfies $\mathcal{C}_\mathrm{G} \leq W_{p,\mathbb{H}}(M, N)$ (Bonet et al., 2025, Props. 17 and 22). For the Klein coordinate $\kappa(x_0, \bar{x}) = \bar{x}/x_0$, it further obeys*
$$\mathrm{SW}_p\left((\kappa \circ \Psi)_\# \mu, (\kappa \circ \Psi)_\# \nu\right) \leq \mathcal{C}_\mathrm{G}. \tag{21}$$
3. *PHCML-Busemann equals $\mathrm{HHSW}_p(M, N)$ and is a finite sliced pseudometric. It satisfies $\mathcal{C}_\mathrm{B} \leq W_{p,\mathbb{H}}(M, N)$ (Bonet et al., 2025, Props. 17 and 22). It becomes a metric only on measure classes for which the horospherical Radon transform is injective. this condition is not assumed for the experimental empirical laws.*

For the Klein bound, let $T = \kappa \circ \Psi$. Then

$$\langle T(C), u \rangle = \tanh\left(s_u^\mathrm{G}(C)\right), \quad W_p\left(\left(\langle T(\cdot), u \rangle\right)_\# \mu, \left(\langle T(\cdot), u \rangle\right)_\# \nu\right) \leq W_p\left((s_u^\mathrm{G})_\# \mu, (s_u^\mathrm{G})_\# \nu\right). \tag{22}$$

Since $\tanh$ is 1-Lipschitz, integration gives Eq. (21). Under $\mu, \nu \in \mathcal{P}_p^\Psi$, $\mathcal{C}_\mathrm{G} = 0$ therefore forces Euclidean SW between the Klein images to vanish. Euclidean SW is a metric and T is a diffeomorphism, so $\mu = \nu$. This is the metricity proof sketch. Supplementary Appendix H.9 gives the full measure-theoretic argument. For the product branch, Product-Log is a metric through its global tangent-coordinate reduction. Product-Busemann is only a finite pseudometric in general and becomes a metric under its product Radon-transform injectivity condition. Product-Geodesic (approx.) is used only as a finite sliced training discrepancy, with no intrinsic geodesic-slicing or metricity claim.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETTINGS

Protocol. We evaluate SPD classification on Radar (Brooks et al., 2019), FPHA (Garcia-Hernando et al., 2018), and HDM05 (Müller et al., 2007), EEG transfer on HINSS2021 (Hinss et al., 2021), BCIC-IV-2a (Brunner et al., 2008), and BNCI2015-001 (Faller et al., 2012), and node classification on Cora, PubMed (Sen et al., 2008; Yang et al., 2016), and Disease-NC (Chami et al., 2019). Within each setting, all methods share the backbone, splits, schedule, and seven seeds, while only the alignment geometry changes. We compare no alignment, SPDSW, RawCorr, generic Corr-SPD, and CorSW. Highlighted CorSW results significantly outperform paired SPDSW results. Full settings are in Supplementary Appendix E.

Setting	None	SPDSW	Mode	Flat CorSW				PHCM-P			PHCML		
				ECM	LECM	OLM	LSM	Log	B	G (approx.)	Log	B	G
R_D	90.85	91.96	C	**92.96**	92.40	92.26	91.57	88.88	88.93	88.69	90.45	89.62	89.23
			M	92.46	91.76	92.27	92.11	88.75	88.40	88.37	90.39	89.67	89.20
R_S	90.45	91.42	C	**92.86**	92.66	92.00	90.91	88.67	88.53	88.77	91.03	90.13	89.68
			M	92.35	92.37	91.98	91.17	88.72	88.05	88.77	90.31	90.02	89.73
F	68.99	61.29	C	66.28	69.02	**70.82**	69.78	53.53	55.41	53.88	70.15	70.40	70.40
			M	66.28	69.05	69.58	69.62	51.86	54.30	52.49	68.82	69.74	69.32
H_{30}	60.69	60.79	C	60.67	61.67	61.86	62.05	62.45	62.32	62.44	62.03	**63.03**	61.71
			M	60.17	61.84	60.44	60.77	62.44	62.34	62.42	60.63	61.69	60.85
H_{50}	61.06	61.48	C	60.35	61.73	62.24	62.61	62.68	62.63	62.68	62.45	**63.39**	62.22
			M	60.59	61.99	61.34	61.71	62.70	62.61	62.72	61.55	62.57	61.51
I_S	54.48	55.33	C	56.06	56.10	55.10	55.59	53.92	**56.99**	54.79	55.24	54.42	54.98
			M	56.11	56.14	55.00	55.48	54.32	56.98	54.88	55.55	54.60	54.82
I_U	49.68	50.83	C	51.68	51.06	51.16	51.50	53.08	50.49	51.84	52.42	51.65	50.63
			M	51.13	51.42	51.62	51.73	**53.75**	50.45	51.78	52.19	51.05	50.62
Cora	80.12	80.18	C	80.18	80.10	80.30	80.04	79.54	80.12	80.20	79.60	80.28	79.82
			M	79.88	80.20	**80.52**	80.12	79.52	79.80	79.62	80.36	79.84	80.04
PubMed	74.92	77.04	C	77.78	77.50	77.92	78.08	77.56	77.70	77.60	78.18	78.10	77.44
			M	78.12	77.64	77.48	**78.56**	77.46	77.70	78.12	77.44	78.48	77.50
Disease-NC	77.09	89.84	C	86.38	92.52	93.15	90.08	93.39	93.46	92.60	91.65	92.05	93.54
			M	86.61	93.31	**95.83**	93.62	93.62	95.35	91.02	95.67	93.54	93.86

Table 1: RDA accuracy (%). PHCM-P: Log/B/Prod-G (approx.). PHCML: Log/B/G. Top-three means are highlighted. All highlighted CorSW results significantly outperform SPDSW.
Notation: C/M = CorSW-only/Mix, R_D/R_S = Radar deep/shallow, F = FPHA, H_{30}/H_{50} = HDM05, I_S/I_U = HINSS2021 inter-session/inter-subject.

5.2 Gradient Flow

We transport identical particles toward fixed interior and near-boundary targets on $\mathrm{Cor}^+(3)$. Across seven paired seeds, all methods reduce the common ambient W_2 while remaining positive definite. Complete endpoint statistics and near-boundary trajectories are reported in Supplementary Appendix E.2.

5.3 Alignment Experiments

Motivation. SPDNet lacks distribution alignment (Huang & Van Gool, 2017), while TSMNet+SPDDSMBN normalizes batchwise Fréchet statistics rather than complete distributions (Kobler et al., 2022). CorSW adds full-distribution alignment without changing either backbone.

RDA pipeline. Let $\widehat{\mu}_\Sigma$, $\widehat{\mu}_C$ be the empirical laws of the SPD features and their correlations. We align each to a domain-independent reference $\rho_{\mathcal{M}} := (\mathrm{Exp}^{\mathcal{M}}_{\mathrm{Id}})_{\#}\mathcal{N}(0, I)$ through $\mathcal{A}_\Sigma = \mathrm{SPDSW}(\widehat{\mu}_\Sigma, \rho_\Sigma)$ and $\mathcal{A}_C = \mathrm{CorSW}(\widehat{\mu}_C, \rho_C)$, and train with $\mathcal{L} = \mathcal{L}_{\mathrm{task}} + \eta\mathcal{L}_{\mathrm{RDA}}$, where

$$\mathcal{L}_{\mathrm{RDA}} = \begin{cases} \mathcal{A}_C, & \text{CorSW-only,} \\ \mathcal{A}_\Sigma, & \text{covariance-only,} \\ \lambda_C\mathcal{A}_C + \lambda_\Sigma\mathcal{A}_\Sigma, & \text{Mix.} \end{cases} \tag{23}$$

The reference samples are detached, so gradients update only the feature extractor. RDA is placed after the last BiMap in SPDNet, before SPDDSMBN in TSMNet, and on the training-node representations in SPDGCN.

Table 1 shows that a fixed LECM or OLM geometry exceeds both references in nine of ten settings, while every setting has at least one CorSW variant above both. On SPDNet, ECM leads both Radar depths, OLM reaches 70.82 on FPHA, and PHCML-Busemann leads HDM05 at 63.03/63.39. On SPDGCN, OLM-Mix reaches 80.52 on Cora and 95.83 on Disease-NC, while LSM-Mix reaches 78.56 on PubMed.

Product vs. HCat. The two curved representations are complementary rather than uniformly ordered. Product-Busemann reaches 56.99 on HINSS2021 inter-session transfer and Product-Log reaches 53.75 inter-subject, whereas PHCML-Busemann leads HDM05 and PHCML-Log reaches 95.67 on

Disease-NC. These results distinguish retaining the PHCM product from joining its factors on one HCat hyperboloid.

5.4 Sliced Alignment for EEG Transfer

Covariance alignment retains absolute channel power, so session- and subject-dependent amplitude changes can dominate the transport objective. Correlation removes this marginal scale while preserving inter-channel dependence, so we align both structures jointly, each on its own manifold.

Pipeline. We follow the SPDSW domain-adaptation setup: a shared SPD backbone maps the source covariances toward the fixed target distribution, and a linear classifier fit on the aligned source features is applied to the target. For each trial the backbone yields an SPD covariance Σ and its correlation $C = \mathbb{D}(\Sigma)^{-1/2}\Sigma\mathbb{D}(\Sigma)^{-1/2}$. Covariance and correlation enter only through two alignment losses, an SPDSW term on $\log \Sigma$ and a CorSW term on $\phi(C)$, combined by learned nonnegative weights $\lambda_\Sigma + \lambda_C = 1$:

$$\mathcal{L}_{\text{align}} = \lambda_\Sigma \text{SPDSW}(\mu^s_\Sigma, \mu^t_\Sigma) + \lambda_C \text{CorSW}(\mu^s_C, \mu^t_C). \tag{24}$$

Each term stays on its own manifold and uses only unlabeled source and target covariances. *No align* omits alignment, ***SPDSW*** uses only the covariance term, and each CorSW variant adds its correlation term. Since $\phi(C)$ is parameter-free, the only additional capacity is two scalar fusion weights.

Learned fusion. OLM leads BCIC-IV-2a and PHCML-Log leads BNCI2015-001, both above covariance-only SPDSW. These results show that correlation and covariance alignment can complement each other.

5.5 Ablation Study

We isolate four design choices: alignment mode, correlation-specific geometry, fusion weight, and high-dimensional slicing cost.

Mode ablation. CorSW-only leads all five SPDNet settings in Table 1, while Mix leads HINSS2021 inter-session transfer and all three graph datasets. Thus covariance-scale alignment complements correlation alignment where useful.

Geometry after correlation normalization. Under the same RDA protocol, at least one CorSW geometry exceeds RawCorr and Corr-SPD in every reported setting. The matched fusion control replaces only CorSW with Corr-SPD. Full results and uncertainty estimates appear in Supplementary Table 21.

Method	BCIC-IV-2a	BNCI2015-001
No align	77.87±3.92	81.50±12.19
SPDSW	82.21±3.06	82.57±12.61
ECM	82.31±3.55	83.99±12.14
LECM	82.44±3.20	83.53±12.96
OLM	**82.69±2.81**	83.40±12.72
LSM	82.52±2.94	83.37±12.85
PHCM-Prod-Log	79.42±3.91	83.61±12.47
PHCM-Prod-B	79.28±4.60	82.34±13.36
PHCM-Prod-G (approx.)	79.09±4.23	81.34±13.35
PHCML-Log	82.49±3.40	**84.20±11.63**
PHCML-B	77.87±3.92	78.53±13.05
PHCML-G	81.98±3.48	83.70±12.71

Table 2: EEG learned-fusion accuracy.

Fusion weight. On FPHA, the correlation-dominant weights $(\lambda_{\text{LOGM}}, \lambda_{\text{CorSW}}) = (0.05, 0.95)$ reach 70.75% versus 66.43% for equal weights, matching the CorSW-only winner in Table 1. The correlation term, rather than the covariance term, supplies the useful signal in this setting.

Efficiency. The complete runtime benchmark, including the $d = 500$ setting, is reported in Supplementary Table 25 under Supplementary Appendix F. PHCML-Log has the lowest Step time, and ECM the lowest Full time.

6 Conclusion

CorSW brings sliced-Wasserstein alignment to full-rank correlation matrices through flat, PHCM-product, and PHCML-HCat representations. Product-Busemann retains cross-factor dependence, while HCat supports closed-form hyperbolic slices. Across SPD classification, EEG transfer, and graph learning, these representations show complementary strengths in aligning channel dependencies.

REFERENCES

Ilya Archakov and Peter Reinhard Hansen. A new parametrization of correlation matrices. *Econometrica*, 89(4):1699–1715, 2021.

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 115–122. Springer, 2005.

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

Ahmad Bdeir, Kristian Schwethelm, and Niels Landwehr. Fully hyperbolic convolutional neural networks for computer vision. In *International Conference on Learning Representations*, 2024.

Benjamin Blankertz, Klaus-Robert Müller, Dean J Krusienski, Gerwin Schalk, Jonathan R Wolpaw, Alois Schlögl, Gert Pfurtscheller, José del R Millán, Michael Schröder, and Niels Birbaumer. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):153–159, 2007.

Clément Bonet, Paul Berg, Nicolas Courty, François Septier, Lucas Drumetz, and Minh-Tan Pham. Spherical Sliced-Wasserstein. arXiv:2206.08780, 2023a.

Clément Bonet, Laetitia Chapel, Lucas Drumetz, and Nicolas Courty. Hyperbolic Sliced-Wasserstein via geodesic and horospherical projections. In *Proceedings of 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*, volume 221 of *Proceedings of Machine Learning Research*, pp. 334–370. PMLR, 28 Jul 2023b.

Clément Bonet, Benoît Malézieux, Alain Rakotomamonjy, Lucas Drumetz, Thomas Moreau, Matthieu Kowalski, and Nicolas Courty. Sliced-Wasserstein on symmetric positive definite matrices for M/EEG signals. In *ICML*, 2023c.

Clément Bonet, Lucas Drumetz, and Nicolas Courty. Sliced-Wasserstein distances and flows on Cartan-Hadamard manifolds. *Journal of Machine Learning Research*, 26(32):1–76, 2025. URL https://jmlr.org/papers/v26/24-0359.html.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

Martin R. Bridson and André Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319. Springer Science & Business Media, 2013.

Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for SPD neural networks. In *NeurIPS*, 2019.

Clemens Brunner, Robert Leeb, Gernot Müller-Putz, Alois Schlögl, and Gert Pfurtscheller. BCI competition 2008–graz data set a. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, 16(1-6):34, 2008.

Rudrasis Chakraborty, Jose Bouza, Jonathan Manton, and Baba C Vemuri. A statistical recurrent model on the manifold of symmetric positive definite matrices. In *NeurIPS*, 2018.

Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2019.

Ines Chami, Albert Gu, Dat Nguyen, and Christopher Ré. HoroPCA: Hyperbolic dimensionality reduction via horospherical projections. In *ICML*, 2021.

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

Justin P. David and Daya Gu. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *SIAM Journal on Matrix Analysis and Applications*, 40(4): 1571–1603, 2019.

Josef Faller, Carmen Vidaurre, Teodoro Solis-Escalante, Christa Neuper, and Reinhold Scherer. Autocalibration and recurrent adaptation: Towards a plug and play online ERD-BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(3):313–319, 2012. doi: 10.1109/TNSRE.2012.2189584.

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 *CVPR*, 2018.

Marcel F Hinss, Ludovic Darmet, Bertille Somon, Emilie Jahanpour, Fabien Lotte, Simon Ladouce, and Raphaelle N Roy. An eeg dataset for cross-session mental workload estimation: passive bci competition of the neuroergonomics conference 2021. Zenodo, 2021.

Chen Hu, Rui Wang, Jiale Zhou, Jingjun Yi, Shaocheng Jin, Yidong Song, and Yefeng Zheng. A sliced-Wasserstein framework on correlation matrices for EEG decoding. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2026. doi: 10.1145/3770855.3818864.

Zhiwu Huang and Luc Van Gool. A Riemannian network for SPD matrix learning. In *AAAI*, 2017.

Reinmar J Kobler, Jun-ichiro Hirayama, Qibin Zhao, and Motoaki Kawanabe. SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG. In *NeurIPS*, 2022.

Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo K. Rohde. Generalized Sliced Wasserstein distances. In *NeurIPS*, 2019.

John M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, 2 edition, 2018. doi: 10.1007/978-3-319-91755-9.

Meinard Müller, Tido Röder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. Documentation mocap database hdm05. Technical report, Universität Bonn, 2007.

Xuan Son Nguyen. The gyro-structure of some matrix manifolds. In *NeurIPS*, 2022.

Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *ICCV*, 2009.

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

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 435–446. Springer, 2011.

David Sabbagh, Pierre Ablin, Gaël Varoquaux, Alexandre Gramfort, and Denis A. Engemann. Predictive regression modeling with meg/eeg: from source power to signals and cognitive states. *NeuroImage*, 222:116893, 2020. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2020.116893. URL https://www.sciencedirect.com/science/article/pii/S1053811920303797.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

Yann Thanwerdas. *Riemannian and Stratified Geometries on Covariance and Correlation Matrices*. Phd thesis, Université Côte d'Azur, 2022.

Yann Thanwerdas and Xavier Pennec. Theoretically and computationally convenient geometries on full-rank correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 43(4):1851–1872, 2022.

Yann Thanwerdas and Xavier Pennec. The geometry of mixed-euclidean metrics on symmetric positive definite matrices. *Linear Algebra and its Applications*, 685:125–163, 2024a.

Yann Thanwerdas and Xavier Pennec. Permutation-invariant log-euclidean geometries on full-rank correlation matrices. *Linear Algebra and its Applications*, 681:54–82, 2024b.

Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer, 2009.

Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In *International Conference on Machine Learning*, 2016.

Riemannian Sliced-Wasserstein Distances on Correlation Matrices

Appendix

Appendix Contents

1 Introduction **1**

2 Preliminaries **2**

2.1 Correlation Geometries 2

2.2 Wasserstein and Sliced Distances 3

3 Slices on Correlation Matrices **3**

3.1 Manifold Slices 3

3.2 Flat Correlation Slices 4

3.3 PHCM Product Identification and Slices 4

3.4 PHCML-HCat Slices 5

4 Correlation Sliced-Wasserstein Distances **6**

4.1 Riemannian Sliced-Wasserstein 6

4.2 Flat CorSW 6

4.3 Product and HCat Branches 7

5 Experiments **7**

5.1 Experimental Settings 7

5.2 Gradient Flow 8

5.3 Alignment Experiments 8

5.4 Sliced Alignment for EEG Transfer 9

5.5 Ablation Study 9

6 Conclusion **9**

A Comprehensive Related Work **15**

B Notation **15**

B.1 Summary of Notation 15

B.2 List of Acronyms 16

C Full-Rank Correlation Geometries **17**

C.1 Pullback metrics 17

C.2 Symmetric matrix functions 17

C.3 Geometries of the correlation manifold 18

D Discussion: From Riemannian Sliced-Wasserstein to Correlation Sliced-Wasserstein 22

D.1 General Framework and SPDSW vs. logSW 22

D.2 Pullback Implementation via Tangent Space Differentials 23

E Experiment Setup and Complete Results 25

E.1 Compute and reporting protocol 25

E.2 Controlled gradient-flow simulation 26

E.3 Datasets and architectures 26

E.4 Complete SPDNet results 27

E.5 Complete HINSS2021 results 27

E.6 Complete SPDGCN results 28

E.7 Fusion and weight ablations 28

F Runtime Benchmark on High-Dimensional Matrices 28

F.1 Experimental Settings 28

F.2 Benchmarking Methodology 28

F.3 Results and Analysis 29

G Extension to Rank-Deficient Correlations 34

H Theoretical Derivations and Proofs 34

H.1 Derivation of the pullback projection specialization 34

H.2 Derivations of the ECM and LECM coordinates 34

H.3 Algebraic expansion of the inherited OLM coordinate 34

H.4 Algebraic expansion of the inherited LSM coordinate 36

H.5 Derivation of the flat distance-preservation corollary 37

H.6 Metricity of the flat CorSW family 38

H.7 Geodesic coordinate of the flat CorSW family 39

H.8 Product-Space Slicing on PHCM 39

H.9 Properties of the PHCML sliced distances 42

A Comprehensive Related Work

This section situates CorSW within the literature and complements the focused positioning in the main text.

Covariance and Correlation Matrices in Machine Learning. Symmetric positive definite (SPD) matrices, often in the form of covariance matrices, have established themselves as robust representations in diverse domains, including brain-computer interfaces (BCI) (Barachant et al., 2012; Sabbagh et al., 2020), pedestrian sequence analysis (Nguyen, 2022), and financial modeling (Archakov & Hansen, 2021). Traditional Euclidean operations on these matrices often suffer from the "swelling effect", leading to poor interpolation and ill-conditioned matrices. To overcome this, researchers have resorted to Riemannian geometries. Early methods (Pennec et al., 2006; Arsigny et al., 2005) introduced the Affine-Invariant (AIM) and Log-Euclidean (LEM) metrics to respect the manifold structure of SPD matrices. More recently, geometry-aware deep learning (Huang & Van Gool, 2017; Kobler et al., 2022) extended these operations to neural networks layer-by-layer.

When cross-domain variation is dominated by global amplitude scaling, full-rank correlation matrices ($\mathrm{Cor}^+(d)$) provide the scale-invariant representation of interest. Explicit Riemannian geometries parameterized for $\mathrm{Cor}^+(d)$ include quotient constructions (David & Gu, 2019) and pullback metrics from the full SPD manifold, such as the Euclidean Correlation Metric (ECM) and the Log-Euclidean Correlation Metric (LECM) (Thanwerdas, 2022; Thanwerdas & Pennec, 2024a). CorSW brings these scale-invariant geometries into a common optimal-transport and deep-learning interface.

Optimal Transport and Sliced-Wasserstein Distances. Optimal Transport (OT) (Peyré & Cuturi, 2019) provides a geometrically intuitive way to compare probability distributions. Standard OT, particularly the p-Wasserstein distance, often imposes a heavy computational burden ($O(n^3 \log n)$). To mitigate this, the Sliced-Wasserstein (SW) distance (Rabin et al., 2011; Bonneel et al., 2015) was proposed, which projects high-dimensional distributions onto one-dimensional Euclidean lines where the optimal transport problem has an efficient $O(n \log n)$ closed-form solution via sorting.

SW has been generalized to spheres, hyperbolic spaces, and SPD manifolds (Bonet et al., 2023a;b;c). More generally, Bonet et al. (2025) established geodesic and horospherical SW on Cartan–Hadamard manifolds, including the pullback-Euclidean theory underlying our flat branch. We formulate ECM, LECM, OLM, and LSM as a unified flat CorSW family, derive Product-Busemann slicing on the original PHCM product space, and introduce PHCML with tangent-logarithmic, horospherical, and geodesic routes on one HCat hyperboloid. Independent concurrent work by Hu et al. (2026) uses the OLM- and LSM-based instances as regularization losses in a specific EEG decoding architecture and domain-shift protocol. Beyond this concurrent setting, we study correlation–covariance fusion and regularization across multiple SPD learning architectures.

B Notation

B.1 Summary of Notation

Table 3 summarizes all the notation used in this paper for better clarity.

Table 3: Summary of notation. All numbers and operators are assumed to be real.

Notation	Explanation
$(\mathcal{M}, g)$ or $\mathcal{M}$	Riemannian manifold
$T_P\mathcal{M}$	Tangent space at $P \in \mathcal{M}$
$\{O_k\}_{k=1}^m$	Orthonormal basis over the m-dimensional $T_E\mathcal{M}$
$g_P(\cdot,\cdot)$ or $\langle\cdot,\cdot\rangle_P$	Riemannian metric at $P \in \mathcal{M}$
$\|\cdot\|_P$	Norm induced by $\langle\cdot,\cdot\rangle_P$ on $T_P\mathcal{M}$
Log_P	Riemannian logarithmic map at P
Exp_P	Riemannian exponential map at P
$\gamma(t, P, Q)$	Geodesic connecting $P, Q \in \mathcal{M}$

Continued on next page

Table 3 – continued from previous page

Notation	Explanation
$\Gamma_{P\to Q}$	Riemannian parallel transportation along the geodesic connecting P and Q
$f_{*,P}$	Differential map of the smooth map f at P
H_x	Horosphere (affine hyperplane) through point x
$\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n\times n}$	Euclidean spaces of real scalars, n-dimensional vectors, and $n \times n$ matrices
$\mathrm{Diag}(n)$	Euclidean space of $n \times n$ diagonal matrices
$\mathrm{Diag}^+(n)$	Manifold of $n \times n$ positive diagonal matrices
S_n	Euclidean space of $n \times n$ symmetric matrices
$\mathrm{Hol}(n)$	Euclidean space of $n \times n$ symmetric matrices with null diagonals
$\mathrm{Row}_0(n)$	Euclidean space of $n \times n$ symmetric matrices with null row sum
S_n^{++}	Manifold of $n \times n$ SPD matrices
$\mathrm{Row}_+^1(n)$	Manifold of $n \times n$ SPD matrices with unit row sum
$\mathrm{Cor}^+(n)$	Manifold of $n \times n$ full rank correlation matrices
$\mathrm{LT}(n)$	Euclidean space of $n \times n$ lower triangular matrices
$\mathrm{LT}_1(n)$	Euclidean space of $n \times n$ lower triangular matrices with unit diagonals
$\mathrm{LT}_0(n)$	Euclidean space of $n \times n$ lower triangular matrices with null diagonals
$\mathrm{LT}^+(n)$	Cholesky manifold of $n \times n$ lower triangular matrices with positive diagonals
$\mathcal{L}_n$	Manifold of $n \times n$ lower triangular matrices with positive diagonals and unit row norm
$\langle\cdot,\cdot\rangle$, $\|\cdot\|$	Canonical Euclidean inner product and norm
$\langle\cdot,\cdot\rangle^{(\alpha,\beta,\gamma)}$	Permutation-invariant inner product over $\mathrm{Hol}(n)$
$\langle\cdot,\cdot\rangle^{(\alpha,\delta,\zeta)}$	Permutation-invariant inner product over $\mathrm{Row}_0(n)$
log, exp, Chol	Matrix logarithm, exponentiation, Cholesky decomposition
$\lfloor\cdot\rfloor$	Returns the strictly lower triangular matrix of a square matrix
$\varphi_{EC}(\cdot)$	$\varphi_{EC}(C) = \lfloor\Theta(C)\rfloor$, the isometry w.r.t. ECM
$D(\cdot)$	Returns a diagonal matrix with diagonals from a square matrix
$\mathrm{diag}(\cdot)$	Returns a diagonal matrix from an input vector
$D_v(\cdot)$	Returns a vector of diagonal elements from a square matrix
$(\cdot)_1$	$(\cdot)_1(S) = \lfloor S\rfloor + \frac{1}{2}D(S)$ for any square matrix S
$\odot$	Hadamard product
Cor	$\mathrm{Cor} : \Sigma \in S_n^{++} \mapsto D(\Sigma)^{-1/2}\Sigma D(\Sigma)^{-1/2} \in \mathrm{Cor}^+(n)$
Θ	$\Theta : C \in \mathrm{Cor}^+(n) \mapsto D(\mathrm{Chol}(C))^{-1}\,\mathrm{Chol}(C) \in \mathrm{LT}_1(n)$
off	Returns a matrix in $\mathrm{Hol}(n)$ consisting of off-diagonal elements
$\mathrm{Log}_\circ$, $\mathrm{Exp}_\circ$	Off-log and its inverse
$\mathrm{Log}_\star$, $\mathrm{Exp}_\star$	Log-scaled and its inverse
$r_i(C)$, $\ell_i(C)$	Normalized Cholesky row coordinate and its row-wise Lorentz lift
HCat	Lorentz direct concatenation of multiple hyperbolic factors
$\Psi(C)$	Joint PHCML representation obtained by applying HCat to the row factors
I or I_n, 0	Identity matrix and Zero matrix or vector
$\mathbf{1}$	Vector with all 1 entities

B.2 List of Acronyms

AISO Affine-Invariant Sinkhorn Optimal Transport
BCI Brain-Computer Interface
CorSW Correlation Matrix Sliced-Wasserstein Distances
ECM Euclidean-Cholesky Metric
EEG Electroencephalography
HCat Lorentz Direct Concatenation

LCM Log-Cholesky Metric
LECM Log-Euclidean-Cholesky Metric
LEM Log-Euclidean Metric
LEW Log-Euclidean Wasserstein
LES Log-Euclidean Sinkhorn
LSM Log-Scaled Metric
OLM Off-Log Metric
PHCM Poly-Hyperbolic Cholesky Metric
PHCML PHCM with a Joint Lorentz Representation
SPD Symmetric Positive Definite
SW Sliced-Wasserstein

C Full-Rank Correlation Geometries

This section follows the notation in Tab. 3 and separates the models used by our method by curvature. ECM, LECM, OLM, and LSM are pullbacks of Euclidean model spaces and therefore have zero curvature Thanwerdas & Pennec (2022; 2024b). The original PHCM uses a product of hyperbolic factors and is generally non-positively curved, including zero-curvature mixed planes. We derive Product-Log and the exact Product-Busemann coordinate on this canonical geometry. PHCML instead maps the same Cholesky row factors to Lorentz coordinates and combines them through HCat into one Lorentz hyperboloid of constant negative curvature. On PHCML, Log applies the Lorentz logarithmic map followed by tangent-space slicing, while Busemann and G apply intrinsic horospherical and geodesic coordinates directly.

C.1 Pullback metrics

Both regimes are constructed through maps to model spaces, so we first recall the general pullback definition. The target metric is Euclidean for the flat branch and Lorentz-hyperbolic for PHCML, and this distinction determines whether the two projection coordinates coincide.

Definition C.1 (Pullback Metrics Lee (2018))**.** Suppose $\mathcal{M}_1, \mathcal{M}_2$ are smooth manifolds, g is a Riemannian metric on $\mathcal{M}_2$, and $f : \mathcal{M}_1 \to \mathcal{M}_2$ is a diffeomorphism. Then the pullback of g by f is defined point-wisely,

$$(f^*g)_p(V, W) = g_{f(p)}(f_{*,p}(V), f_{*,p}(W)), \tag{25}$$

where $f_{*,p}(\cdot)$ is the differential map of f at $p \in \mathcal{M}_1$, and $V, W \in T_p\mathcal{M}_1$. f^*g is a Riemannian metric on $\mathcal{M}_1$, called the pullback metric of g by f. Here, f is also called a Riemannian isometry.

Although pullback metrics can also be defined by smooth maps Lee (2018), this paper focuses on diffeomorphisms.

C.2 Symmetric matrix functions

We denote $\mathcal{S}^n$ as the Euclidean space of $n \times n$ real symmetric matrices, and $\mathcal{S}^n_{++}$ as the SPD manifold of $n \times n$ SPD matrices. Let I be an open interval of $\mathbb{R}$ and $f : I \to \mathbb{R}$ be a smooth function. The smooth map induced by f for any symmetric matrix S with all eigenvalues in I is defined as

$$f : S \longmapsto U f(\Sigma) U^\top \in \mathcal{S}^n, \text{ with } S = U\Sigma U^\top \text{ as the eigendecomposition.} \tag{26}$$

Its differential is known as the Daleckii-Krein formula:

$$f_{*,S}(V) = U\left(L \odot \left(U^\top V U\right)\right) U^\top, \quad \forall V \in \mathcal{S}^n, \tag{27}$$

$$L_{i,j} = \begin{cases} \frac{f(\sigma_i) - f(\sigma_j)}{\sigma_i - \sigma_j}, & \text{if } \sigma_i \neq \sigma_j \\ f'(\sigma_i), & \text{otherwise} \end{cases} \tag{28}$$

where L is called the Loewner matrix with the (i, j)-th element defined in Eq. (28), and $\odot$ denotes the Hadamard product. Two special cases are the matrix logarithm: $\log : \mathcal{S}^n_{++} \to \mathcal{S}^n$ and its inverse, the matrix exponentiation $\exp : \mathcal{S}^n \to \mathcal{S}^n_{++}$.

Table 4: Zero-curvature pullback-Euclidean metrics used by the flat CorSW branch.

Metric	Prototype space	Diffeomorphisms	Properties
ECM Thanwerdas & Pennec (2022)	$\mathrm{LT}^1(n) = \mathrm{LT}^0(n) + I_n$	$\Theta : C \longmapsto \mathbb{D}(\mathrm{Chol}(C))^{-1}\mathrm{Chol}(C)$ $\Theta^{-1} = \mathrm{Cor} \circ \mathrm{Chol}^{-1}$	Null curvature
LECM Thanwerdas & Pennec (2022)	$\mathrm{LT}^0(n)$	$\log \circ \Theta : \mathrm{Cor}^+(n) \longmapsto \mathrm{LT}^0(n)$ $(\log \circ \Theta)^{-1} = \mathrm{Cor} \circ \mathrm{Chol}^{-1} \circ \exp$	Null curvature
OLM Thanwerdas & Pennec (2024b)	$\mathrm{Hol}(n)$	$\mathrm{Log}_\circ : C \longmapsto \mathrm{off}(\log(C))$ $(\mathrm{Log}_\circ)^{-1} = \exp \circ \mathbb{D}$	Permutation-inv. Null curvature
LSM Thanwerdas & Pennec (2024b)	$\mathrm{Row}_0(n)$	$\mathrm{Log}_\star : C \longmapsto \mathrm{row}_0(\log(C))$ $(\mathrm{Log}_\star)^{-1} = \exp \circ \mathbb{R}$	Permutation-inv. Null curvature

C.3 Geometries of the correlation manifold

Following Sec. 2, we use two geometric regimes: (i) the zero-curvature pullback-Euclidean metrics ECM, LECM, OLM, and LSM, and (ii) the constant-negative-curvature PHCML representation obtained by applying HCat to row-wise PHCM Lorentz factors. Permutation invariance provides a second, independent distinction. OLM and LSM are permutation-invariant, whereas the Cholesky-based constructions are order dependent.

Zero-curvature Cholesky pullback metrics

The zero-curvature Cholesky metrics ECM and LECM Thanwerdas & Pennec (2022) are defined by pullback to Euclidean triangular-matrix spaces:

$$\text{ECM:}\quad \mathrm{Cor}^+(n) \overset{\Theta=\mathbb{D}^{-1}(\mathrm{Chol}(\cdot))\mathrm{Chol}(\cdot)}{\longrightarrow} \mathrm{LT}^1(n) = I_n + \mathrm{LT}^0(n), \tag{29}$$

$$\text{LECM:}\quad \mathrm{Cor}^+(n) \overset{\log\circ\Theta}{\longrightarrow} \mathrm{LT}^0(n), \tag{30}$$

The following formulas review their associated maps and differentials.

ECM and LECM. For any $C \in \mathrm{Cor}^+(n)$, $V \in T_C\mathrm{Cor}^+(n) \cong \mathrm{Hol}(n)$, $K \in \mathrm{LT}^1(n)$ and $X, \xi \in \mathrm{LT}^0(n)$, the involved maps and their differentials in ECM and LECM are

$$\Theta(C) = \mathbb{D}(L)^{-1}L \tag{31}$$

$$\Theta^{-1}(K) = \mathbb{D}\left(KK^\top\right)^{-\frac{1}{2}} KK^\top \mathbb{D}\left(KK^\top\right)^{-\frac{1}{2}} \tag{32}$$

$$\log(K) = \sum_{k=1}^{n-1} \frac{(-1)^{k-1}}{k}(K - I_n)^k \tag{33}$$

$$\exp(\xi) = \sum_{k=0}^{n-1} \frac{1}{k!}\xi^k \tag{34}$$

$$\Theta_{*,C}(V) = \Theta(C)\left(L^{-1}VL^{-\top}\right)_{\frac{1}{2}} - \frac{1}{2}\mathbb{D}\left(L^{-1}VL^{-\top}\right)\Theta(C) \tag{35}$$

$$(\Theta_{*,C})^{-1}(\xi) = \left(L\xi^\top - C\mathbb{D}\left(L\xi^\top\right)\right)\mathbb{D}(L) + \mathbb{D}(L)\left(\xi L^\top - \mathbb{D}\left(L\xi^\top\right)C\right) \tag{36}$$

$$\log_{*,K}(\xi) = \sum_{k=1}^{n-1} \frac{(-1)^{k-1}}{k}\left[(K - I_n)^{k-1}\xi + \cdots + \xi(K - I_n)^{k-1}\right] \tag{37}$$

$$\exp_{*,X}(\xi) = \sum_{k=1}^{n-1} \frac{1}{k!}\left(X^{k-1}\xi + X^{k-2}\xi X + \cdots + \xi X^{k-1}\right) \tag{38}$$

$$(\log\circ\Theta)_{*,C}(V) = \log_{*,\Theta(C)}(\Theta_{*,C}(V)) \tag{39}$$

$$\mathrm{Chol}_{*,C}(V) = L\left(L^{-1}VL^{-\top}\right)_{\frac{1}{2}} \tag{40}$$

$$(\mathrm{Chol}_{*,C})^{-1}(Z) = LZ^\top + ZL^\top, \quad \forall Z \in T_L\mathrm{LT}^+(n) \cong \mathrm{LT}(n) \tag{41}$$

Table 5: Riemannian operators under the non-permutation-invariant log-Euclidean Metrics. Here, $C, C' \in \mathrm{Cor}^+(n)$ are correlation matrices and $V, W \in T_C\mathrm{Cor}^+(n) \cong \mathrm{Hol}(n)$ are tangent vectors. Although the inner product $\langle \cdot, \cdot \rangle$ could be any Euclidean inner product, this paper focuses on the canonical one.

Operation	ECM	LECM
$g_C(V, W)$	$\langle \Theta_{*,C}(V), \Theta_{*,C}(W) \rangle$	$\langle (\log \circ \Theta)_{*,C}(V), (\log \circ \Theta)_{*,C}(W) \rangle$
$\mathrm{Exp}_C(V)$	$\Theta^{-1}(\Theta(C) + \Theta_{*,C}(V))$	$(\log \circ \Theta)^{-1}(\log \circ \Theta(C) + (\log \circ \Theta)_{*,C}(V))$
$\mathrm{Log}_C(C')$	$\Theta^{-1}_{*,\Theta(C)}(\Theta(C') - \Theta(C))$	$(\log \circ \Theta)^{-1}_{*,\log \circ \Theta(C)}(\log \circ \Theta(C') - \log \circ \Theta(C))$
$\gamma(t, C, C')$	$\Theta^{-1}((1-t)\Theta(C) + t\Theta(C'))$	$(\log \circ \Theta)^{-1}((1-t)\log \circ \Theta(C) + t \log \circ \Theta(C'))$
$d(C, C')$	$\lVert \Theta(C) - \Theta(C') \rVert$	$\lVert \log \circ \Theta(C) - \log \circ \Theta(C') \rVert$
Fréchet mean	$\Theta^{-1}\left(\frac{1}{k}\sum_{i=1}^k \Theta(C_i)\right)$	$(\log \circ \Theta)^{-1}\left(\frac{1}{k}\sum_{i=1}^k \log \circ \Theta(C_i)\right)$
Curvature	0	0
$\Gamma_{C \to C'}(V)$	$(\Theta_{*,C'})^{-1}(\Theta_{*,C}(V))$	$((\log \circ \Theta)_{*,C'})^{-1}((\log \circ \Theta)_{*,C}(V))$

where L is the Cholesky factor of C and I_n is the $n \times n$ identity matrix. Due to the nilpotency of $\mathrm{LT}^0(n)$, the matrix logarithm over $\mathrm{LT}^1(n)$ and exponentiation over $\mathrm{LT}^0(n)$ are free from eigendecomposition. With the above equations, Table 5 summarizes the Riemannian operators under ECM and LECM.

Negative-curvature PHCML Lorentz lift

Let $C = LL^\top$ and define the PHCM row coordinates $r_i = L_{i,1:i-1}/L_{ii}$ and Lorentz factors $\ell_i = (\sqrt{1 + \lVert r_i \rVert_2^2}, r_i)$ as in Eq. (9). For $N = d - 1$ points $q_i = (q_{i,0}, \bar{q}_i) \in \mathbb{H}^{m_i}_{-1}$, Lorentz direct concatenation Bdeir et al. (2024) is

$$\mathrm{HCat}(q_1, \ldots, q_N) = \left(\sqrt{\sum_{i=1}^N q_{i,0}^2 - (N-1)}, \mathrm{Concat}(\bar{q}_1, \ldots, \bar{q}_N) \right). \tag{42}$$

Indeed, if $y = \mathrm{HCat}(q_1, \ldots, q_N)$, then

$$\begin{aligned} \langle y, y \rangle_L &= -\sum_{i=1}^N q_{i,0}^2 + (N-1) + \sum_{i=1}^N \lVert \bar{q}_i \rVert_2^2 \\ &= \sum_{i=1}^N \langle q_i, q_i \rangle_L + (N-1) = -1, \end{aligned} \tag{43}$$

so y lies on one Lorentz hyperboloid. Applying Eq. (42) to $\{\ell_i\}_{i=2}^d$ yields

$$\begin{aligned} \Psi(C) &= \left(\sqrt{\sum_{i=2}^d (1 + \lVert r_i \rVert_2^2) - (d-2)}, \mathrm{Concat}(r_2, \ldots, r_d) \right) \\ &= \left(\sqrt{1 + \lVert z(C) \rVert_2^2}, z(C) \right), \end{aligned} \tag{44}$$

which proves the closed form in Eq. (12). Hence PHCML preserves the row-wise PHCM construction in its input pipeline but replaces the product target by a joint constant-curvature Lorentz representation. It is therefore not an isometric coordinate realization of the original PHCM product metric: the latter has zero-curvature mixed planes, whereas every tangent two-plane of $\mathbb{H}^D_{-1}$ has curvature -1. The joint representation then supports the three slice-coordinate routes in Sec. 3.4.

Permutation-invariant metrics

Let $\mathcal{S}^n$ be the group of permutation matrices $P_\sigma = [\delta_{i,\sigma(j)}]_{1 \le i,j \le n}$ by permutation σ, and $\mathcal{D}^{\pm}(n) = \{\mathrm{diag}((\varepsilon_1, \ldots, \varepsilon_n)), \varepsilon \in \{-1, 1\}^n\}$ be the group of diagonal matrices with coefficients in $\{-1, 1\}$.

Thanwerdas Thanwerdas & Pennec (2024b) (Prop. 1.1) showed that the largest congruence action on full-rank correlation matrices is the action of signed permutation matrices:

$$\star : (A, C) \in \widehat{\mathcal{S}}^{\pm}(n) \times \mathrm{Cor}^{+}(n) \longmapsto ACA^{\top} \in \mathrm{Cor}^{+}(n) \tag{45}$$

with $\widehat{\mathcal{S}}^{\pm}(n) = \mathcal{D}^{\pm}(n)\mathcal{S}^{n}$. Based on this finding, Thanwerdas Thanwerdas & Pennec (2024b) proposed two permutation-invariant metrics, namely OLM and LSM, by pulling back permutation-invariant inner products via the following permutation-equivariant diffeomorphisms:

$$\mathrm{Cor}^{+}(n) \underset{\mathrm{Exp}^{\circ}}{\overset{\mathrm{Log}^{\circ}=\mathrm{off}\circ\log}{\longleftrightarrow}} \mathrm{Hol}(n) \tag{46}$$

$$\mathrm{Cor}^{+}(n) \underset{\mathrm{Exp}^{\star}=\mathrm{Cor}\circ\exp}{\overset{\mathrm{Log}^{\star}}{\longleftrightarrow}} \mathrm{Row}_0(n) \tag{47}$$

$$\mathrm{Exp}^{\circ} : \mathrm{Hol}(n) \ni H \longmapsto \exp(\mathcal{D}(H) + H) \tag{48}$$

$$\mathrm{Log}^{\star} : \mathrm{Cor}^{+}(n) \ni C \longmapsto \log(\mathcal{D}^{\star}(C)C\mathcal{D}^{\star}(C)) \in \mathrm{Row}_0(n) \tag{49}$$

where $\log(\cdot)$ and $\exp(\cdot)$ are symmetric matrix logarithm and exponentiation. The involved $\mathcal{D}$ and $\mathcal{D}^{\star}$ can be formally expressed as $\mathcal{D} : \mathrm{Hol}(n) \to \mathrm{Diag}(n)$ and $\mathcal{D}^{\star} : \mathrm{Cor}^{+}(n) \to \mathrm{Diag}^{+}(n)$, where $\mathrm{Diag}(n)$ denotes the Euclidean space of $n \times n$ diagonal matrices, and $\mathrm{Diag}^{+}(n)$ is a submanifold of $\mathrm{Diag}(n)$, consisting of positive diagonal matrices.

As discussed in Section 2.1 of the main text, for any symmetric hollow matrix $H \in \mathrm{Hol}(d)$, the operator $\mathcal{D}(H)$ returns a unique diagonal matrix, which can be computed by the exponentially converging algorithm (Archakov & Hansen, 2021, Sec. 5): $D_{k+1} = D_k - \log(\mathbb{D}(\exp(D_k + H)))$, with $D_0 = \mathbf{0}$. For LSM, $\mathcal{D}^{*}(C)$ corresponds to the unique zero of $f : x \in \mathbb{R}_{+}^{d} \mapsto Cx - \frac{1}{x}$, which could be solved by damped Newton's method (Thanwerdas & Pennec, 2024b, Sec. 3.5).

The differentials of Log° and $\mathrm{Log}^{\star}$ and their inverses can be calculated by the differential of symmetric matrix logarithm and exponentiation Thanwerdas & Pennec (2024b) (Props. 2.4 and 4.1). Given $C \in \mathrm{Cor}^{+}(n)$, tangent vector $V \in T_C\mathrm{Cor}^{+}(n) \cong \mathrm{Hol}(n)$, $H, W \in \mathrm{Hol}(n)$, and $S = H + \mathcal{D}(H) = U\Lambda U^{\top}$, the differential of Log° and its inverse Exp° are

$$\mathrm{Log}^{\circ}_{\cdot,C}(V) = \mathrm{off}\left(\log_{\cdot,C}(V)\right) \tag{50}$$

$$\mathrm{Exp}^{\circ}_{\cdot,H}(W) = \exp_{\cdot,S}\left(W + \mathcal{D}_{\cdot,H}(W)\right) \tag{51}$$

$$\mathcal{D}_{\cdot,H}(W) = -\mathrm{diag}\left((H^0)^{-1}\mathbb{D}\left(\exp_{\cdot,S}(W)\right)\mathbf{1}\right) \tag{52}$$

$$\mathcal{S}^{n}_{++} \ni H^0_{ij} = \sum_{j,k} P_{ij}P_{ik}P_{tj}P_{tk}L_{j,k} \tag{53}$$

where L is the Loewner matrix of $\exp_{\cdot,S}$, and $\mathbf{1}$ is the vector of all 1 entities. Here, $\log_{\cdot}$ and $\exp_{\cdot}$ can be calculated using the Daleckii-Krein formula of the symmetric matrix, while $\mathrm{diag}(\cdot) : \mathbb{R}^n \to \mathrm{Diag}(n)$ returns a diagonal matrix from an input vector. Further denoting $X, Y \in \mathrm{Row}_0(n)$ and $\Sigma = \mathcal{D}^{\star}(C)C\mathcal{D}^{\star}(C)$, the differentials of $\mathrm{Log}^{\star}$ and its inverse $\mathrm{Exp}^{\star}$ are

$$\mathrm{Log}^{\star}_{\cdot,C}(V) = \log_{\cdot,\Sigma}\left(\Delta V\Delta + \frac{1}{2}\left(V^0\Sigma + \Sigma V^0\right)\right) \tag{54}$$

$$\mathrm{Exp}^{\star}_{\cdot,X}(Y) = \Delta^{-1}\left[\exp_{\cdot,X}(Y) - \frac{1}{2}\left(\Delta^{-2}\mathbb{D}\left(\exp_{\cdot,X}(Y)\right)\Sigma + \Sigma\mathbb{D}\left(\exp_{\cdot,X}(Y)\right)\Delta^{-2}\right)\right]\Delta^{-1} \tag{55}$$

with $\Delta = \mathbb{D}(\Sigma)^{1/2}$ and $V^0 = -2\mathrm{diag}\left((I_n + \Sigma)^{-1}\Delta V\Delta\mathbf{1}\right)$.

As both $\mathrm{Log}^{\star}_{\cdot}$ and $\mathrm{Log}^{\circ}_{\cdot}$ are permutation-equivariant Thanwerdas & Pennec (2024b), permutation-invariant metrics for the correlation manifold can be induced by permutation-invariant inner products on $\mathrm{Hol}(n)$ and $\mathrm{Row}_0(n)$, respectively. The following two propositions review such inner products.

Table 6: Riemannian geometries under the permutation-invariant log-Euclidean Metrics.

Operation	OLM	LSM
$g_C(V,W)$	$\langle \mathrm{Log}^{\circ}_{\cdot,C}(V), \mathrm{Log}^{\circ}_{\cdot,C}(W)\rangle^{(\alpha,\beta,\gamma)}$	$\langle \mathrm{Log}^{\star}_{\cdot,C}(V), \mathrm{Log}^{\star}_{\cdot,C}(W)\rangle^{(\alpha,\delta,\zeta)}$
$\mathrm{Exp}_C(V)$	$\mathrm{Exp}^{\circ}(\mathrm{Log}^{\circ}(C) + \mathrm{Log}^{\circ}_{\cdot,C}(V))$	$\mathrm{Exp}^{\star}(\mathrm{Log}^{\star}(C) + \mathrm{Log}^{\star}_{\cdot,C}(V))$
$\mathrm{Log}_C(C')$	$\mathrm{Exp}^{\circ}_{\cdot,\mathrm{Log}^{\circ}(C)}(\mathrm{Log}^{\circ}(C') - \mathrm{Log}^{\circ}(C))$	$\mathrm{Exp}^{\star}_{\cdot,\mathrm{Log}^{\star}(C)}(\mathrm{Log}^{\star}(C') - \mathrm{Log}^{\star}(C))$
$\gamma(t,C,C')$	$\mathrm{Exp}^{\circ}((1-t)\mathrm{Log}^{\circ}(C) + t\mathrm{Log}^{\circ}(C'))$	$\mathrm{Exp}^{\star}((1-t)\mathrm{Log}^{\star}(C) + t\mathrm{Log}^{\star}(C'))$
$d(C,C')$	$\|\mathrm{Log}^{\circ}(C) - \mathrm{Log}^{\circ}(C')\|^{(\alpha,\beta,\gamma)}$	$\|\mathrm{Log}^{\star}(C) - \mathrm{Log}^{\star}(C')\|^{(\alpha,\delta,\zeta)}$
Fréchet mean	$\mathrm{Exp}^{\circ}\left(\frac{1}{k}\sum_{i=1}^{k}\mathrm{Log}^{\circ}(C_i)\right)$	$\mathrm{Exp}^{\star}\left(\frac{1}{k}\sum_{i=1}^{k}\mathrm{Log}^{\star}(C_i)\right)$
Curvature	0	0
$\Gamma_{C\to C'}(V)$	$(\mathrm{Log}^{\circ}_{\cdot,C'})^{-1}(\mathrm{Log}^{\circ}_{\cdot,C}(V))$	$(\mathrm{Log}^{\star}_{\cdot,C'})^{-1}(\mathrm{Log}^{\star}_{\cdot,C}(V))$
Properties	Permutation-invariance, Null curvature Signed-perm.-inv. ($\beta=\gamma=0$ for OLM) Inverse-consistency (LSM only)	

Proposition C.2 (Permutation-invariant inner products on $\mathrm{Hol}(n)$ Thanwerdas & Pennec (2024b))**.** *Supposing $n \geq 4$, permutation-invariant inner products on Hol(n) are:*

$$\langle X_1, X_2\rangle^{(\alpha,\beta,\gamma)} = \alpha tr(X_1X_2) + \beta Sum(X_1X_2) + \gamma Sum(X_1)Sum(X_2), \quad \forall X_1, X_2 \in Hol(n) \tag{56}$$

with $\alpha > 0$, $2\alpha + (n-2)\beta > 0$, and $\alpha + (n-1)(\beta + n\gamma) > 0$. For $n = 3$, permutation-invariant inner products have the same form with $\alpha = 0$:

$$\langle X_1, X_2\rangle^{(\alpha,\beta,\gamma)} = \beta Sum(X_1X_2) + \gamma Sum(X_1)Sum(X_2), \quad \text{with } \beta > 0 \text{ and } \beta + 3\gamma > 0 \tag{57}$$

For $n = 2$, they have the same form with $\alpha = \beta = 0$:

$$\langle X_1, X_2\rangle^{(\alpha,\beta,\gamma)} = \gamma Sum(X_1)Sum(X_2), \quad \text{with } \gamma > 0 \tag{58}$$

Proposition C.3 (Permutation-invariant inner products on $\mathrm{Row}_0(n)$ Thanwerdas & Pennec (2024b))**.** *For $n \geq 4$, permutation-invariant inner products on Row$_0(n)$ are*

$$\langle Y_1, Y_2\rangle^{(\alpha,\delta,\zeta)} = \alpha tr(Y_1Y_2) + \delta tr(\mathbb{D}(Y_1)\mathbb{D}(Y_2)) + \zeta tr(Y_1)tr(Y_2), \quad \forall Y_1, Y_2 \in Row_0(n) \tag{59}$$

with $\alpha > 0$, $n\alpha + (n-2)\delta > 0$, and $n\alpha + (n-1)(\delta + n\zeta) > 0$. For $n = 3$, the permutation-invariant inner products have the same form with $\alpha = 0$. For $n = 2$, they have the same form with $\alpha = \delta = 0$.

As shown by Thanwerdas Thanwerdas & Pennec (2024b), OLM is further invariant to signed-permutation under $\beta = \gamma = 0$, where the associated $\langle\cdot,\cdot\rangle^{(\alpha,0,0)}$ is reduced to the scaled canonical Euclidean inner product:

$$\langle V, W\rangle^{(\alpha,0,0)} = \alpha\langle V, W\rangle, \quad \forall V, W \in \mathrm{Hol}(n) \tag{60}$$

In the main paper, we assume that $\langle\cdot,\cdot\rangle^{(\alpha,\beta,\gamma)}$ and $\langle\cdot,\cdot\rangle^{(\alpha,\delta,\zeta)}$ are the canonical Euclidean inner product.

Lastly, we briefly review inverse-consistency, a property exclusive to LSM. The cor-inversion is defined as $\mathcal{I} : \mathrm{Cor}^{+}(n) \ni C \mapsto \mathrm{Cor}(C^{-1}) \in \mathrm{Cor}^{+}(n)$ Thanwerdas & Pennec (2024b) (Def. 1.4). It corresponds to the matrix inversion $\mathrm{inv} : \mathcal{S}^n_{++} \ni \Sigma \mapsto \Sigma^{-1} \in \mathcal{S}^n_{++}$, as represented on the following commuting diagram:

$$\begin{array}{ccc} \mathcal{S}^n_{++} & \xrightarrow{\text{inv}} & \mathcal{S}^n_{++} \\ {\scriptstyle \mathrm{Cor}}\downarrow & & \downarrow{\scriptstyle \mathrm{Cor}} \\ \mathrm{Cor}^{+}(n) & \xrightarrow[\mathcal{I}]{} & \mathrm{Cor}^{+}(n) \end{array} \tag{61}$$

As shown by Thanwerdas Thanwerdas & Pennec (2024b) (Prop. 1.7), LSM enjoys inverse-consistency:

$$\mathrm{Log}^{\star}(\mathcal{I}(C)) = -\mathrm{Log}^{\star}(C), \quad \forall C \in \mathrm{Cor}^{+}(n). \tag{62}$$

Table 6 summarizes the Riemannian structures of OLM and LSM.

Remark C.4. We make the following remarks w.r.t. OLM and LSM.

1. **Invariance and dimension:** Propositions C.2 and C.3 imply that when $n \leq 3$, the canonical inner products over $\mathrm{Hol}(n)$ and $\mathrm{Row}_0(n)$, as well as the induced OLM and LSM, are no longer invariant metrics. However, our main paper focuses on cases where $n > 3$.

2. $\mathcal{D}$ **and** $\mathcal{D}^\star$**:** $\mathcal{D}$ is also well-defined over $\mathcal{S}^n$, a surjective map $\mathcal{D} : \mathcal{S}^n \to \mathrm{Diag}(n)$. In this way, $\mathrm{Exp}^\circ : \mathcal{S}^n \to \mathrm{Cor}^+(n)$ is no longer bijective Thanwerdas & Pennec (2024b) (Prop. 2.1). Similarly, $\mathcal{D}^\star$ is well defined over $\mathcal{S}^n_{++}$, a surjective map $\mathcal{D}^\star : \mathcal{S}^n_{++} \to \mathrm{Diag}^+(n)$. Consequently, $\mathrm{Log}^\star : \mathcal{S}^n_{++} \to \mathrm{Row}_0(n)$ is no longer bijective Thanwerdas & Pennec (2024b) (Prop. 3.5).

D Discussion: From Riemannian Sliced-Wasserstein to Correlation Sliced-Wasserstein

We contextualize CorSW within Riemannian optimal transport and separate inherited Cartan–Hadamard theory from our correlation-specific instantiations. We first review the general construction and the SPDSW vs. logSW comparison, then record its pullback implementation via tangent-space differentials.

D.1 General Framework and SPDSW vs. logSW

Riemannian Sliced-Wasserstein Construction

For a Cartan–Hadamard manifold $\mathcal{M}$ with base point o, Bonet et al. (2025, Def. 5) define both geodesic and horospherical CHSW by integrating one-dimensional Wasserstein distances over the corresponding coordinates. Writing π_v for either choice,

$$\mathrm{RSW}_p^p(\mu, \nu) = \int_{S_o} W_p^p(\pi_v \sharp \mu, \pi_v \sharp \nu) d\lambda_o(v), \tag{63}$$

where $S_o = \{v \in T_o\mathcal{M} : \|v\|_o = 1\}$ is the unit tangent sphere, $\pi_v : \mathcal{M} \to \mathbb{R}$ denotes either the signed geodesic coordinate P^v or the Busemann coordinate B^v, and λ_o is the uniform measure on S_o.

SPDSW vs. logSW on SPD Manifolds

For SPD^d with the Log-Euclidean metric, two approaches exist Bonet et al. (2023c):

SPDSW. Samples unit-norm symmetric matrices $A \in \mathcal{S}^d$ via QR decomposition, projecting as $t_A(M) = \mathrm{tr}(A \log M)$. This respects SPD geometry but costs $O(d^3)$ per projection.

logSW. Maps via $\log : \mathrm{SPD}^d \to \mathcal{S}^d$, then applies Euclidean SW with spherical sampling in $\mathbb{R}^{d \times d}$:

$$\mathrm{logSW}_p^p(\mu, \nu) = \mathrm{SW}_p^p(\log_\sharp \mu, \log_\sharp \nu). \tag{64}$$

Motivation. The SPDSW/logSW comparison of Bonet et al. (2023c) illustrates the value of geometry-respecting directions while motivating an efficient sliced construction. CorSW addresses this opportunity on correlation matrices through geometry-aware prototype spaces whose directions retain $\mathcal{O}(d^2)$ sampling cost.

Correlation Matrices: Isometric Embeddings

Correlation matrices $C \in \mathrm{Cor}(d)$ admit *isometric embeddings* into Euclidean spaces Thanwerdas & Pennec (2022; 2024b):

$$\varphi : \mathrm{Cor}(d) \xrightarrow{\text{isometry}} V \subset \mathbb{R}^D, \quad D = \frac{d(d-1)}{2}, \tag{65}$$

where V is the model space (e.g., $\mathrm{LT}_0(d)$ for LECM, $\mathrm{Hol}(d)$ for OLM). The isometry ensures:

$$d_{\mathrm{Cor}}(C, C') = \|\varphi(C) - \varphi(C')\|_{\text{Euclidean}}. \tag{66}$$

GEOMETRY-RESPECTING EUCLIDEAN IMPLEMENTATION

A Euclidean implementation applies SW to $\varphi_\sharp\mu, \varphi_\sharp\nu$:

$$\text{CorSW}_p^p(\mu,\nu) = \int_{\mathbb{S}^{D-1}} W_p^p(\langle\varphi,\theta\rangle_\sharp\mu, \langle\varphi,\theta\rangle_\sharp\nu)d\theta, \tag{67}$$

with θ sampled from $\mathbb{S}^{D-1}$. The implementation respects CorSW geometry when directions are sampled in the Euclidean prototype space V with its induced inner product and constraints, including the strictly lower-triangular ECM/LECM space and the hollow or zero-row-sum OLM/LSM spaces.

Once directions are sampled uniformly from the unit sphere of the correct prototype space V, this implementation is mathematically equivalent to tangent-sphere sampling via $d\varphi_I$ by Bonet et al. (2025, Lem. 25).

D.2 PULLBACK IMPLEMENTATION VIA TANGENT SPACE DIFFERENTIALS

SPECIALIZATION TO CORRELATION EMBEDDINGS

The inherited pullback procedure is:

1. Sample $v \sim \lambda_I$ uniformly from $S_I = \{v \in T_I \operatorname{Cor}(d) : \|v\|_I = 1\}$
2. Compute $u = \operatorname{pushforward}\varphi_I(v) \in \mathbb{R}^D$ via the differential at I
3. Project as $B_v^\varphi(C) = -\langle\varphi(C) - \varphi(I), u\rangle_V$

The *Correlation Sliced-Wasserstein distance* is:

$$\text{CorSW}_p^p(\mu,\nu) = \int_{S_I} W_p^p(B_v^\varphi\sharp\mu, B_v^\varphi\sharp\nu)d\lambda_I(v). \tag{68}$$

The differential $\operatorname{pushforward}\varphi_I$ ensures: (i) directions respect $T_I \operatorname{Cor}(d)$ structure, (ii) sampling is geometrically uniform, (iii) distances reflect the true Riemannian metric.

EXPLICIT DIFFERENTIALS INHERITED FROM THE CORRELATION EMBEDDINGS

The following formulas collect the differentials at the identity of the correlation embeddings introduced in Thanwerdas & Pennec (2022; 2024b).

Proposition D.1 (Collected differentials at I)**.** *For $V \in T_I \operatorname{Cor}(n) \cong$ Hol(n) (symmetric hollow matrices), the differentials are:*

$$\textit{ECM:} \quad \operatorname{pushforward}\varphi_{EC,I}(V) = \lfloor V \rfloor, \tag{69}$$

$$\textit{LECM:} \quad \operatorname{pushforward}(\log\circ\Theta)_I(V) = \lfloor V \rfloor, \tag{70}$$

$$\textit{OLM:} \quad \operatorname{pushforward}\textit{Log}^\circ{}_I(V) = V, \tag{71}$$

$$\textit{LSM:} \quad \operatorname{pushforward}\textit{Log}^\star{}_I(V) = V - \textit{diag}(V\mathbf{1}), \tag{72}$$

where $\lfloor V \rfloor$ is the strictly lower-triangular part, $\mathbf{1} = (1,\ldots,1)^\top$.

Proof. These are obtained by differentiating the cited embeddings at I: ECM and LECM retain the strictly lower-triangular part, OLM is the identity on symmetric hollow tangent matrices, and LSM enforces the zero-row-sum constraint. See Thanwerdas & Pennec (2022; 2024b) for the parent maps and their geometry. □

Geometric Interpretation. **ECM/LECM**: Extract strictly lower-triangular entries, with prototype space $\text{LT}_0(d)$ and $D = d(d-1)/2$. **OLM**: Identity on symmetric hollow matrices $\text{Hol}(d)$, and permutation-invariant. **LSM**: Enforce row-sum-zero via $V - \text{diag}(V\mathbf{1})$, with prototype space $\text{Row}_0(d)$, and permutation-invariant.

AMBIENT AND GEOMETRY-RESPECTING PROJECTION

Remark D.2 (Constraint Enforcement). For LSM, geometry-respecting sampling via $\operatorname{pushforward}\varphi_I(v) = v - \text{diag}(v\mathbf{1})$ automatically enforces the row-sum-zero structure.

Table 7: Comparison of ambient Euclidean and geometry-respecting tangent projection.

Aspect	**Ambient**	**Geometry-respecting** ($\mathrm{pushforward}\ \varphi_I$)
Sampling	$\mathbb{S}^{D-1} \subset \mathbb{R}^D$	$S_I \subset T_I \operatorname{Cor}(d)$
Direction	Arbitrary θ	$v \in T_I \to u = \mathrm{pushforward}\ \varphi_I(v)$
Geometry	Ambient coordinates	Respects Riemannian metric
Invariance (OLM/LSM)	Ambient sampling	Yes (permutation)
Constraints	Ambient-space dependent	Automatically satisfied

Computational Complexity

We distinguish two levels of complexity: (i) the *per-projection* cost (one random direction, one pass over all samples) and (ii) the *total* cost over L projections and n_s, n_t samples. Following Bonet et al. (2023c), we focus on the asymptotic dependence on the matrix size d and treat the cost of matrix logarithms and Cholesky factorizations as $\mathcal{O}(d^3)$.

Per-projection cost. Table 8 summarizes the cost of generating one projection direction and projecting one SPD / correlation matrix onto that direction (the one-dimensional scalar coordinate). For SPDSW, each projection direction is a Frobenius-unit symmetric matrix $A = P\,\mathrm{Diag}(\theta)P^\top$, where $\theta \sim \mathrm{Unif}(S^{d-1})$ and $P \sim \mathrm{Haar}(O(d))$, as established in Bonet et al. (2023c, Lem. 2.4). A standard QR implementation of the Haar draw, together with forming A, costs $\mathcal{O}(d^3)$ per direction, and the subsequent matrix inner product $\langle A, \log X\rangle_F$ costs $\mathcal{O}(d^2)$ per sample. In Euclidean logSW, directions are sampled on the unit sphere in $\mathbb{R}^{d^2}$ and projection is a standard Euclidean inner product, both of order $\mathcal{O}(d^2)$.

For CorSW, an ambient-vectorization baseline simply applies Euclidean SW to the strictly lower-triangular correlation coordinates. Geometry-respecting CorSW samples unit tangent vectors $v \in S_I \subset T_I \operatorname{Cor}^+(d)$, pushes them forward via the differential $\varphi_{*,I}$ to obtain directions $u \in V$, and computes Busemann coordinates by prototype-space inner products, see Sec. 3.1. These operations are linear in the ambient dimension $D = \frac{d(d-1)}{2}$ and therefore $\mathcal{O}(d^2)$ per projection.

Key insight. Geometry-respecting CorSW retains the $\mathcal{O}(d^2)$ per-projection cost of an ambient Euclidean implementation while enforcing the Riemannian correlation geometry. In particular, it avoids the $\mathcal{O}(d^3)$ QR factorization required by SPDSW to generate symmetric directions on $O(d)$.

Total cost over L projections and n_s, n_t samples. Let n_s and n_t denote the number of source and target matrices, and $N = n_s + n_t$. For SPDSW on SPD matrices of size $d \times d$, Bonet et al. (2023c) show that one call to the distance has complexity

$$T_{\text{SPDSW}} = \mathcal{O}\big(LN(d^2 + \log N) + (L+N)d^3\big), \tag{73}$$

where the $\mathcal{O}(Nd^3)$ term comes from applying the matrix logarithm to all samples, the $\mathcal{O}(Ld^3)$ term from Haar-orthogonal direction sampling and forming $A = P\,\mathrm{Diag}(\theta)P^\top$, the $\mathcal{O}(LNd^2)$ term from matrix inner products, and the $\mathcal{O}(LN\log N)$ term from sorting 1D projections to compute Wasserstein distances Bonet et al. (2023c); Peyré & Cuturi (2019); Pele & Werman (2009).

For CorSW on correlation matrices, under any of the four pullback metrics (ECM, LECM, OLM, LSM), computing the embedding $\varphi(C)$ for all N samples requires $\mathcal{O}(KNd^3)$ time, where $K = 1$ for ECM/LECM (single Cholesky/logarithm) and K is a small constant ($\approx$ 3-5) for OLM/LSM due to the diagonal/row-sum solvers Thanwerdas & Pennec (2022; 2024b); Archakov & Hansen (2021). Once these embeddings are cached, each call to CorSW costs

$$T_{\text{CorSW}} = \mathcal{O}\big(LNd^2 + LN\log N\big), \tag{74}$$

corresponding to L Gaussian samples in $D = \mathcal{O}(d^2)$, LN inner products in V, and L one-dimensional Wasserstein computations via sorting.

Combining the two expressions, the full cost of CorSW including the initial embedding is

$$T_{\text{CorSW,total}} = \mathcal{O}\big(KNd^3 + LN(d^2 + \log N)\big), \tag{75}$$

which mirrors the structure of Eq. (73) but replaces SPDSW's $\mathcal{O}(Ld^3)$ QR sampling term with a cheaper $\mathcal{O}(Ld^2)$ Gaussian sampling in the pullback space. In practice, when $\varphi(C)$ is computed once

Table 8: Asymptotic computational complexity *per projection direction* (ignoring the initial $\mathcal{O}(d^3)$ embedding cost). Here d denotes the matrix size.

Method	Sampling	Projection	Total
SPDSW Bonet et al. (2023c)	$\mathcal{O}(d^3)$	$\mathcal{O}(d^2)$	$\mathcal{O}(d^3)$
logSW Bonneel et al. (2015)	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$
CorSW (ambient)	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$
CorSW (geometry-respecting)	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$

per dataset and reused throughout training, the dominant per-iteration cost of CorSW is linear in both the number of samples N and the number of projections L, with a quadratic dependence on d.

Hierarchy of Geometric Fidelity

Beyond raw complexity, the different sliced distances occupy a clear hierarchy in terms of how faithfully they respect the underlying manifold geometry and the scale structure of the data. Table 9 summarizes this trade-off between geometric fidelity, scale invariance, and computational cost.

logSW operates purely in a Euclidean tangent space at the identity and ignores the curvature of the SPD cone, so it is computationally cheap but geometrically coarse. SPDSW restores intrinsic SPD geometry by sampling symmetric directions on the manifold of SPD matrices and projecting along geodesics, but this comes with an $\mathcal{O}(d^3)$ sampling overhead per projection due to QR factorization on $O(d)$.

On the correlation side, the way Euclidean SW is applied matters. An ambient Euclidean SW that samples directions in the full matrix space and ignores the inner product and constraints of each model space (for example, the strictly lower-triangular structure for ECM/LECM, or the hollow and zero-row-sum constraints for OLM/LSM) does not act on $\mathrm{Cor}^+(d)$ itself.

However, by the inherited pullback-isometry argument of Bonet et al. (2025, Lem. 25), if directions are sampled uniformly from the unit sphere of the *correct prototype space* V (with the proper inner product), this Euclidean SW implementation is mathematically identical to tangent-sphere CorSW. Thus, proper CorSW retains the same $\mathcal{O}(d^2)$ per-projection complexity of Euclidean sliced distances while preserving two useful properties: (i) it respects the geometry of $\mathrm{Cor}^+(d)$ through the pullback isometry φ and its differential $\mathrm{pushforward}\,\varphi_I$, and (ii) it inherits the scale invariance of correlation matrices, which we empirically observe to be useful for cross-session and cross-subject BCI transfer.

Conclusion. By using the differential $\mathrm{pushforward}\,\varphi_I$ to map tangent directions to the correct prototype space, CorSW combines geometry-aware sampling, $\mathcal{O}(d^2)$ cached projections, scale invariance, and symmetry preservation. The relevant distinction from ambient Euclidean SW is not merely the existence of an isometric embedding, but the use of its differential to sample and project in the corresponding model space.

E Experiment Setup and Complete Results

E.1 Compute and reporting protocol

The performance sweeps were run on the Leonardo GPU cluster using the supplied TSMNet, SPDNet, and SPDGCN code. The standalone runtime benchmark in App. F was run on an NVIDIA RTX 4090. All reported accuracy and controlled-transport results aggregate seven independent seeds. Tables in this supplement report the mean and population standard deviation. A diagonal jitter of 10^{-6} is used before Cholesky or matrix-logarithm operations.

For SPDGCN PHCML-Busemann, all 126 runs in the input-scale sweep completed: scales $\{0.25, 0.50, 0.75\}$ across Cora, Disease-NC, and PubMed, under CorSW-only and Mix, and over seven model seeds. For each dataset, we selected the mode and scale with the largest validation accuracy, without using test accuracy for selection, and report the corresponding mean test accuracy with population standard deviation. The selected settings are Cora CorSW-only at 0.25 (validation $76.80 \pm 1.75\%$, loss $0.8105 \pm 0.0531\%$, and test $80.28 \pm 0.75\%$), Disease-NC Mix at 0.75 (validation

Table 9: Hierarchy of geometric fidelity and computational cost for different sliced distances. "Geometry" indicates the space where projections are defined, and "Scale-Inv." denotes whether the representation is invariant to channel-wise rescaling (correlation vs. covariance).

Method	Geometry	Sampling	Scale-Inv.	Cost
logSW	Tangent to $\mathrm{SPD}(d)$	Spherical in $\mathbb{R}^{d^2}$	No	$\mathcal{O}(d^2)$
SPDSW Bonet et al. (2023c)	SPD manifold ($\mathrm{SPD}(d)$)	Symmetric on $O(d)$	No	$\mathcal{O}(d^3)$
CorSW (ambient)	Off-diagonal vectorization	Spherical in $\mathbb{R}^D$	Yes	$\mathcal{O}(d^2)$
CorSW (proper)	Full Riem. via pushforward φ_I	Tangent sphere S_I	Yes	$\mathcal{O}(d^2)$

$93.81 \pm 3.56\%$, loss 0.2366 ± 0.0805, and test $93.54 \pm 3.89\%$), and PubMed Mix at 0.25 (validation $78.56 \pm 1.64\%$, loss 0.6508 ± 0.0129, and test $78.48 \pm 0.99\%$).

The fixed-scale PHCML-Busemann report uses scale 0.25 and seven seeds. For RMLR, values are the maximum validation accuracy per seed, aggregated with population standard deviation. For HINSS transfer, values are test scores averaged over the subject-session records per mode, and the standard deviation is across seven seed means. Throughout the paper, `logmap0` followed by Euclidean SW denotes the PHCML-Log route, whereas PHCML-Busemann uses the horospherical coordinate directly.

For the EEG fusion experiment, all methods share one pipeline and one setting. A shared SPD backbone (translation and rotation) maps the source covariances toward the fixed target distribution using only source and target covariances, without labels, and a linear classifier fit on the aligned source features is evaluated on the target. *No align* classifies the unaligned features, *SPDSW* trains the backbone with the covariance alignment loss alone, and each CorSW variant adds its correlation alignment loss on $\phi(C)$, combined with the covariance term by learned nonnegative weights. The correlation maps are parameter-free, so all rows share the same backbone and evaluation, and a CorSW row adds only the two scalar fusion weights. All methods are run under the same hyperparameter setting, so the rows in the fusion tables are directly comparable.

For regularization experiments, **CorSW-only** denotes the correlation-geometry sliced term, **LOGM-only** denotes covariance SPDSW, and **Mix** combines them. TSMNet and SPDGCN use the fixed correlation-dominant weights $(\lambda_{\mathrm{CorSW}}, \lambda_{\mathrm{LOGM}}) = (0.95, 0.05)$ in Mix mode. SPDNet uses the learned-weight implementation initialized according to its manifest. The proposed family includes ECM, LECM, OLM, LSM, PHCM-Prod, and PHCML-HCat. PHCM-Prod retains the product factors and uses Product-Log, Product-Busemann, or Product-Geodesic (approx.). PHCML first applies the same row-wise PHCM encoding and then HCat. PHCML-Log performs Euclidean slicing after the Lorentz logarithmic map, PHCML-Busemann applies the intrinsic Busemann coordinate, and PHCML-G uses the intrinsic geodesic coordinate.

E.2 Controlled gradient-flow simulation

For each of seven paired seeds, the simulation uses the same 192 source correlation matrices for an interior and a near-boundary 192-particle target on $\mathrm{Cor}^+(3)$. Every method runs 400 steps with step size 0.10, 256 training directions per step, and 512 fixed evaluation directions per checkpoint. The compared methods are ECM, LECM, OLM, LSM, PHCM-Log, PHCM-G, and PHCM-Busemann. The suffixes Log, Busemann, and G consistently identify the slice coordinate. Cross-method comparisons use empirical W_2 in ambient off-diagonal coordinates, and native sliced losses are retained only as within-method diagnostics because their units differ.

The empirical correlation invalid rate and maximum symmetry error are zero at every checkpoint, and the maximum unit-diagonal error is 1.4×10^{-6}. The mean Lorentz constraint residual is at most 8.4×10^{-6}, and all reconstructed correlation matrices remain positive definite. Complete per-seed traces, metadata, and aggregation code are included in the reproducibility bundle.

E.3 Datasets and architectures

EEG. BCIC-IV-2a Brunner et al. (2008) and BNCI2015-001 Faller et al. (2012) are used for learned fusion. HINSS2021 Hinss et al. (2021) is used for inter-session and inter-subject transfer with

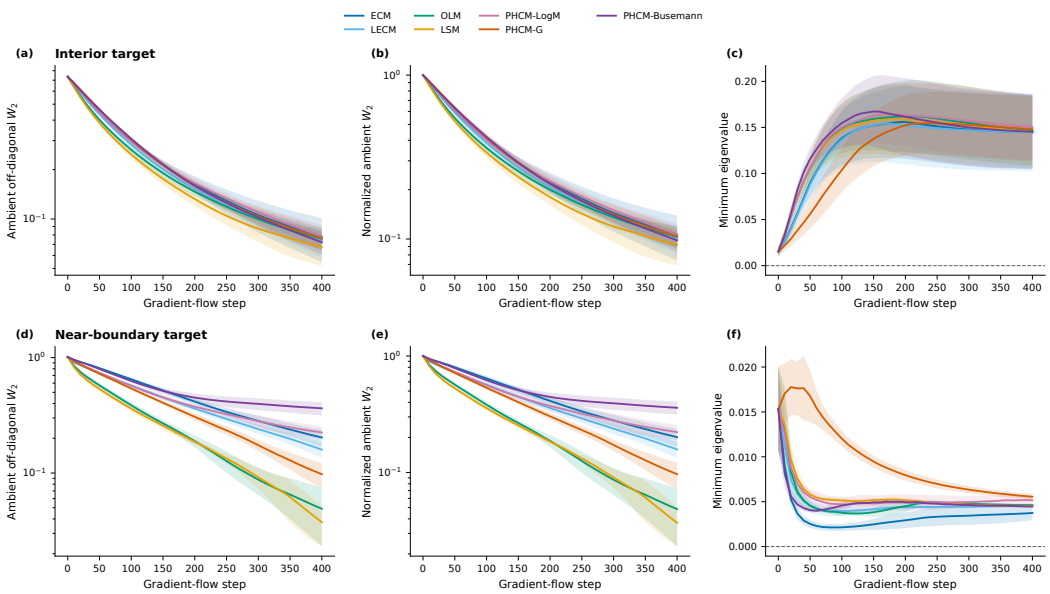


Figure 3: Complete controlled-flow curves. Rows are interior and near-boundary targets, and columns are ambient W_2, normalized ambient W_2, and minimum eigenvalue (mean $\pm$ std over seven seeds).

TSMNet+SPDDSMBN Kobler et al. (2022). Signals are processed with the MOABB/MNE pipeline used by TSMNet, including resampling, 4–36 Hz filtering for HINSS2021/BNCI2015-001, temporal-spatial filtering, and covariance pooling. All transfer tables in this appendix refer to HINSS2021.

SPD classification. Radar Brooks et al. (2019) contains 20×20 covariance descriptors. FPHA Garcia-Hernando et al. (2018) is evaluated with architecture $[63, 20, 16, 8]$. HDM05 Müller et al. (2007) uses 93×93 covariance descriptors and the reported architectures $[93, 70, 30]$ and $[93, 70, 50, 30]$. Radar is reported for $[20, 16, 8]$ and $[20, 16, 14, 12, 10, 8]$. The classifier applies LogEig followed by a linear prediction head, and the regularizer is attached to the selected SPD feature layer.

Graphs. Cora, PubMed, and Disease-NC are evaluated with the supplied SPDGCN implementation. The classifier maps the final SPD feature through LogEig, vectorizes the upper triangle, and applies a linear head. The official settings use learning rate 0.01, dropout 0.5, weight decay 0.005, SPD hidden dimension 3, ReEig nonlinearity, 300 epochs, and early-stopping patience 200. RDA is evaluated on training nodes only, and validation and test losses remain cross entropy.

E.4 Complete SPDNet results

The split tables make the family-level pattern explicit: ECM is strongest on Radar, OLM on FPHA, and LSM on HDM05 within the flat branch. In the PHCML branch, the fixed-scale Busemann route leads both reported HDM05 architectures. PHCM-Prod is competitive on HDM05 but underperforms both baselines on Radar and FPHA. The product branch is therefore not a uniformly stronger replacement for HCat. Its value is task-dependent and is clearest in the transfer and graph results below.

E.5 Complete HINSS2021 results

For inter-session transfer, all four pullback-flat metrics improve over the None baseline in CorSW-only mode, with LECM-Mix reaching the highest flat-branch result of $56.14 \pm 0.79\%$. The PHCM-Prod branch achieves the overall best inter-session result with Prod-B at $56.99 \pm 0.52\%$, surpassing all flat and PHCML variants. On inter-subject transfer, Prod-Log reaches $53.08 \pm 0.76\%$ in CorSW-only mode and $53.75 \pm 0.26\%$ in Mix mode, improving over the strongest PHCML result of $52.42 \pm 0.71\%$. Prod-G (approx.) reaches $51.84 \pm 0.76\%$ and $51.78 \pm 0.85\%$, while Prod-B reaches $50.49 \pm 0.56\%$

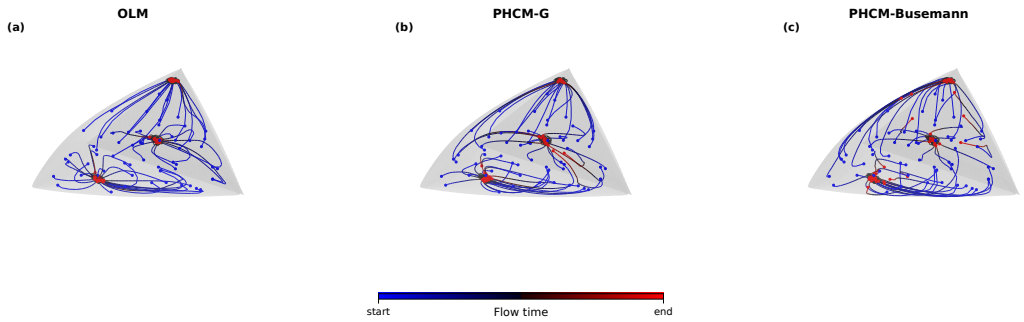


Figure 4: Near-boundary trajectories for seed 0. The same source and target particles are used for OLM, PHCM-G, and PHCM-Busemann, and color runs from the initial to final flow step.

and $50.45 \pm 0.66\%$. The complete tables therefore expose distinct coordinate preferences within the product branch across the two transfer protocols.

E.6 Complete SPDGCN results

These PHCML-branch results report CorSW-only and Mix for all three slice-coordinate routes on the shared HCat representation, making the effect of both slice choice and covariance–correlation fusion directly inspectable. The PHCM-Prod branch achieves competitive performance on Cora and PubMed, and attains the highest Disease-NC Mix result at $95.35 \pm 2.57\%$ with Prod-B, closely matching PHCML-Log Mix at $95.67 \pm 2.54\%$.

E.7 Fusion and weight ablations

PHCML-B is reported here and in the main table for completeness. Its $77.87 \pm 3.92\%$ and $78.53 \pm 13.05\%$ results do not improve the corresponding unaligned means, reinforcing that the preferred slice coordinate is task-dependent rather than uniformly ordered.

This ablation isolates the scale-invariant correlation term. Together with Tab. 20, it motivates fusing the two alignment losses: the CorSW term captures transferable correlation structure while the covariance SPDSW term retains complementary amplitude information.

The monotonic degradation as the LOGM share grows shows that equal-weight correlation–covariance regularization is not generally complementary in this setting.

F Runtime Benchmark on High-Dimensional Matrices

F.1 Experimental Settings

We benchmark wall-clock runtime under single-GPU CUDA execution using PyTorch 2.1.2. Each configuration uses a batch of $b = 16$ synthetic SPD matrices at dimensions $d \in \{50, 100, 200, 500\}$. We draw $X \in \mathbb{R}^{b\times d\times d}$ with independent standard Gaussian entries and form $XX^\top + \epsilon I$, where $\epsilon = 10^{-3}$ ensures strict positive definiteness. SPDSW receives these SPD matrices, whereas the CorSW routes receive their correlation normalizations, matching the native input of each objective.

All methods use $L = 50$ directions and $\texttt{power} = 1$, thereby excluding the cost of an optional fractional matrix power. The RTX 4090 benchmark includes SPDSW, ECM, LECM, OLM, and LSM, all three PHCM-Product coordinates (Log, Busemann, and Product-Geodesic (approx.)), and all three PHCML-HCat coordinates (Log, Busemann, and G). Each method is evaluated through its native mapping and sliced objective rather than through a shared projection head.

F.2 Benchmarking Methodology

We decompose runtime into one cached setup phase and two per-step phases:

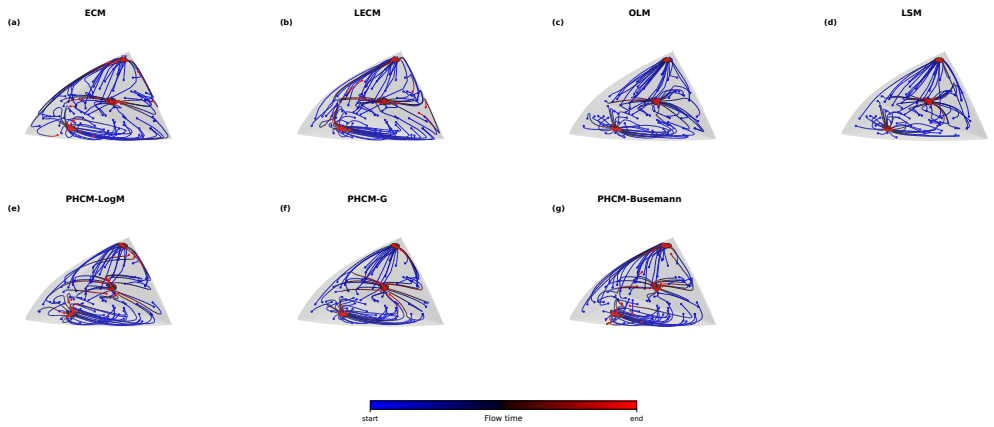


Figure 5: Near-boundary trajectories for all seven geometries (seed 0). Viewpoint, elliptope, source particles, target particles, and flow budget are fixed across panels.

- **Sampling** constructs the loss module and materializes its directions. Following Bonet et al. (2023c, Lem. 2.4), SPDSW samples a Frobenius-unit symmetric direction as $A = P\,\mathrm{Diag}(\theta)P^\top$, where $P \sim \mathrm{Haar}(O(d))$. In the standard implementation, drawing P by QR factorization and forming A costs $\mathcal{O}(d^3)$ per direction. CorSW instead draws and normalizes Gaussian directions in a $D = d(d-1)/2 = \mathcal{O}(d^2)$ model space. Directions are reused across subsequent loss evaluations.
- **Forward propagation (FP)** includes the input-to-coordinate mapping, scalar projections, and one-dimensional Wasserstein losses.
- **Backward propagation (BP)** measures automatic differentiation of the loss with respect to the input matrices.

We report **Step**= FP+BP and **Full**= Sampling+FP+BP. Only the sampled directions are cached. In particular, FP includes correlation normalization and every native Cholesky, matrix-logarithm, diagonal solver, Lorentz map, scalar projection, and one-dimensional Wasserstein operation used by that method. Step isolates repeated optimization cost, whereas Full is a complete one-shot loss call with direction materialization.

For each configuration, ten warm-up iterations initialize the CUDA context and stabilize execution, followed by 100 measured iterations. We call `torch.cuda.synchronize()` immediately before and after every measured phase to exclude asynchronous-launch artifacts. The tables report mean wall-clock time. Gradient norms are not part of the runtime comparison.

F.3 Results and Analysis

Tables 24 and 25 gives both timing definitions. Among the flat routes, LECM moves from approximate parity with SPDSW at $d = 50$ to Step times of 0.0536 and 0.1668 seconds at $d = 200$ and 500, compared with 0.1086 and 0.3896 seconds for SPDSW. OLM remains near SPDSW, while the more involved LSM coordinate map dominates its Step time.

The curved routes are light under their native implementations. PHCM-Product evaluates normalized Cholesky-row factors and combines closed-form factor coordinates under one global direction, while PHCML concatenates the same rows and applies a vectorized Lorentz coordinate. At $d = 500$, the Product routes require 0.0121–0.0142 seconds per Step, while PHCML requires 0.0103–0.0125 seconds.

These timings reflect realized native implementations, not a stronger asymptotic claim. PHCM-Product and PHCML still require a Cholesky-based map with cubic worst-case cost, while SPDSW evaluates a full SPD matrix logarithm and uses dense matrix-valued coordinates. Full captures direction construction in addition to the complete native FP/BP path. It remains a kernel-level one-shot comparison rather than end-to-end training or accuracy-versus-wall-clock evidence.

Table 10: Gradient-flow endpoints over seven paired seeds (mean ± population std). W_2 is computed in ambient off-diagonal coordinates, and $\lambda_{\min}$ is the smallest particle eigenvalue.

Method	Interior target		Near-boundary target	
	Final W_2	Final $\lambda_{\min}$	Final W_2	Final $\lambda_{\min}$
ECM	0.0780±0.0234	0.1450±0.0405	0.2030±0.0342	0.0037±0.0008
LECM	0.0737±0.0150	0.1443±0.0414	0.1594±0.0241	0.0046±0.0002
OLM	0.0757±0.0108	0.1468±0.0387	0.0491±0.0260	0.0046±0.0002
LSM	**0.0675±0.0155**	0.1450±0.0389	**0.0375±0.0141**	0.0045±0.0002
PHCM-Log	0.0784±0.0133	**0.1504±0.0361**	0.2241±0.0279	0.0052±0.0005
PHCM-G	0.0763±0.0117	0.1481±0.0353	0.0976±0.0241	**0.0056±0.0003**
PHCM-Busemann	0.0721±0.0112	0.1452±0.0399	0.3635±0.0482	0.0045±0.0001

Table 11: Complete SPDNet results for the flat branch (accuracy, mean ± std over seven seeds, %). CorSW and Mix denote CorSW-only and CorSW+LOGM regularization, respectively.

Dataset	Architecture	None	SPDSW	ECM		LECM		OLM		LSM	
				CorSW	Mix	CorSW	Mix	CorSW	Mix	CorSW	Mix
Radar	[20, 16, 14, 12, 10, 8]	90.85±0.84	91.956±1.068	**92.960±1.336**	92.456±1.629	92.40±1.62	91.758±0.861	92.264±1.308	92.266±1.749	91.574±1.517	92.108±0.971
Radar	[20, 16, 8]	90.45±0.84	91.422±1.040	**92.856±1.161**	92.346±1.200	92.664±0.956	92.374±1.733	92.000±1.311	91.976±1.884	90.906±1.390	91.172±0.584
FPHA	[63, 20, 16, 8]	68.992±1.885	61.292±0.310	66.282±2.150	66.280±2.031	69.016±1.987	69.050±1.377	**70.816±1.680**	69.584±3.244	69.782±2.346	69.618±1.558
HDM05	[93, 70, 30]	60.690±0.660	60.786±1.049	60.671±0.821	60.173±0.517	61.668±0.922	61.841±1.317	61.860±0.668	60.441±1.164	**62.052±0.586**	60.767±0.674
HDM05	[93, 70, 50, 30]	61.060±1.310	61.477±0.906	60.345±1.402	60.594±1.273	61.726±1.012	61.994±0.524	62.244±0.692	61.342±1.087	**62.608±0.673**	61.707±0.952

Table 12: Complete SPDNet results for the PHCML representation branch (accuracy, mean ± std over seven seeds, %). PHCML-Busemann entries use the fixed-scale 0.25 report.

Dataset	Architecture	PHCML-Log		PHCML-Busemann		PHCML-G	
		CorSW	Mix	CorSW	Mix	CorSW	Mix
Radar	[20, 16, 14, 12, 10, 8]	90.448±1.722	90.392±1.932	89.62±1.30	89.67±1.47	89.23±1.65	89.20±1.68
Radar	[20, 16, 8]	91.032±1.349	90.312±1.298	90.13±0.66	90.02±0.97	89.68±1.26	89.73±1.54
FPHA	[63, 20, 16, 8]	70.150±1.660	68.818±1.935	**70.40±1.25**	69.74±2.09	**70.40±1.24**	69.32±2.29
HDM05	[93, 70, 30]	62.033±0.588	60.633±0.724	**63.03±0.51**	61.69±0.86	61.71±0.54	60.85±0.75
HDM05	[93, 70, 50, 30]	62.454±0.905	61.553±0.899	**63.39±0.53**	62.57±0.71	62.22±0.43	61.51±1.03

Table 13: Complete SPDNet results for the PHCM-Prod (product-space) branch (accuracy, mean ± std over seven seeds, %).

Dataset	Architecture	PHCM-Prod-Log		PHCM-Prod-B		PHCM-Prod-G (approx.)	
		CorSW	Mix	CorSW	Mix	CorSW	Mix
Radar	[20, 16, 14, 12, 10, 8]	88.88±0.88	88.75±0.75	88.93±1.50	88.40±1.00	88.69±0.80	88.37±1.03
Radar	[20, 16, 8]	88.67±2.33	88.72±1.56	88.53±1.86	88.05±2.33	88.77±2.24	88.77±1.47
FPHA	[63, 20, 16, 8]	53.53±2.13	51.86±2.03	55.41±2.32	54.30±1.77	53.88±2.02	52.49±2.15
HDM05	[93, 70, 30]	62.45±0.69	62.44±0.68	62.32±0.63	62.34±0.66	62.44±0.67	62.42±0.66
HDM05	[93, 70, 50, 30]	62.68±0.74	62.70±0.70	62.63±0.86	62.61±0.79	62.68±0.74	62.72±0.70

Table 14: HINSS2021 TSMNet+SPDDSMBN transfer accuracy for the flat branch (mean ± std over seven seeds, %). LOGM is the metric-independent SPDSW baseline. Mix uses 0.95 CorSW + 0.05 LOGM.

Task	None	LOGM	ECM		LECM		OLM		LSM	
			CorSW	Mix	CorSW	Mix	CorSW	Mix	CorSW	Mix
Inter-session	54.48±0.00	55.33±0.75	56.06±0.72	56.11±0.72	56.10±0.90	**56.14±0.79**	55.10±0.92	55.00±0.85	55.59±0.58	55.48±0.61
Inter-subject	49.68±7.88	50.83±0.79	51.68±0.47	51.13±0.44	51.06±0.39	51.42±0.85	51.16±0.89	51.62±0.42	51.50±0.65	**51.73±0.42**

Table 15: HINSS2021 TSMNet+SPDDSMBN transfer accuracy for the PHCML representation branch (mean ± std over seven seeds, %). PHCML-Busemann entries use fixed-scale-0.25 test scores.

Task	PHCML-Log		PHCML-Busemann		PHCML-G	
	CorSW	Mix	CorSW	Mix	CorSW	Mix
Inter-session	55.24±0.94	**55.55±0.90**	54.42±0.51	54.60±0.41	54.98±0.55	54.82±0.52
Inter-subject	**52.42±0.71**	52.19±0.72	51.65±1.08	51.05±1.00	50.63±0.68	50.62±0.46

Table 16: HINSS2021 TSMNet+SPDDSMBN transfer accuracy for the PHCM-Prod (product-space) branch (mean ± std, %). Inter-session uses seven seeds, while inter-subject uses five seeds over 15 held-out subjects.

Task	PHCM-Prod-Log		PHCM-Prod-B		PHCM-Prod-G (approx.)	
	CorSW	Mix	CorSW	Mix	CorSW	Mix
Inter-session	53.92±0.40	54.32±0.45	**56.99±0.52**	**56.98±0.53**	54.79±0.63	54.88±0.73
Inter-subject	**53.08±0.76**	**53.75±0.26**	50.49±0.56	50.45±0.66	51.84±0.76	51.78±0.85

Table 17: SPDGCN node-classification accuracy for the flat branch (mean ± std over seven seeds, %), including the supplied None and SPDSW baselines.

Dataset	Baselines		ECM		LECM		OLM		LSM	
	None	SPDSW	CorSW	Mix	CorSW	Mix	CorSW	Mix	CorSW	Mix
Cora	80.12±0.40	80.18±0.76	80.18±0.95	79.88±0.76	80.10±0.69	80.20±0.72	80.30±0.93	**80.52±1.16**	80.04±1.05	80.12±0.64
PubMed	74.92±0.45	77.04±1.06	77.78±0.82	78.12±0.59	77.50±0.98	77.64±0.35	77.92±0.69	77.48±0.93	78.08±0.74	**78.56±0.48**
Disease-NC	77.09±20.77	89.84±1.38	86.38±12.76	86.61±10.67	92.52±1.32	93.31±2.27	93.15±4.08	**95.83±1.49**	90.08±7.26	93.62±4.62

Table 18: SPDGCN test accuracy for the complete PHCML representation branch (mean ± std over seven seeds, %).

Dataset	PHCML-Log		PHCML-Busemann		PHCML-G	
	CorSW	Mix	CorSW	Mix	CorSW	Mix
Cora	79.60±1.37	**80.36±1.13**	80.28±0.75	79.84±0.68	79.82±1.00	80.04±0.58
PubMed	78.18±0.71	77.44±1.26	78.10±0.73	**78.48±0.99**	77.44±0.87	77.50±0.88
Disease-NC	91.65±4.48	**95.67±2.54**	92.05±4.65	93.54±3.89	93.54±2.25	93.86±4.03

Table 19: SPDGCN test accuracy for the PHCM-Prod (product-space) branch (mean ± std over seven seeds, %).

Dataset	PHCM-Prod-Log		PHCM-Prod-B		PHCM-Prod-G (approx.)	
	CorSW	Mix	CorSW	Mix	CorSW	Mix
Cora	79.54±0.94	79.52±1.28	80.12±0.81	79.80±1.20	80.20±0.50	79.62±1.32
PubMed	77.56±0.57	77.46±0.76	77.70±1.04	77.70±0.45	77.60±1.06	78.12±0.53
Disease-NC	93.39±1.06	93.62±0.85	93.46±3.88	**95.35±2.57**	92.60±1.96	91.02±5.60

Table 20: Complete covariance–correlation fusion accuracy (mean ± std over seven seeds, %). Every highlighted CorSW result is significantly better than SPDSW.

Method	BCIC-IV-2a	BNCI2015-001
No align	77.87±3.92	81.50±12.19
SPDSW	82.21±3.06	82.57±12.61
ECM	82.31±3.55	83.99±12.14
LECM	82.44±3.20	83.53±12.96
OLM	**82.69±2.81**	83.40±12.72
LSM	82.52±2.94	83.37±12.85
PHCM-Prod-Log	79.42±3.91	83.61±12.47
PHCM-Prod-B	79.28±4.60	82.34±13.36
PHCM-Prod-G (approx.)	79.09±4.23	81.34±13.35
PHCML-Log	82.49±3.40	**84.20±11.63**
PHCML-B	77.87±3.92	78.53±13.05
PHCML-G	81.98±3.48	83.70±12.71

Notation: Shaded entries indicate the top three complete means in each dataset column. All shaded CorSW entries significantly outperform SPDSW.

Table 21: Complete matched geometry controls after correlation normalization (mean ± std over seven paired seeds, %). These entries use the same RDA protocol as the corresponding CorSW-only and Mix results. Corr+Cov retains the exact CorSW-Mix covariance term, fusion weights, and tuning while replacing only CorSW by Corr-SPD.

Setting	R_S	R_D	H_{30}	H_{50}	Cora	PubMed	Disease-NC
RawCorr	89.17±1.62	89.31±1.72	59.10±0.62	61.76±0.60	79.08±1.06	77.74±0.79	92.68±5.49
Corr-SPD	89.78±1.04	89.73±0.60	61.11±0.44	61.36±0.81	79.88±0.80	77.62±0.90	91.34±0.28
Corr+Cov	89.64±1.20	89.65±0.77	61.14±0.94	61.49±0.93	80.24±0.89	78.00±0.76	83.86±18.94

Table 22: Correlation-only EEG accuracy (%). These runs remove learned covariance–correlation fusion.

Dataset	No align	ECM	LECM	OLM	LSM
BCIC-IV-2a	**77.87**	56.15	73.19	77.71	75.70
BNCI2015-001	78.47	78.03	78.47	**79.30**	59.00

Table 23: FPHA validation accuracy for fixed CorSW/LOGM mixture weights (seven seeds).

λ_{LOGM}	λ_{CorSW}	Accuracy (%)
0.05	0.95	**70.75±1.60**
0.10	0.90	69.91±2.25
0.25	0.75	68.56±2.27
0.50	0.50	66.43±1.71

Table 24: RTX 4090 runtime for $d \in \{50, 100\}$ with $b = 16$, $L = 50$, and $\texttt{power} = 1$. Step= FP+BP and Full= Sampling+Step. FP includes the native matrix-to-coordinate map. All values are seconds.

Method	Step	Full
$d = 50$		
SPDSW	0.026373	0.069155
CorSW (ECM)	0.008621	0.010519
CorSW (LECM)	0.026760	0.028766
CorSW (OLM)	0.026001	0.029316
CorSW (LSM)	0.049578	0.052625
CorSW (PHCM-Prod-Log)	0.010079	0.021281
CorSW (PHCM-Prod-B)	0.010035	0.016708
CorSW (PHCM-Prod-G (approx.))	0.010407	0.017715
CorSW (PHCML-Log)	0.005974	0.008447
CorSW (PHCML-B)	0.008059	0.010717
CorSW (PHCML-G)	0.008998	0.011474
$d = 100$		
SPDSW	0.047346	0.081900
CorSW (ECM)	0.007845	0.011689
CorSW (LECM)	0.040643	0.042871
CorSW (OLM)	0.049997	0.057831
CorSW (LSM)	0.082688	0.089565
CorSW (PHCM-Prod-Log)	0.011910	0.026757
CorSW (PHCM-Prod-B)	0.012381	0.028301
CorSW (PHCM-Prod-G (approx.))	0.012392	0.028885
CorSW (PHCML-Log)	0.007949	0.016259
CorSW (PHCML-B)	0.010314	0.016938
CorSW (PHCML-G)	0.011085	0.018399

Table 25: RTX 4090 runtime for $d \in \{200, 500\}$ under the same protocol and definitions as Tab. 24. All values are seconds.

Method	Step	Full
	$d = 200$	
SPDSW	0.108647	0.200589
CorSW (ECM)	0.010463	0.019376
CorSW (LECM)	0.053640	0.065273
CorSW (OLM)	0.104358	0.131853
CorSW (LSM)	0.220950	0.240446
CorSW (PHCM-Prod-Log)	0.010980	0.046641
CorSW (PHCM-Prod-B)	0.011016	0.041715
CorSW (PHCM-Prod-G (approx.))	0.011513	0.041685
CorSW (PHCML-Log)	0.006926	0.026119
CorSW (PHCML-B)	0.009475	0.019329
CorSW (PHCML-G)	0.010056	0.028356
	$d = 500$	
SPDSW	0.389551	0.820124
CorSW (ECM)	0.015454	0.063379
CorSW (LECM)	0.166789	0.246230
CorSW (OLM)	0.365906	0.452678
CorSW (LSM)	0.622185	0.750617
CorSW (PHCM-Prod-Log)	0.012074	0.115153
CorSW (PHCM-Prod-B)	0.013055	0.142183
CorSW (PHCM-Prod-G (approx.))	0.014152	0.137586
CorSW (PHCML-Log)	0.010275	0.138422
CorSW (PHCML-B)	0.011614	0.096535
CorSW (PHCML-G)	0.012489	0.133754

G Extension to Rank-Deficient Correlations

CorSW currently targets strictly full-rank correlation matrices in the interior of the elliptope. Extending the family to rank-deficient empirical correlations will bring the same geometry-aware alignment principle to boundary and lower-dimensional strata, using positive-semidefinite correlation geometries and slice coordinates such as stratified or quotient constructions.

H Theoretical Derivations and Proofs

H.1 Derivation of the pullback projection specialization

Proof. This calculation substitutes the correlation origin into the pullback-Euclidean parent identity of Bonet et al. (2025, Prop. 7). According to (Bridson & Haefliger, 2013, Chapter II, Section 8), for all $x \in \mathcal{M}$, we have:

$$B^v(x) = \lim_{t\to\infty} \left(d_{\mathcal{M}}(\gamma_v(t), x) - t\right) = \lim_{t\to\infty} \frac{d_{\mathcal{M}}(\gamma_v(t), x)^2 - t^2}{2t}. \tag{76}$$

Let $\gamma_v : t \mapsto \phi^{-1}(\phi(I_d) + t\phi_{*,I_d}(v))$ be the geodesic associated with $\mathcal{G}^v$. Then, we obtain:

$$\begin{aligned}
\frac{d_{\mathcal{M}}(\gamma_v(t), x)^2 - t^2}{2t} &= \frac{1}{2t}\left(\|\phi(\gamma_v(t)) - \phi(x)\|_V^2 - t^2\right) \\
&= \frac{1}{2t}\left(\|\phi(I_d) + t\phi_{*,I_d}(v) - \phi(x)\|_V^2 - t^2\right) \\
&= \frac{1}{2t}\Big(t^2\|\phi_{*,I_d}(v)\|_V^2 - 2t\langle\phi_{*,I_d}(v), \phi(x) - \phi(I_d)\rangle_V \\
&\qquad + \|\phi(x) - \phi(I_d)\|_V^2 - t^2\Big) \\
&= -\langle\phi_{*,I_d}(v), \phi(x) - \phi(I_d)\rangle_V + \frac{1}{2t}\|\phi(x) - \phi(I_d)\|_V^2,
\end{aligned}$$

where we have used the fact that $\|v\|_{I_d} = \|\phi_{*,I_d}(v)\|_V = 1$.

Taking the limit as $t \to \infty$, we find:

$$B^v(x) = -\langle\phi_{*,I_d}(v), \phi(x) - \phi(I_d)\rangle_V. \tag{77}$$

□

H.2 Derivations of the ECM and LECM coordinates

For the Euclidean-Cholesky metric, $\phi = \Theta$, and for the Log-Euclidean-Cholesky metric, $\phi = \log \circ \Theta$. Both formulas follow by substitution into the inherited pullback identity in Cor. 3.1.

H.3 Algebraic expansion of the inherited OLM coordinate

The explicit busemann projection for the Off-Log Metric (OLM) across all dimensions is given by:

$$\Pi^v_{\mathrm{OLM}}(x) = \begin{cases} -\gamma\mathrm{Sum}(\log x) & \text{if } d = 2, \\ -2\gamma\mathrm{Sum}(v)\mathrm{Sum}(\log x) - \frac{\beta}{2}\mathrm{Sum}(v \circ \log x + (\log x) \circ v) & \text{if } d = 3, \\ -\frac{1}{2}\big(\alpha\,\mathrm{tr}(v^\top \log x + (\log x)^\top v) + 2\gamma\mathrm{Sum}(v)\mathrm{Sum}(\log x) & \\ \quad +\beta\mathrm{Sum}(v \circ \log x + (\log x) \circ v)\big) & \text{if } d \geq 4. \end{cases} \tag{78}$$

Case $n = 2$. The squared distance between a point x and a point on the geodesic $\gamma(t)$ can be expressed as:

$$d^2(x, \gamma(t)) = \gamma\left(t\,\mathrm{Sum}(V) - \mathrm{Sum}(\mathrm{Log}(x))\right)^2. \tag{79}$$

For the busemann function, we compute:

$$\begin{aligned} B(x) &= \lim_{t\to\infty} \left(d(x,\gamma(t)) - t\right) \\ &= \lim_{t\to\infty} \frac{d(x,\gamma(t))^2 - t^2}{2t} \\ &= \lim_{t\to\infty} \frac{1}{2t}\Big(\gamma t^2(\operatorname{Sum}(V))^2 + \gamma(\operatorname{Sum}(\operatorname{Log}(x)))^2 \\ &\qquad - 2t\gamma \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x)) - t^2\Big). \end{aligned} \tag{80}$$

For $n = 2$ with unit norm, $\|V\| = \operatorname{Sum}(V) = 1$, the terms $\gamma t^2(\operatorname{Sum}(V))^2 = \gamma t^2$ cancels with $-t^2$. Thus:

$$\begin{aligned} B(x) &= \lim_{t\to\infty} \frac{1}{2t}\Big(- 2t\gamma \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x)) + \gamma(\operatorname{Sum}(\operatorname{Log}(x)))^2\Big) \\ &= -\gamma \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x)). \end{aligned} \tag{81}$$

Case $n = 3$**.** The squared distance is given by:

$$\begin{aligned} d^2 = \gamma\big(&t^2 \operatorname{Sum}^2(V) + \operatorname{Sum}^2(\operatorname{Log}(x)) - 2t \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x))\big) \\ &+ t^2\beta \operatorname{Sum}(V^2) - t\beta\big(\operatorname{Sum}(V\operatorname{Log}(x) + \operatorname{Log}(x)V)\big) \\ &+ \beta \operatorname{Sum}(\operatorname{Log}^2(x)). \end{aligned} \tag{82}$$

For the busemann function, we compute:

$$\begin{aligned} B(x) &= \lim_{t\to\infty} \frac{d^2 - t^2}{2t} \\ &= \lim_{t\to\infty} \frac{1}{2t}\Big(t^2(\gamma \operatorname{Sum}^2(V) + \beta \operatorname{Sum}(V^2)) - t^2 \\ &\qquad - 2\gamma t \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x)) \\ &\qquad - t\beta \operatorname{Sum}(V\operatorname{Log}(x) + \operatorname{Log}(x)V) + \text{const}\Big). \end{aligned} \tag{83}$$

For $n = 3$ with unit norm, $\gamma \operatorname{Sum}^2(V) + \beta \operatorname{Sum}(V^2) = 1$, the t^2 terms cancel. Thus:

$$\begin{aligned} B(x) = &-\gamma \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x)) \\ &- \frac{\beta}{2} \operatorname{Sum}(V\operatorname{Log}(x) + \operatorname{Log}(x)V). \end{aligned} \tag{84}$$

Case $n \geq 4$**.** The squared distance is:

$$\begin{aligned} d^2 = &\alpha\Big(t^2 \operatorname{tr}(V^2) - t \operatorname{tr}(V\operatorname{Log}(x) + \operatorname{Log}(x)V) + \operatorname{tr}(\operatorname{Log}^2(x))\Big) \\ &+ \gamma\Big(t^2 \operatorname{Sum}^2(V) + \operatorname{Sum}^2(\operatorname{Log}(x)) - 2t \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x))\Big) \\ &+ \beta\Big(t^2 \operatorname{Sum}(V^2) - t \operatorname{Sum}(V\operatorname{Log}(x) + \operatorname{Log}(x)V) + \operatorname{Sum}(\operatorname{Log}^2(x))\Big). \end{aligned} \tag{85}$$

For the busemann function with unit norm $\alpha \operatorname{tr}(V^2) + \gamma \operatorname{Sum}^2(V) + \beta \operatorname{Sum}(V^2) = 1$:

$$\begin{aligned} B(x) &= \lim_{t\to\infty} \frac{d^2 - t^2}{2t} \\ &= -\frac{1}{2}\Big(\alpha \operatorname{tr}(V\operatorname{Log}(x) + \operatorname{Log}(x)V) \\ &\qquad + 2\gamma \operatorname{Sum}(V)\operatorname{Sum}(\operatorname{Log}(x)) \\ &\qquad + \beta \operatorname{Sum}(V\operatorname{Log}(x) + \operatorname{Log}(x)V)\Big). \end{aligned} \tag{86}$$

H.4 Algebraic expansion of the inherited LSM coordinate

The explicit busemann projection for the Log-Scaled Metric (LSM) across all dimensions is given by:

$$\Pi^v_{\mathrm{LSM}}(x) = \begin{cases} -\zeta \operatorname{tr}(\tilde{X}) \operatorname{tr}(V) & \text{if } d = 2, \\ -\delta \operatorname{tr}(\operatorname{Diag}(\tilde{X})\operatorname{Diag}(V)) - \zeta \operatorname{tr}(\tilde{X}) \operatorname{tr}(V) & \text{if } d = 3, \\ -\alpha \operatorname{tr}(\tilde{X}V) - \delta \operatorname{tr}(\operatorname{Diag}(\tilde{X})\operatorname{Diag}(V)) - \zeta \operatorname{tr}(\tilde{X}) \operatorname{tr}(V) & \text{if } d \geq 4. \end{cases} \tag{87}$$

From the exponential map at the identity, we have:

$$\operatorname{Exp}^*_{I_n}(tv) = \operatorname{Exp}^* \left(\operatorname{Log}^*(I_n) + d_{I_n} \operatorname{Log}^*(tv) \right) = \operatorname{Exp}^*(tv), \tag{88}$$

where

$$\operatorname{Log}^*(I_n) = \log(I_n) = 0, \tag{89}$$

and

$$d_{I_n} \operatorname{Log}^*(tv) = d_{I_n} \log(tv + \mathbf{x}_0) = tv, \tag{90}$$

with

$$\mathbf{x}_0 = -2 \operatorname{diag}\left(\frac{t}{2} v \mathbf{1} \right) = 0, \quad v \in \operatorname{Row}_0(n). \tag{91}$$

The squared distance can be expressed as:

$$d^*(x, \gamma(t))^2 = q^* \left(\log^*(x) - tv \right). \tag{92}$$

Let $\log^*(x) = \tilde{x}$. Then:

$$d^*(x, \gamma(t))^2 = q^* \left(\tilde{x} - tv \right). \tag{93}$$

Case $n = 2$**.** The squared distance is:

$$\begin{aligned} d^*(x, \gamma(t))^2 &= \zeta \operatorname{tr}((\tilde{x} - tV)^2) \\ &= \zeta \left(\operatorname{tr}(\tilde{x}) - t \operatorname{tr}(V) \right)^2 \\ &= \zeta \left((\operatorname{tr}(\tilde{x}))^2 + (t \operatorname{tr}(V))^2 - 2t \operatorname{tr}(\tilde{x}) \operatorname{tr}(V) \right). \end{aligned} \tag{94}$$

For the busemann function:

$$\begin{aligned} B(x) &= \lim_{t \to \infty} \frac{d^2 - t^2}{2t} \\ &= \lim_{t \to \infty} \frac{1}{2t} \Big(\zeta t^2 (\operatorname{tr}(V))^2 - t^2 - 2\zeta t \operatorname{tr}(\tilde{x}) \operatorname{tr}(V) + \zeta (\operatorname{tr}(\tilde{x}))^2 \Big). \end{aligned} \tag{95}$$

For $n = 2$ with unit norm, $\zeta(\operatorname{tr}(V))^2 = 1$, the t^2 terms cancel. Thus:

$$B(x) = -\zeta \operatorname{tr}(\tilde{x}) \operatorname{tr}(V). \tag{96}$$

Case $n = 3$**.** The squared distance is:

$$\begin{aligned} d^2 &= \delta \operatorname{tr}\left((\operatorname{Diag}(\tilde{x}) - t \operatorname{Diag}(V))^2 \right) + \zeta \operatorname{tr}((\tilde{x} - tV)^2) \\ &= \delta \left(\operatorname{tr}(\operatorname{Diag}(\tilde{x}))^2 + t^2 \operatorname{tr}(\operatorname{Diag}(V))^2 - 2t \operatorname{tr}(\operatorname{Diag}(\tilde{x}) \operatorname{Diag}(V)) \right) \\ &\quad + \zeta \left(\operatorname{tr}(\tilde{x}^2) + t^2 \operatorname{tr}(V^2) - 2t \operatorname{tr}(\tilde{x}) \operatorname{tr}(V) \right) \\ &= t^2 + \delta \operatorname{tr}(\operatorname{Diag}(\tilde{x}))^2 - 2t\delta \operatorname{tr}(\operatorname{Diag}(\tilde{x}) \operatorname{Diag}(V)) \\ &\quad + \zeta \operatorname{tr}(\tilde{x}^2) - 2t\zeta \operatorname{tr}(\tilde{x}) \operatorname{tr}(V). \end{aligned} \tag{97}$$

For the busemann function with unit norm $\delta \operatorname{tr}(\operatorname{Diag}(V))^2 + \zeta \operatorname{tr}(V^2) = 1$:

$$\begin{aligned} B(x) &= \lim_{t \to \infty} \frac{d^2 - t^2}{2t} \\ &= -\delta \operatorname{tr}(\operatorname{Diag}(\tilde{x}) \operatorname{Diag}(V)) - \zeta \operatorname{tr}(\tilde{x}) \operatorname{tr}(V). \end{aligned} \tag{98}$$

Case $n \geq 4$**.** Note that:

$$(\tilde{x} - tV)(\tilde{x} - tV) = \tilde{x}^2 - t\tilde{x}V - tV\tilde{x} + t^2V^2. \tag{99}$$

The squared distance is:

$$\begin{aligned} d^2 &= \alpha \operatorname{tr}((\tilde{x} - tV)^2) + \delta \left(\operatorname{tr}(\operatorname{Diag}(\tilde{x}))^2 + t^2 \operatorname{tr}(\operatorname{Diag}(V))^2 - 2t \operatorname{tr}(\operatorname{Diag}(\tilde{x}) \operatorname{Diag}(V))\right) \\ &\quad + \zeta \left(\operatorname{tr}(\tilde{x})^2 + t^2 \operatorname{tr}(V)^2 - 2t \operatorname{tr}(\tilde{x}) \operatorname{tr}(V)\right) \\ &= \alpha \operatorname{tr}(\tilde{x}^2) - \alpha t \operatorname{tr}(\tilde{x}V + V\tilde{x}) + t^2 \\ &\quad + \delta \operatorname{tr}(\operatorname{Diag}(\tilde{x}))^2 - 2t\delta \operatorname{tr}(\operatorname{Diag}(\tilde{x}) \operatorname{Diag}(V)) \\ &\quad + \zeta \operatorname{tr}(\tilde{x})^2 - 2t\zeta \operatorname{tr}(\tilde{x}) \operatorname{tr}(V). \end{aligned} \tag{100}$$

For the busemann function with unit norm $\alpha \operatorname{tr}(V^2) + \delta \operatorname{tr}(\operatorname{Diag}(V))^2 + \zeta \operatorname{tr}(V)^2 = 1$:

$$\begin{aligned} B(x) &= \lim_{t \to \infty} \frac{d^2 - t^2}{2t} \\ &= -\frac{1}{2}\Big(\alpha \operatorname{tr}(\tilde{x}V + V\tilde{x}) + 2\delta \operatorname{tr}(\operatorname{Diag}(\tilde{x}) \operatorname{Diag}(V)) \\ &\quad + 2\zeta \operatorname{tr}(\tilde{x}) \operatorname{tr}(V)\Big). \end{aligned} \tag{101}$$

H.5 Derivation of the flat distance-preservation corollary

Corollary H.1 (Distance preservation under the flat pullback)**.** *Let* $(\mathrm{Cor}^+(d), d_\phi)$ *be globally isometric through* ϕ *to a Euclidean model space* $(V, \langle\cdot,\cdot\rangle_V)$*. Let* $\gamma : \mathbb{R} \to \mathrm{Cor}^+(d)$ *be a unit-speed geodesic line such that* $\phi(\gamma(t)) = \phi(\gamma(0)) + tu$ *for some* $\|u\|_V = 1$*, let* $p = \gamma(+\infty)$*, and normalize the associated Busemann function by* $B_p(\gamma(t)) = -t$*. If* $\pi_{\gamma,p}(x)$ *is the intersection of* $\gamma(\mathbb{R})$ *with the horosphere through* x*, then, for all* $x, y \in \mathrm{Cor}^+(d)$*,*

$$d_\phi\big(\pi_{\gamma,p}(x), \pi_{\gamma,p}(y)\big) = \big|\langle \phi(y) - \phi(x), u\rangle_V\big| \leq d_\phi(x, y). \tag{102}$$

Equality holds whenever $\phi(y) - \phi(x) \in \operatorname{span}(u)$*, equivalently when* x *and* y *lie on a common geodesic line with forward endpoint* p*.*

Proof. Let ϕ: $(\mathrm{Cor}^+(n), g) \to (V, \langle\cdot,\cdot\rangle_V)$ be a global isometry onto a Euclidean model space V. Fix $o \in \mathrm{Cor}^+(n)$ and let $\gamma : \mathbb{R} \to \mathrm{Cor}^+(n)$ be the unit-speed geodesic line with direction $u \in V$, i.e.,

$$\phi(\gamma(t)) = \phi(o) + tu, \quad \|u\|_V = 1.$$

For the Busemann function associated with the ideal endpoint p of γ, we may choose the additive constant so that

$$B_p(x) = -\langle \phi(x) - \phi(o), u\rangle_V.$$

Hence the horosphere through x is mapped to the affine hyperplane

$$H_x = \{z \in V : \langle z - \phi(o), u\rangle_V = \langle \phi(x) - \phi(o), u\rangle_V\}.$$

By definition of the horospherical (Busemann) projection, $\phi(\pi_{\gamma,p}(x))$ is the unique intersection of the line $\phi(\gamma)$ with H_x, which yields the explicit formula

$$\phi(\pi_{\gamma,p}(x)) = \phi(o) + \langle \phi(x) - \phi(o), u\rangle_V u. \tag{103}$$

For arbitrary $x, y \in \mathrm{Cor}^+(n)$, Eq. (103) and Cauchy–Schwarz give

$$\begin{aligned} d(\pi_{\gamma,p}(x), \pi_{\gamma,p}(y)) &= |\langle \phi(y) - \phi(x), u\rangle_V| && (104) \\ &\leq \|\phi(y) - \phi(x)\|_V = d(x, y). && (105) \end{aligned}$$

This is the flat specialization of the general 1-Lipschitz property in Bonet et al. (2025, Lem. 38).

Now assume x and y lie on a common geodesic line ξ with forward endpoint p. Since ϕ is an isometry and geodesics map to straight lines, $\phi(\xi)$ is a line with the same direction u as $\phi(\gamma)$. Therefore there exists $\alpha \in \mathbb{R}$ such that

$$\phi(y) - \phi(x) = \alpha u.$$

Using Eq. (103) we obtain

$$\begin{aligned}\phi(\pi_{\gamma,p}(y)) - \phi(\pi_{\gamma,p}(x)) &= \left[\langle \phi(y) - \phi(o), u\rangle_V - \langle \phi(x) - \phi(o), u\rangle_V\right] u \\ &= \langle \phi(y) - \phi(x), u\rangle_V u \\ &= \alpha u,\end{aligned}$$

and thus

$$\|\phi(\pi_{\gamma,p}(y)) - \phi(\pi_{\gamma,p}(x))\|_V = |\alpha| = \|\phi(y) - \phi(x)\|_V.$$

Finally, since ϕ is an isometry,

$$d(\pi_{\gamma,p}(x), \pi_{\gamma,p}(y)) = d(x, y).$$

This proves the equality case in Eq. (102), and strict contraction is possible outside this case. □

H.6 Metricity of the flat CorSW family

We work on the manifold of full-rank correlation matrices Cor_n^+ endowed with one of the pullback Riemannian metrics $g \in \{\mathrm{ECM}, \mathrm{LECM}, \mathrm{OLM}, \mathrm{LSM}\}$ introduced in Thanwerdas & Pennec (2022; 2024b). For each such metric g, there exists a smooth injective map

$$\Phi_g : (\mathrm{Cor}_n^+, d_g) \to (E_g, \|\cdot\|)$$

into a Euclidean space $(E_g, \|\cdot\|)$ such that d_g is the pullback of the Euclidean distance, i.e.

$$d_g(C_1, C_2) = \left\|\Phi_g(C_1) - \Phi_g(C_2)\right\| \quad \forall C_1, C_2 \in \mathrm{Cor}_n^+. \tag{106}$$

Flatness and the relevant correlation embeddings are prior geometric results Thanwerdas & Pennec (2022; 2024b).

Let $X_g := (\mathrm{Cor}_n^+, d_g)$ and $\Phi := \Phi_g$. We denote by $\mathcal{P}_p(X_g)$ the set of Borel probability measures on X_g with finite p-th moment with respect to d_g. By Eq. (106) and the finite-dimensionality of X_g, we have

$$\mu \in \mathcal{P}_p(X_g) \quad \Longleftrightarrow \quad \Phi_{\#}\mu \in \mathcal{P}_p(E_g),$$

where $\Phi_{\#}\mu$ is the pushforward measure on E_g.

Definition of CorSW_p For $\mu, \nu \in \mathcal{P}_p(X_g)$, we define the correlation sliced-Wasserstein distance by

$$\mathrm{CorSW}_{p,g}^p(\mu, \nu) := \int_{\mathbb{S}^{D-1}} W_p^p\big((\theta^\top \Phi(\cdot))_{\#}\mu, (\theta^\top \Phi(\cdot))_{\#}\nu\big)\mathrm{d}\theta, \tag{107}$$

where $D = \dim(E_g)$ and W_p is the one-dimensional p-Wasserstein distance on $\mathbb{R}$.

Noting that $(\theta^\top \Phi(\cdot))_{\#}\mu = (\pi_\theta)_{\#}(\Phi_{\#}\mu)$ with $\pi_\theta(x) = \langle \theta, x\rangle$, we can rewrite Eq. (107) as

$$\mathrm{CorSW}_{p,g}(\mu, \nu) = \mathrm{SW}_p\big(\Phi_{\#}\mu, \Phi_{\#}\nu\big), \tag{108}$$

where SW_p is the standard Euclidean sliced-Wasserstein distance. Applying the pullback identity of Bonet et al. (2025, Lem. 25) to the four correlation embeddings yields this common reduction.

By Bonet et al. (2025, Prop. 26), the right-hand side is a finite metric. Injectivity of Φ transfers identity of indiscernibles back to X_g, while non-negativity, symmetry, and the triangle inequality are inherited directly. This proves Cor. 4.2 for all four members of the flat CorSW family. The PHCML family has its own route-specific reductions, established separately in Prop. 4.3. □

H.7 GEODESIC COORDINATE OF THE FLAT CORSW FAMILY

Corollary H.2 (Geodesic–Busemann sign relation). *Let ϕ be a diffeomorphism from $\mathrm{Cor}^+(d)$ to a Euclidean model space $(V, \langle\cdot,\cdot\rangle_V)$ inducing a pullback metric. For $v \in T_{I_d}\mathrm{Cor}^+(d)$ with $\|v\|_{I_d} = 1$, the geodesic projection coordinate satisfies*

$$P^v(x) = \big\langle \phi(x) - \phi(I_d), \phi_{*,I_d}(v)\big\rangle_V = -B^v(x) = -\Pi^v_\phi(x). \tag{109}$$

Consequently, with $\widetilde{P}^v(x) = \exp_{I_d}(P^v(x)v)$,

$$P^v(x) = \mathrm{sgn}\big(\langle \phi(x) - \phi(I_d), \phi_{*,I_d}(v)\rangle_V\big)\, d_\phi\big(\widetilde{P}^v(x), I_d\big). \tag{110}$$

Proof. Let $x \in \mathcal{M}$. Denote $f : \mathbb{R} \to \mathbb{R}$ such that

$$\begin{aligned} f(t) &= d_{\mathcal{M}}(\gamma(t), x)^2 \\ &= d_{\mathcal{M}}(\phi^{-1}(\phi(I_d) + t\phi_{*,I_d}(v)), x)^2 \\ &= \|\phi(\phi^{-1}(\phi(I_d) + t\phi_{*,I_d}(v))) - \phi(x)\|_V^2 \\ &= t^2\|\phi_{*,I_d}(v)\|_V^2 - 2t\langle \phi(x) - \phi(I_d), \phi_{*,I_d}(v)\rangle_V + \|\phi(I_d) - \phi(x)\|_V^2 \\ &= t^2 - 2t\langle \phi(x) - \phi(I_d), \phi_{*,I_d}(v)\rangle_V + \|\phi(I_d) - \phi(x)\|_V^2. \end{aligned}$$

Using in the last line that $\|\phi_{*,I_d}(v)\|_V^2 = 1$ since $v \in S_{I_d}$, we have

$$f'(t) = 0 \iff t = \langle \phi(x) - \phi(I_d), \phi_{*,I_d}(v)\rangle_V.$$

Therefore,

$$P^v(x) = \arg\min_{t\in\mathbb{R}} f(t) = \langle \phi(x) - \phi(I_d), \phi_{*,I_d}(v)\rangle_V.$$

By Bonet et al. (2025, Prop. 7), $B^v(x) = -\langle \phi_{*,I_d}(v), \phi(x) - \phi(I_d)\rangle_V$. With our convention $\Pi^v_\phi = B^v$, we obtain $P^v(x) = \langle \phi(x) - \phi(I_d), \phi_{*,I_d}(v)\rangle_V = -B^v(x) = -\Pi^v_\phi(x)$. Because v is unit norm, $t \mapsto \exp_{I_d}(tv)$ is unit speed and therefore $d_\phi(\widetilde{P}^v(x), I_d) = |P^v(x)|$. Combining this identity with the sign of the inner product proves Eq. (110). □

H.8 PRODUCT-SPACE SLICING ON PHCM

The product-hyperbolic Cholesky metric represents $C \in \mathrm{Cor}^+(d)$ by normalized Cholesky rows. Let $L = \mathrm{Chol}(C)$ and set

$$r_i = \frac{L_{i,1:i-1}}{L_{ii}}, \qquad \ell_{i,0} = \sqrt{1 + \|r_i\|_2^2}, \qquad x_i = (\ell_{i,0}, r_i) \in \mathbb{H}^{i-1}. \tag{111}$$

The representation is the product Cartan–Hadamard manifold

$$\mathcal{M}_{\mathrm{PHCM}} = \prod_{i=2}^{d} \mathbb{H}^{i-1}, \tag{112}$$

whose product distance is

$$d_{\mathrm{Prod}}(x, y)^2 = \sum_{i=2}^{d} d_{\mathbb{H}}(x_i, y_i)^2. \tag{113}$$

Each factor has curvature -1, while tangent planes mixing different factors have curvature zero (Thanwerdas & Pennec, 2022).

GLOBAL PRODUCT DIRECTIONS

A unit tangent direction decomposes as

$$u = (u_2, \ldots, u_d), \qquad u_i \in \mathbb{R}^{i-1}, \qquad \sum_{i=2}^{d} \|u_i\|_2^2 = 1. \tag{114}$$

Let $a_i = \|u_i\|_2$. For $a_i > 0$, define $v_i = u_i/a_i$. The product geodesic ray is

$$\gamma_u(t) = (\gamma_{v_2}(a_2 t), \ldots, \gamma_{v_d}(a_d t)), \qquad \gamma_{v_i}(s) = (\cosh(s), \sinh(s)v_i). \tag{115}$$

Thus, one global direction controls all factors, and a_i is the speed assigned to factor i.

Proposition H.3 (PHCM product coordinates). *Define*

$$d_i = \operatorname{asinh}\left(\|r_i\|_2\right), \qquad \Xi_i = \begin{cases} d_i r_i/\|r_i\|_2, & r_i \neq 0, \\ 0, & r_i = 0. \end{cases} \tag{116}$$

Then the product Log and Busemann coordinates are

$$\begin{aligned} s_u^{\text{Prod-Log}}(C) &= \sum_{i=2}^{d} \langle \Xi_i, u_i \rangle, \\ s_u^{\text{Prod-B}}(C) &= \sum_{i:a_i>0} a_i \log\left(\ell_{i,0} - \langle r_i, v_i \rangle\right). \end{aligned} \tag{117}$$

The second expression is the exact Busemann function associated with the unit-speed product ray γ_u. The Product-Geodesic (approx.) coordinate used in our experiments is

$$s_u^{\text{Prod-G}}(C) = \sum_{i:a_i>0} a_i \operatorname{arctanh}\left(\frac{\langle r_i, v_i \rangle}{\ell_{i,0}}\right). \tag{118}$$

It is a closed-form speed-weighted combination of exact factor geodesic coordinates. For general product points, it differs from the nearest-point coordinate on the shared product geodesic, and their precise relationship is derived below.

Proof. The logarithmic map of x_i at the hyperbolic origin is Ξ_i. Since the product tangent space is the orthogonal direct sum of the factor tangent spaces, its projection onto u is $\sum_i \langle \Xi_i, u_i \rangle$.

For Product-Busemann, write

$$b_{v_i}(x_i) = \log\left(\ell_{i,0} - \langle r_i, v_i \rangle\right). \tag{119}$$

For every $a_i > 0$, the definition of the factor Busemann function gives

$$d_{\mathbb{H}}\left(x_i, \gamma_{v_i}(a_i t)\right) = a_i t + b_{v_i}(x_i) + o(1) \tag{120}$$

as $t \to \infty$. Factors with $a_i = 0$ contribute only bounded terms. Therefore,

$$d_{\text{Prod}}\left(x, \gamma_u(t)\right) = \left(\sum_{i=2}^{d} d_{\mathbb{H}}\left(x_i, \gamma_{v_i}(a_i t)\right)^2\right)^{1/2} \tag{121}$$

$$= \left(t^2 + 2t \sum_{i:a_i>0} a_i b_{v_i}(x_i) + \mathcal{O}(1)\right)^{1/2} \tag{122}$$

$$= t + \sum_{i:a_i>0} a_i b_{v_i}(x_i) + o(1), \tag{123}$$

where $\sum_i a_i^2 = 1$ is used in the second line. Subtracting t and taking the limit proves the Product-Busemann formula.

The signed geodesic coordinate of factor i is

$$c_i = \operatorname{arctanh}\left(\frac{\langle r_i, v_i \rangle}{\ell_{i,0}}\right). \tag{124}$$

Equation (118) is the speed-weighted aggregation of these factor coordinates. □

AGGREGATION BEFORE ONE-DIMENSIONAL WASSERSTEIN MATCHING

For a fixed global direction u, Product-Busemann first forms one scalar per sample. For empirical samples x_b and y_b, define

$$p_{b,i} = b_{v_i}(x_{b,i}), \qquad q_{b,i} = b_{v_i}(y_{b,i}), \qquad S_b = \sum_i a_i p_{b,i}, \qquad T_b = \sum_i a_i q_{b,i}. \tag{125}$$

The sliced loss for this direction is one Wasserstein distance,

$$W_2^2\left(\{S_b\}_{b=1}^B, \{T_b\}_{b=1}^B\right) = \min_{\pi} \frac{1}{B}\sum_{b=1}^{B}\left(\sum_i a_i\left(p_{b,i} - q_{\pi(b),i}\right)\right)^2. \tag{126}$$

The permutation π is shared by all factors because sorting is performed after scalar aggregation. In contrast, a factor-wise loss takes the form

$$\sum_i a_i^2 \min_{\pi_i} \frac{1}{B}\sum_{b=1}^{B}\left(p_{b,i} - q_{\pi_i(b),i}\right)^2, \tag{127}$$

where each factor has an independent matching π_i. Equations (126) and (127) are generally unequal, since sorting does not commute with summation.

A two-factor example makes the distinction explicit. Let $a_1 = a_2 = 1/\sqrt{2}$, with source factor coordinates $(0,1)$ and $(1,0)$ and target coordinates $(0,0)$ and $(1,1)$. Both factor marginals are $\{0,1\}$, so the factor-wise loss is zero. The aggregated source scores are $\{1/\sqrt{2}, 1/\sqrt{2}\}$, whereas the target scores are $\{0, \sqrt{2}\}$. Their one-dimensional Wasserstein loss is $1/2$. Product-Busemann therefore retains cross-factor dependence that independent marginal matching cannot detect.

Exact product geodesic projection

For each active factor, define

$$b_i = \langle r_i, v_i\rangle, \qquad c_i = \operatorname{arctanh}\left(\frac{b_i}{\ell_{i,0}}\right), \qquad m_i = \sqrt{\ell_{i,0}^2 - b_i^2}. \tag{128}$$

Since $\ell_{i,0} = m_i\cosh(c_i)$ and $b_i = m_i\sinh(c_i)$, the hyperbolic distance from x_i to the factor geodesic satisfies

$$d_{\mathbb{H}}\left(x_i, \gamma_{v_i}(a_i t)\right) = \operatorname{arccosh}\left(\ell_{i,0}\cosh(a_i t) - b_i\sinh(a_i t)\right) \tag{129}$$

$$= \operatorname{arccosh}\left(m_i\cosh\left(a_i t - c_i\right)\right). \tag{130}$$

Consequently, the exact nearest-geodesic coordinate is

$$t_u^\star(C) = \arg\min_{t\in\mathbb{R}} F_u(t), \qquad F_u(t) = \sum_{i:a_i>0} \operatorname{arccosh}^2\left(m_i\cosh\left(a_i t - c_i\right)\right). \tag{131}$$

This is a one-dimensional strictly convex problem because it is the squared distance to a geodesic on a Hadamard product manifold. Writing

$$z_i(t) = m_i\cosh\left(a_i t - c_i\right), \tag{132}$$

its first-order condition is

$$F_u'(t) = 2\sum_{i:a_i>0} a_i \operatorname{arccosh}\left(z_i(t)\right)\frac{m_i\sinh\left(a_i t - c_i\right)}{\sqrt{z_i(t)^2 - 1}} = 0, \tag{133}$$

where the ratio is interpreted continuously as one when $z_i(t) = 1$. Every summand has the sign of $a_i t - c_i$. Hence the unique solution lies in

$$\min_{i:a_i>0}\frac{c_i}{a_i} \le t_u^\star(C) \le \max_{i:a_i>0}\frac{c_i}{a_i}, \tag{134}$$

which provides a stable interval for bisection.

Product-Geodesic (approx.) retains the exact factor coordinates c_i and combines them according to the factor speeds:

$$s_u^{\text{Prod-G}}(C) = \sum_{i:a_i>0} a_i c_i. \tag{135}$$

The relationship to the exact coordinate is particularly transparent when $m_i = 1$ for every active factor. This condition is equivalent to r_i being parallel to v_i, so each factor lies on its corresponding slicing geodesic. In this case,

$$F_u(t) = \sum_{i:a_i>0}\left(a_i t - c_i\right)^2, \tag{136}$$

and $\sum_i a_i^2 = 1$ gives

$$t_u^\star(C) = \sum_{i:a_i>0} a_i c_i = s_u^{\text{Prod-G}}(C). \tag{137}$$

Thus, Product-Geodesic (approx.) equals the exact nearest-point coordinate on factor-aligned product points and remains closed form for general product points. The nearest-point coordinate $t_u^\star$ can be evaluated through the one-dimensional solve in Eqs. (133) and (134), whereas the reported Prod-G (approx.) experiments use Eq. (118), avoiding a separate numerical optimization for every sample and slicing direction.

MOMENT CLASSES AND METRICITY

Let $o = (o_2, \dots, o_d)$ be the product origin and define

$$\mathcal{P}_p^{\text{Prod}} = \left\{ \mu \mid \int d_{\text{Prod}}\,(o, x(C))^p \,\mathrm{d}\mu(C) < \infty \right\}. \tag{138}$$

The product logarithmic coordinate $C \mapsto \text{Concat}(\Xi_2, \dots, \Xi_d)$ is a global diffeomorphism onto $\mathbb{R}^D$. Hence Product-Log reduces to Euclidean SW and is a finite metric on $\mathcal{P}_p^{\text{Prod}}$. Product-Busemann is finite because its normalized Busemann function is 1-Lipschitz and vanishes at the origin. It is a sliced pseudometric in general and becomes a metric on measure classes for which the product horospherical Radon transform is injective up to null directions. Product-Geodesic (approx.) is finite on the same moment class and defines a sliced training discrepancy. We do not claim identity of indiscernibles or intrinsic geodesic-slicing metricity for Product-Geodesic (approx.).

H.9 PROPERTIES OF THE PHCML SLICED DISTANCES

For $r \in \{\text{Log}, \text{G}, \text{B}\}$, let $\mu_u^r = (s_u^r)_\# \mu$ and $\nu_u^r = (s_u^r)_\# \nu$. The three PHCML objectives are

$$\mathcal{C}_{\text{Log}}^p = \int_{S^{D-1}} W_p^p(\mu_u^{\text{Log}}, \nu_u^{\text{Log}}) \mathrm{d}\sigma(u), \quad \mathcal{C}_{\text{G}}^p = \int_{S^{D-1}} W_p^p(\mu_u^{\text{G}}, \nu_u^{\text{G}}) \mathrm{d}\sigma(u), \quad \mathcal{C}_{\text{B}}^p = \int_{S^{D-1}} W_p^p(\mu_u^{\text{B}}, \nu_u^{\text{B}}) \mathrm{d}\sigma(u). \tag{139}$$

Proof. We first verify the global coordinates used in the proposition. Given arbitrary row coordinates $r_i \in \mathbb{R}^{i-1}$, define

$$L_{ii} = \left(1 + \|r_i\|_2^2\right)^{-1/2}, \qquad L_{i,1:i-1} = L_{ii} r_i, \tag{140}$$

with $L_{11} = 1$. Every row of the resulting lower-triangular matrix L has unit Euclidean norm and positive diagonal, so $C = LL^\top$ is the unique matrix in $\text{Cor}^+(d)$ with these coordinates. The construction and its inverse are smooth. Hence $C \mapsto z(C)$ is a global diffeomorphism onto $\mathbb{R}^D$. The graph map $z \mapsto (\sqrt{1 + \|z\|_2^2}, z)$ is a global diffeomorphism from $\mathbb{R}^D$ onto $\mathbb{H}_{-1}^D$, proving the claim for Ψ. Since the hyperbolic logarithmic map at $o_{\mathbb{H}}$ is a global diffeomorphism onto $T_{o_{\mathbb{H}}}\mathbb{H}_{-1}^D \simeq \mathbb{R}^D$, the composition $\Xi = \text{Log}_{o_{\mathbb{H}}}^{\mathbb{H}} \circ \Psi$ is also a global diffeomorphism. Moreover, $\|\Xi(C)\|_2 = d_{\mathbb{H}}(o_{\mathbb{H}}, \Psi(C))$. Thus, for $M = \Psi_\# \mu$,

$$\int_{\mathbb{H}_{-1}^D} d_{\mathbb{H}}(o_{\mathbb{H}}, x)^p \,\mathrm{d}M(x) = \int_{\text{Cor}^+(d)} \|\Xi(C)\|_2^p \,\mathrm{d}\mu(C) < \infty, \tag{141}$$

and likewise for $N = \Psi_\# \nu$. Hence $M, N \in \mathcal{P}_p(\mathbb{H}_{-1}^D)$, while $\Xi_\# \mu, \Xi_\# \nu \in \mathcal{P}_p(\mathbb{R}^D)$.

For PHCML-Log, substituting $s_u^{\text{Log}}(C) = \langle \Xi(C), u \rangle$ into Eq. (16) yields

$$\text{CorSW}_{p,\text{PHCML-Log}}(\mu, \nu) = \text{SW}_p(\Xi_\# \mu, \Xi_\# \nu). \tag{142}$$

Euclidean sliced-Wasserstein is a finite metric on $\mathcal{P}_p(\mathbb{R}^D)$, and injectivity of Ξ transfers identity of indiscernibles to the original correlation measures. This proves the first item.

The intrinsic formulas in Eq. (13) are exactly the Lorentz horospherical and signed geodesic coordinates for curvature -1 from Bonet et al. (2025, Prop. 9), and the latter projects onto the complete line $\gamma_u(\mathbb{R})$. Substitution into Eq. (139) gives the HHSW and GHSW identities in Prop. 4.3. Their finiteness, symmetry, and triangle inequalities follow from Bonet et al. (2025, Prop. 17).

Write $\pi_u^r : \mathbb{H}^D_{-1} \to \mathbb{R}$ for the corresponding hyperbolic coordinate, so $s_u^r = \pi_u^r \circ \Psi$. Both π_u^{B} and π_u^{G} are 1-Lipschitz and vanish at $o_{\mathbb{H}}$. Their pushforwards therefore have finite p-moments, and Wasserstein contraction gives, for every $u \in S^{D-1}$,

$$W_p((\pi_u^r)_\# M, (\pi_u^r)_\# N) \le W_{p,\mathbb{H}}(M, N), \qquad r \in \{\mathrm{B}, \mathrm{G}\}. \tag{143}$$

Raising this inequality to the p-th power, integrating over u, and taking the p-th root yields the two Wasserstein upper bounds in Prop. 4.3, equivalently the bound of Bonet et al. (2025, Prop. 22).

The PHCML-G lower bound and metricity are established independently in Lem. H.4 below. PHCML-Busemann is finite by the argument above. On a measure subclass for which equality of $(\pi_u^{\mathrm{B}})_\# M$ and $(\pi_u^{\mathrm{B}})_\# N$ for σ_{D-1}-almost every u implies $M = N$, the horospherical Radon-transform argument of Bonet et al. (2025, Sec. 5.1) gives identity of indiscernibles and hence metricity. □

Lemma H.4 (Klein lower bound and metricity of PHCML-G)**.** *Let $p \ge 1$ and $\mu, \nu \in \mathcal{P}_p^\Psi$. With $T = \kappa \circ \Psi : \mathrm{Cor}^+(d) \to \mathbb{B}^D$,*

$$\mathrm{SW}_p(T_\# \mu, T_\# \nu) \le \mathrm{CorSW}_{p,\text{PHCML-G}}(\mu, \nu). \tag{144}$$

Consequently, the population PHCML-G distance is a finite metric on $\mathcal{P}_p^\Psi$.

Proof. Fix $C \in \mathrm{Cor}^+(d)$, write $x = \Psi(C) = (x_0, \bar{x})$, and let $u \in S^{D-1}$. Along the complete geodesic $\gamma_u(t) = (\cosh t, \sinh t\, u)$,

$$\cosh d_{\mathbb{H}}(x, \gamma_u(t)) = x_0 \cosh t - \langle \bar{x}, u \rangle \sinh t. \tag{145}$$

Since $\arccosh$ is increasing, the projection coordinate minimizes the right-hand side. Its derivative vanishes exactly when

$$x_0 \sinh t - \langle \bar{x}, u \rangle \cosh t = 0, \qquad \tanh t = \frac{\langle \bar{x}, u \rangle}{x_0} = \langle T(C), u \rangle. \tag{146}$$

Thus the signed geodesic coordinate satisfies

$$s_u^{\mathrm{G}}(C) = \operatorname{arctanh}\langle T(C), u \rangle, \qquad \langle T(C), u \rangle = \tanh\big(s_u^{\mathrm{G}}(C)\big). \tag{147}$$

Moreover, $\|T(C)\|_2 = \tanh d_{\mathbb{H}}(o_{\mathbb{H}}, x)$. Monotonicity of $\operatorname{arctanh}$ therefore gives

$$|s_u^{\mathrm{G}}(C)| \le d_{\mathbb{H}}(o_{\mathbb{H}}, \Psi(C)) = \|\Xi(C)\|_2. \tag{148}$$

Hence the defining moment condition makes every geodesic pushforward a member of $\mathcal{P}_p(\mathbb{R})$, and the Klein pushforwards have finite moments automatically because $T(C) \in \mathbb{B}^D$.

The map $\tanh : \mathbb{R} \to (-1, 1)$ is 1-Lipschitz. For any coupling of two one-dimensional measures, its pushforward by $(a, b) \mapsto (\tanh a, \tanh b)$ has no larger p-transport cost. Taking the infimum over couplings and applying Eq. (147) yields, for every u,

$$W_p^p((\langle T(\cdot), u \rangle)_\# \mu, (\langle T(\cdot), u \rangle)_\# \nu) \le W_p^p\big((s_u^{\mathrm{G}})_\# \mu, (s_u^{\mathrm{G}})_\# \nu\big)\,. \tag{149}$$

Integrating both sides with respect to the same normalized law σ_{D-1} gives

$$\mathrm{SW}_p^p(T_\# \mu, T_\# \nu) = \int_{S^{D-1}} W_p^p((\langle T(\cdot), u \rangle)_\# \mu, (\langle T(\cdot), u \rangle)_\# \nu)\ \mathrm{d}\sigma_{D-1}(u) \tag{150}$$

$$\le \int_{S^{D-1}} W_p^p\big((s_u^{\mathrm{G}})_\# \mu, (s_u^{\mathrm{G}})_\# \nu\big)\ \mathrm{d}\sigma_{D-1}(u) = \mathcal{C}_{\mathrm{G}}^p. \tag{151}$$

Taking the p-th root proves the claimed lower bound with constant one.

The remaining metric axioms follow directionwise. Define $d_u(\mu, \nu) = W_p((s_u^{\mathrm{G}})_\# \mu, (s_u^{\mathrm{G}})_\# \nu)$. The one-dimensional Wasserstein triangle inequality gives $d_u(\mu, \eta) \le d_u(\mu, \nu) + d_u(\nu, \eta)$, and Minkowski's inequality yields

$$\|d_u(\mu, \eta)\|_{L^p(\sigma_{D-1})} \le \|d_u(\mu, \nu)\|_{L^p(\sigma_{D-1})} + \|d_u(\nu, \eta)\|_{L^p(\sigma_{D-1})}. \tag{152}$$

Non-negativity and symmetry also hold directionwise, while finiteness follows from Eq. (148). Finally, if $\mathcal{C}_{\mathrm{G}} = 0$, the lower bound and metricity of Euclidean SW imply $T_\# \mu = T_\# \nu$. Both $\Psi : \mathrm{Cor}^+(d) \to \mathbb{H}^D_{-1}$ and $\kappa : \mathbb{H}^D_{-1} \to \mathbb{B}^D$ are diffeomorphisms, so T is injective and $\mu = \nu$. Together with the sliced triangle inequality, this proves metricity. □