Riemannian Flow Matching for Brain Connectivity Matrices via Pullback Geometry

Antoine Collas, **Ce Ju**, **Nicolas Salvy**, **Bertrand Thirion**

Inria, CEA, University Paris-Saclay

Code:

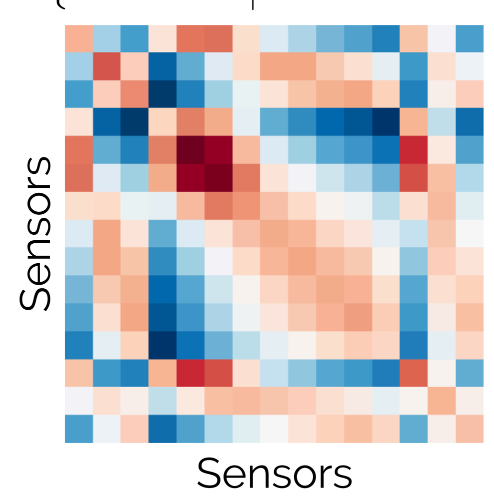
Brain Connectivity & Manifolds

Modern neuroimaging analyzes signals via connectivity matrices:

EEG Spatial Covariance

Symmetric Positive Definite (SPD) matrix

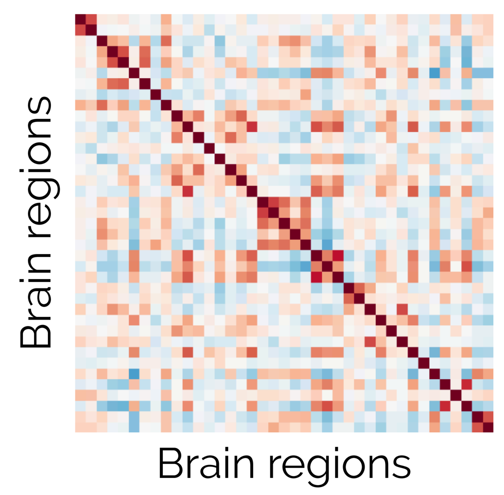
$\mathbb{S}_d^{++} = \left\{ \boldsymbol{\Sigma} \in \mathbb{R}^{d \times d} \,\middle|\, \boldsymbol{\Sigma}^\top = \boldsymbol{\Sigma}, \boldsymbol{\Sigma} \succ 0 \right\}$



fMRI Functional Connectivity

Correlation matrix

$\mathrm{Corr}_d = \{ \boldsymbol{\Sigma} \in \mathbb{S}_d^{++} \,|\, \mathrm{diag}(\boldsymbol{\Sigma}) = \mathbf{1} \}$



Challenges of brain connectivity generation:

1. **Euclidean methods** (e.g., standard Flow Matching) generate non-SPD matrices.
2. **Riemannian methods** require expensive geodesics and Riemannian norms, making training slow ($10\times$ slower).

Contribution:
We propose `DiffeoCFM` that enables **Euclidean** Flow Matching on matrix manifolds with **pullback metrics** induced by global diffeomorphisms.

- **Fast:** Train and sample in Euclidean space.
- **Exact:** Equivalent to Riemannian Flow Matching on the pullback manifold.
- **Valid:** Guarantees manifold constraints by construction.

Riemannian Flow Matching

Learn a vector field $u_\theta^{\mathcal{M}}$ on the manifold $\mathcal{M}$ to match geodesic velocities [1].

Training: Requires computing geodesics $\gamma(t)$ and Riemannian norms $\|\cdot\|_{\gamma(t)}$:

$$\mathcal{L}(\theta) \triangleq \mathbb{E}_{t,x_0,x_1} \left\| u_\theta^{\mathcal{M}}(t, \gamma(t)) - \dot{\gamma}(t) \right\|_{\gamma(t)}^2$$

Sampling: Requires solving Riemannian ODEs on $\mathcal{M}$

$$\dot{x}(t) = u_\theta^{\mathcal{M}}(t, x(t)), \quad x(0) = x_0$$

Pullback Manifold

$\phi : \mathcal{M} \to E$ a **global diffeomorphism** to a Euclidean space E,

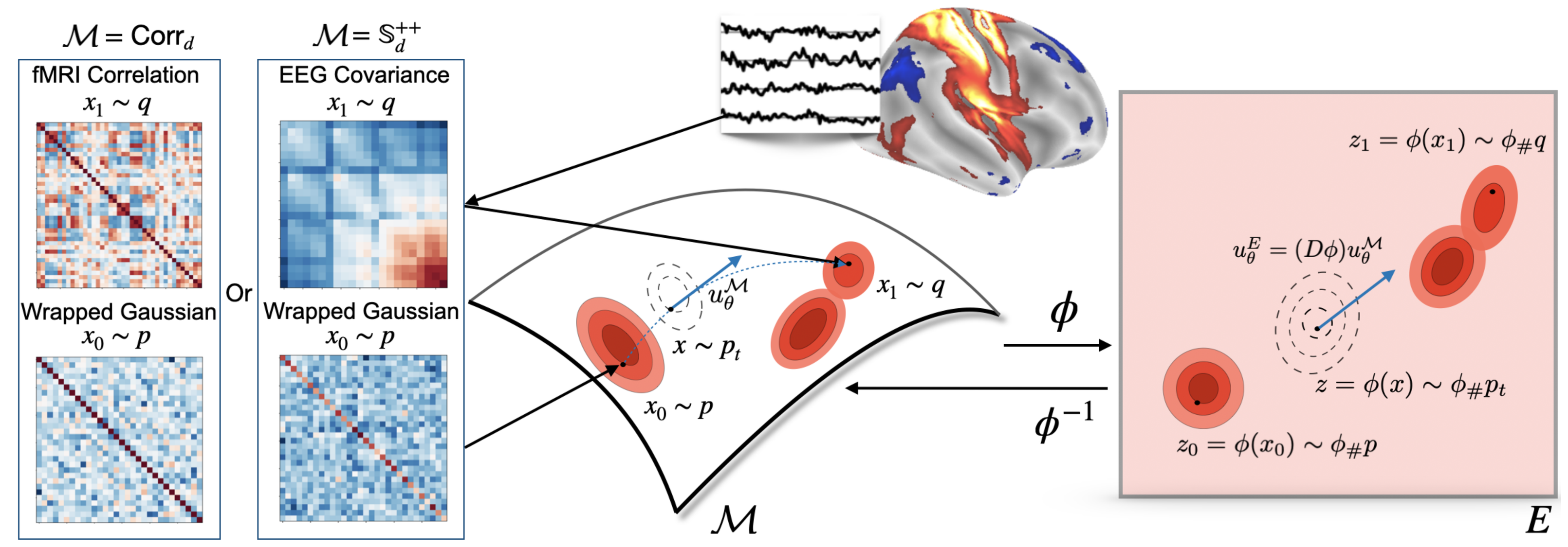
$$g_x(\xi, \eta) \triangleq \langle \mathrm{D}\,\phi(x)[\xi], \mathrm{D}\,\phi(x)[\eta] \rangle, \quad \xi, \eta \in T_x\mathcal{M}$$

Key Insight: we can compute everything in E and map back to $\mathcal{M}$ via ϕ:

(Geodesic) $\gamma(t) = \phi^{-1}\left((1-t)\phi(x_0) + t\phi(x_1)\right)$

(Distance) $d_{\mathcal{M}}(x_0, x_1) = \|\phi(x_0) - \phi(x_1)\|_E$

`DiffeoCFM`: Overview



Assumption: There exists a global diffeomorphism $\phi : \mathcal{M} \to E$ to a Euclidean space E.

- **Training:** Map data to E via ϕ and learn a Euclidean vector field u_θ^E to match straight-line velocities.
- **Sampling:** Solve Euclidean ODE $\dot{z}(t) = u_\theta^E(t, z)$ in E and map back to $\mathcal{M}$ via ϕ^{-1}.

Theoretical Guarantees

We prove that training in Euclidean space is **exactly equivalent** to Riemannian Flow Matching on $(\mathcal{M}, g)$ by defining the pulled-back vector field:

$$u_\theta^E(t, z) \triangleq \mathrm{D}\,\phi(\phi^{-1}(z)) \left[u_\theta^{\mathcal{M}}(t, \phi^{-1}(z)) \right]$$

Proposition 1 (Training Equivalence): The Riemannian loss simplifies to the standard Euclidean Flow Matching loss on transformed data $z = \phi(x)$:

$$\mathcal{L}(\theta) = \mathbb{E}_{t,\,z_0,\,z_1} \underbrace{\left\| u_\theta^E\left(t, (1-t)z_0 + tz_1\right) - (z_1 - z_0) \right\|_E^2}_{\text{Standard Euclidean Flow Matching Loss}}$$

Proposition 2 (Sampling Equivalence): Let $z(t)$ be the solution to $\dot{z}(t) = u_\theta^E(t, z)$ in E. Then

$$x(t) = \phi^{-1}(z(t))$$

is the exact solution to the Riemannian ODE on $\mathcal{M}$.

$\rightarrow$ We **avoid all manifold-specific operations (Exp, Log, Parallel Transport) during training and sampling** while being exactly equivalent to Riemannian Flow Matching on $(\mathcal{M}, g)$.

Diffeomorphisms for covariance and correlation matrices

We use specific global diffeomorphisms for neuroimaging data:

- **Covariance ($\mathbb{S}_d^{++}$):** Log-Euclidean map $\phi_{\mathbb{S}_d^{++}} : \mathbb{S}_d^{++} \to \mathbb{R}^{d(d+1)/2}$.

$$\phi_{\mathbb{S}_d^{++}}(\boldsymbol{\Sigma}) = \mathrm{vec}(\log(\boldsymbol{\Sigma}))$$

- **Correlation (Corr_d):** Normalized Cholesky map $\phi_{\mathrm{Corr}_d} : \mathrm{Corr}_d \to \mathbb{R}^{d(d-1)/2}$.

$$\phi_{\mathrm{Corr}_d}(\boldsymbol{\Sigma}) = \mathrm{vec}(\mathrm{NormChol}(\boldsymbol{\Sigma}))$$

Experiments

Datasets:

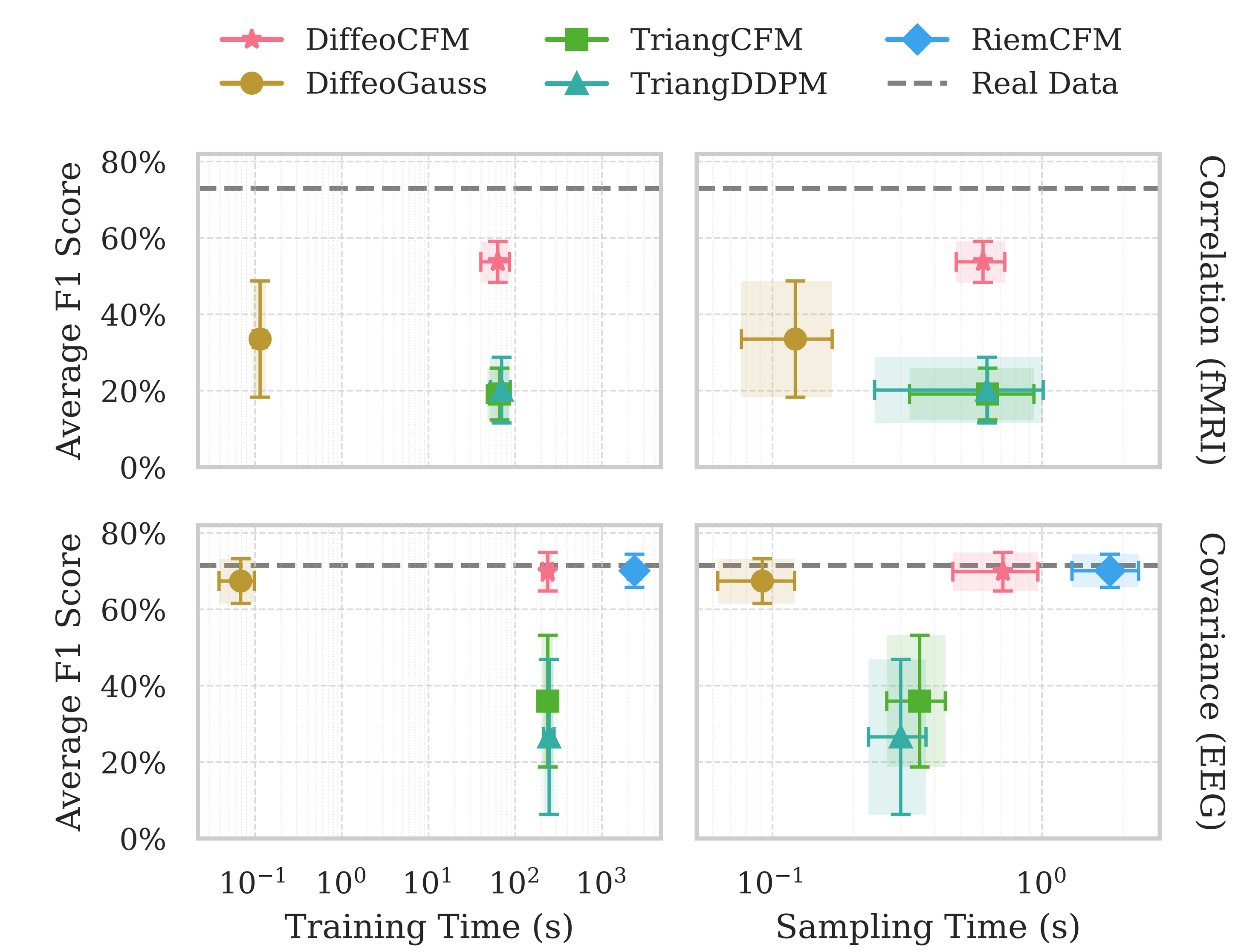
- fMRI (correlation): ABIDE, ADNI, OASIS-3, 4600 scans from 2800 subjects with disease labels (Autism, Alzheimer's, Control) and 10 subject-level splits.
- EEG (covariance): BNCI2014, BNCI2015, $> 30k$ trials from 26 subjects with right hand vs. feet motor imagery labels and cross-session splits.

Baselines:

- DiffeoGauss: one Gaussian per class in E with ϕ [2].
- TriangDDPM / TriangCFM: generates triangular part with DDPM or FM + projection $\boldsymbol{\Sigma} \leftarrow (1-\alpha)\boldsymbol{\Sigma} + \alpha\mathbf{I}$.
- RiemCFM: Riemannian Flow Matching (affine-invariant metric) [1].

Evaluation Metrics:

- Distribution Quality: α-Precision (fidelity) & β-Recall (diversity) [3].
- Utility (CAS): train a classifier on *generated* data $\rightarrow$ test on *real* data [4].
- Efficiency: training and sampling time.



References

[1] Ricky TQ Chen and Yaron Lipman. "Flow matching on general geometries". In: *The Twelfth International Conference on Learning Representations*. 2024.

[2] Thibault de Surrel et al. "Wrapped Gaussian on the manifold of Symmetric Positive Definite Matrices". In: *Forty-second International Conference on Machine Learning*. 2025.

[3] Ahmed Alaa et al. "How faithful is your synthetic data? sample-level metrics for evaluating and auditing generative models". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 290–306.

[4] Suman Ravuri and Oriol Vinyals. "Classification accuracy score for conditional generative models". In: *Advances in neural information processing systems* 32 (2019).

antoine.collas@inria.fr