IOCFF 2026 Conference

The 1st International Online Conference on Fractal and Fractional

13–15 April 2026 | Online

MDPI

Fractional-Order Learning Dynamics on SPD Manifolds

İnan Ünal, Özal Yıldırım

Department of Computer Engineering, Faculty of Engineering, Munzur University, Tunceli, 62000, Turkey

Department of Software Engineering, Faculty of Technology, Fırat University, Elazığ, 23000, Turkey

INTRODUCTION & AIM

The aim is to go beyond classical gradient descent and introduce a memory effect into the learning process.

- Multichannel time-series data (e.g., Human Activity Recognition) are effectively represented as covariance matrices. These matrices naturally reside on a non-Euclidean, curved space known as the Symmetric Positive Definite (SPD) manifold.
- Applying standard Euclidean optimization directly to SPD matrices violates their geometric structure. While classical Riemannian Gradient Descent (RGD) preserves this geometry, its integer-order derivatives make it strictly local and memoryless, leading to slower convergence and sensitivity to noisy gradients.
- Fractional calculus introduces a powerful "memory effect" into optimization. However, integrating it with non-Euclidean geometry is challenging, as applying fractional operators directly to matrices forces them outside the manifold boundaries.

The primary objective of this study is to bridge the gap between differential geometry and fractional calculus by proposing a novel, memory-aware learning paradigm. Specifically, we aim to:

- ✓ Develop a New Framework: Introduce a Fractional Riemannian Gradient Descent (FRGD) algorithm that operates intrinsically on the SPD manifold.
- ✓ Preserve Geometric Constraints: Formulate the fractional-order update rule strictly within the tangent bundle of the manifold, utilizing the Affine-Invariant Riemannian Metric (AIRM) and Exponential map to ensure all updates remain strictly positive definite.
- ✓ Enhance Optimization Dynamics: Utilize the fractional order $(\nu > 1)$ as an implicit temporal regularizer to incorporate past gradient history.

This geometric "memory" aims to dampen high-frequency oscillations, bypass local minima, and yield smoother, more stable convergence for HAR classification.

METHOD

Our core contribution is the Fractional Riemannian Gradient Descent (FRGD) algorithm, designed to compute the Fréchet mean for the Minimum Distance to Mean (MDM) classifier.

The Riemannian Objective: HAR sensor signals are mapped to 9×9 SPD covariance matrices (X_i). Instead of the arithmetic mean, we compute the intrinsic Fréchet mean (M) that minimizes the Affine-Invariant Riemannian Metric (AIRM) distance:

$$L(M) = \frac{1}{2N}\sum_{i=1}^{N} \left\|\log\left(M^{-\frac{1}{2}}X_i M^{-\frac{1}{2}}\right)\right\|_F^2$$

Tangent Space Gradient: The classical gradient is computed by mapping target matrices into the flat tangent space using the Riemannian Logarithmic map:

$$\nabla_R L(M_t) = -\frac{1}{N}\sum_{i=1}^{N} \log_{M_t}(X_i)$$

Fractional Modulation (The Core Innovation): Applying fractional derivatives directly to matrices destroys their SPD structure. Instead, we lift a Caputo-type fractional modifier into the tangent space to modulate the gradient's magnitude:

$$V_t = \frac{1}{\Gamma(2-\nu)}\|\nabla_R L(M_t)\|_{M_t}^{1-\nu}\nabla_R L(M_t)$$

(For $1 < \nu < 2$, this acts as a geometric low-pass filter, penalizing oscillatory gradients and mimicking momentum).

Geodesic Projection: Finally, the fractionally modulated update (V_t) is mapped back onto the curved manifold via the Exponential map, ensuring the new matrix remains strictly positive definite:

$$M_{t+1} = \mathrm{Exp}_{M_t}(-\eta V_t)$$

Dataset & Representation (UCI HAR)

We utilized the UCI Human Activity Recognition with Smartphones dataset, segmenting multi-channel smartphone sensor data into 128-reading temporal windows across 9 channels $(S_i \in \mathbb{R}^{128\times 9})$. To encapsulate the cross-channel statistical dependencies, we computed the empirical covariance matrix for each window. By applying a minimal ridge regularization $(+10^{-4}\mathrm{I})$ to prevent computational singularities, we successfully transformed each raw time-series instance into a strictly positive definite matrix. This pipeline embedded the entire temporal dataset as geometric points residing on the curved 9×9 SPD manifold (S_{++}^{9}), providing the mathematical foundation for our Fractional Riemannian optimization.

To execute the fractional-order learning framework, this implementation relies on specialized mathematical packages beyond the standard data science stack. The geomstats library is central to the methodology, providing the necessary infrastructure to define the Symmetric Positive Definite (SPD) manifolds and natively compute the intrinsic Fréchet mean that serves as the optimization target. Furthermore, advanced modules from scipy are utilized to handle the complex matrix calculus; specifically, scipy.linalg.sqrtm and scipy.linalg.logm are strictly required to compute the affine-invariant Riemannian distance for the loss function, while scipy.special.gamma robustly evaluates the Gamma function governing the Caputo-type fractional update rule.

RESULTS & DISCUSSION

Both classical RGD $(\nu = 1.0)$ and our FRGD $(\nu > 1.0)$ kept trajectories strictly inside the manifold. However, FRGD exhibited a distinct "sweeping curve," visually proving that the fractional memory effect allows the path to intelligently bypass sharply conditioned regions.

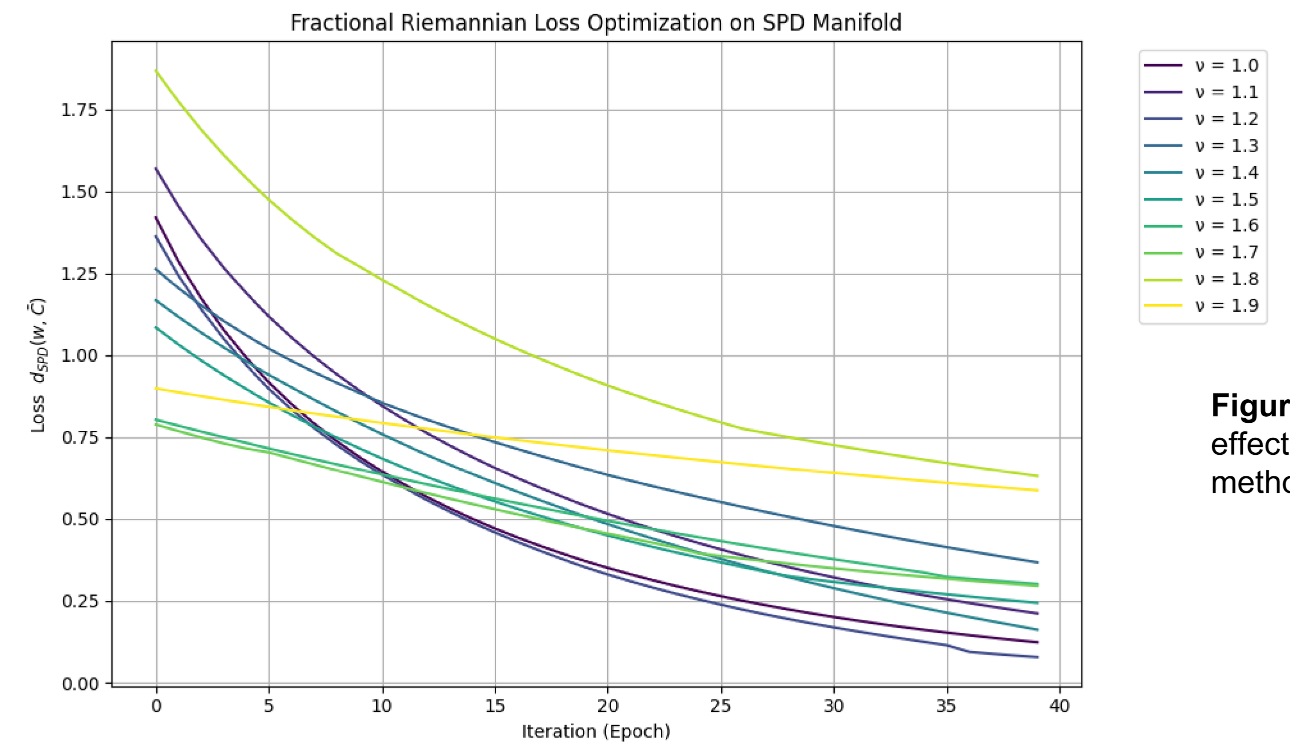

Figure-1: Memory effect of FRGD method.

Convergence of the FRGD algorithm computing the Fréchet mean for the "WALKING" class on the 9×9 empirical covariance manifold. Classical RGD $(\nu = 1.0)$ exhibits slow descent. In contrast, the fractional-order updates ($\nu = 1.2$ and 1.5) successfully leverage gradient memory to significantly accelerate convergence. Notably, the fractional memory acts as an implicit regularizer, dampening the high-frequency noise inherent in real sensor data and ensuring a steep, stable descent without overshoot (Figure-1).

Unlike classical models, the risk of overfitting in our framework is directly tied to the fractional 'memory effect'. Operating within the optimal fractional range $(\nu \in [1.2, 1.5])$ acts as a natural low-pass filter against high-frequency sensor noise, preventing the model from memorizing spurious artifacts and thereby enhancing generalization. Conversely, excessively high fractional orders $(\nu \to 2.0)$ cause the algorithm to 'over-remember' past gradients. This memory over-accumulation forces the model to overfit historical noise rather than the true manifold geometry, causing the optimization trajectory to deviate from the Fréchet mean with severe, undamped oscillations.

CONCLUSION

This study successfully bridges fractional calculus and information geometry by introducing the Fractional Riemannian Gradient Descent (FRGD) framework. By formulating fractional-order updates strictly within the tangent space, we endowed the optimization process with a robust "memory effect" while preserving the strict geometric constraints of the SPD manifold. Empirical results on the HAR dataset prove that optimal fractional orders $(\nu \in [1.2, 1.5])$ act as an implicit temporal regularizer—accelerating convergence, dampening high-frequency sensor noise, and preventing overfitting. Ultimately, FRGD provides a highly stable, memory-aware paradigm that significantly enhances manifold-based machine learning.

FUTURE WORK / REFERENCES

Future research will explore dynamically adapting the fractional order ν during training to balance momentum-based exploration with precise exploitation, and generalizing the framework to other non-Euclidean spaces like Grassmann or Stiefel manifolds. Most compellingly, extending this manifold-aware memory effect to Natural Language Processing—specifically to stabilize and accelerate the optimization of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) architectures—represents a major frontier for integrating information geometry with generative AI.

[1] Tuzel, O., Porikli, F., & Meer, P. (2006). Region covariance: A fast descripto for detection and classification. In European Conference on Computer Vision (pp. 589–600). Springer.
[2] Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2011). Multiclass brain–computer interface classification by Riemannian geometry. IEEE Transactions on Biomedical Engineering, 59 (4), 920–928.
[3] Amari, S. (2016). Information Geometry and Its Applications (Vol. 194). Springer.
[4] Absil, P.-A., Mahony, R., & Sepulchre, R. (2008). Optimization Algorithms on Matrix Manifolds. Princeton University Press.
[5] Pennec, X., Fillard, P., & Ayache, N. (2006). A Riemannian framework for tensor computing. International Journal of Computer Vision, 66 (1), 41–66.
[6] Podlubny, I. (1998). Fractional Differential Equations. Academic Press.
[7] Pu, Y.-F., Yi, Z., & Zhou, J.-L. (2014). Fractional-order gradient descent algorithm. IEEE Transactions on Neural Networks and Learning Systems, 26 (9), 2087–2101.
[8] Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM Journal on Matrix Analysis and Applications, 26 (3), 735–747.
[9] Bhatia, R. (2009). Positive Definite Matrices. Princeton University Press.
[10] Chen, Y., Wei, Y., Wang, Y., et al. (2020). Fractional-order gradient descent learning of neural networks: discrete-time dynamics and experimental validation. Neural Networks, 130, 142–152.
[11] Fletcher, P. T., Lu, C., Pizer, S. M., & Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging, 23 (8), 995–1005.
[12] Bini, D. A., & Iannazzo, B. (2013). Computing the Karcher mean of symmetric positive definite matrices. Linear Algebra and its Applications, 438 (4), 1700–1710.
[13] Wang, J., Wen, Y., Gou, Y., Ye, Z., & Chen, H. (2017). Fractional-order gradient descent learning of BP neural networks with Caputo derivative. Neural Networks, 89, 19–30.
[14] Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). A public domain dataset for human activity recognition using smartphones. ESANN, 3, 3.