A GENERAL RIEMANNIAN PRINCIPAL COMPONENT ANALYSIS FRAMEWORK

OLDEMAR RODRÍGUEZ

Abstract. In this talk we propose an innovative extension of Principal Component Analysis (PCA) that transcends the traditional assumption of data lying in Euclidean space, enabling its application to data on Riemannian manifolds. The primary challenge addressed is the lack of vector space operations on such manifolds. Fletcher et al., in their work *Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape*, proposed Principal Geodesic Analysis (PGA) as a geometric approach to analyze data on Riemannian manifolds, particularly effective for structured datasets like medical images, where the manifold's intrinsic structure is apparent. However, PGA's applicability is limited when dealing with general datasets that lack an implicit local distance notion. In this work, we introduce a generalized framework, termed *Riemannian Principal Component Analysis (R-PCA)*, to extend PGA for any data endowed with a local distance structure. Specifically, we adapt the PCA methodology to Riemannian manifolds by equipping data tables with local metrics, enabling the incorporation of manifold geometry. This framework provides a unified approach for dimensionality reduction and statistical analysis directly on manifolds, opening new possibilities for datasets with region-specific or part-specific distance notions, ensuring respect for their intrinsic geometric and topological properties.

Keywords: domain decomposition, BDDC deluxe preconditioners, adaptive primal constraints, elliptic problems.

Mathematics Subject Classifications (2010): 65F08, 65N30, 65N35, 65N55.

References

[1] 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. https://doi.org/10.1109/TMI.2004.831793

[2] Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9 (Nov), 2579-2605. Submitted 5/08; Revised 9/08; Published 11/08.

[3] McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *cite arxiv: 1802.03426 Comment: Reference implementation available at http://github.com/lmcinnes/umap / 6*, 352–357. http://arxiv.org/abs/1802.03426

[4] Pennec X. Sommer S. and Fletcher T. (Eds). (2020). *Riemannian Geometric Statistics in Medical Image Analysis.* Academic Press, Elsevier.

[5] Rodríguez, O. (2025). *riemannian_stats: Tools for Statistical Analysis of Riemannian Data* (v1.0) [Python package]. GitHub. Retrieved from https://github.com/OldemarRodriguez/riemannian_stats.

[6] Rodríguez, O. (2025). Riemannian Statistics for Any Type of Data. *Data Science, Classification, and Artificial Intelligence for Modeling Decision Making*, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 151–160. Springer Nature Switzerland, Cham. doi:10.1007/978-3-031-85870-3_17.

[7] Samaria, F. S., & Harter, A. C. (1994). *Parameterisation of a stochastic model for human face identification.* Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision, 138–142. https://doi.org/10.1109/ACV.1994.341300

Centro de Investigación en Matemática Pura y Aplicada, Universidad de Costa Rica, San Jose, Costa Rica, 11501

Email address: oldemar.rodriguez@ucr.ac.cr