Proceedings of the 4th International Conference on Statistics: Theory and Applications (ICSTA'22)
Prague, Czech Republic – July 28- 30, 2022
Paper No. 143
DOI: 10.11159/icsta22.143

A Unified Framework for Principal Subspace Analysis from the Hamiltonian Viewpoint

Ke Yu
University of Oxford
Park End Street, Oxford, UK
ke.yu.dphil@said.oxford.edu

Extended Abstract

In this work, we develop a general and unified framework for principal component analysis (PCA) applicable to Riemannian, sub-Riemannian and symplectic manifold-valued data and functional data. Almost all existing statistical methods for manifold data rely on the tangent bundle of the manifold, with the purpose of transforming the nonlinear manifold to the linear tangent spaces. However, such methods become invalid when the tangent vectors are constrained to lie in a subspace of the tangent space since the exponential map will no longer be a local diffeomorphism. This scenario, known as the sub-Riemannian geometry, has attracted considerable attention in recent years.

We propose to shift the tangent space viewpoint and move towards the dual spaces of the tangent spaces, i.e., the cotangent spaces, and build subspaces based on initial covectors. More generally, motivated by the Arnold-Liouville theorem we propose the anchor-compatible identification for subspaces with first integrals (ACISFI), which constructs a properly nested sequence of subspaces as the fibres of a carefully chosen set of functionally independent functions defined on the cotangent bundle, i.e., the first integrals of the Hamiltonian system, generalising the ideas of obtaining subspaces from linearly independent tangent vectors [1] or from affinely independent points [2]. There are several advantages of the ACISFI over the existing PCA on manifolds. First, the subspaces can be learnt from sample points in a completely data-driven way. We do not impose a particular form for the Hamiltonian, e.g., the Hamiltonian which induces the geodesic flow, but it can be chosen as any smooth function on the manifold, and there is not any a priori restriction on the form of first integrals as well. Second, the submanifolds can be defined globally. There is no need to be constrained within the complement of the cut loci of data points. Third, the user-defined anchor point can be guaranteed to be included in the subspace, whereas a particular point may appear or disappear as the dimension of the subspace varies in the exponential barycentric subspaces (EBS) [2]. Fourth, it is computationally simpler to implement in practice. Furthermore, with an alternative definition using foliations and Bott partial connections of the EBS, we prove that our proposed approaches would generate subspaces which coincide with the EBS in some cases, which ultimately leads to a unified framework for subspace analysis. We also extend this framework to PCA for manifold-valued functional data. We transform the random processes on the manifold to function-valued stochastic processes [3] and prove that it yields upper bounds to residual variances without the need of an assumption on the sectional curvatures on the manifold, which is required in [4]. We illustrate our methods through simulated data in the symplectic manifold R^{2n} and the cotangent bundle of the Stiefel manifold, and brain imaging data from the the Developing Human Connectome Project (dHCP).

References

[1] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, “Principal geodesic analysis for the study of nonlinear statistics of shape,” *IEEE Trans. Med. Imaging*, vol. 23, no. 8, pp. 995-1005, Aug. 2004.
[2] X. Pennec, “Barycentric subspace analysis on manifolds,” *Ann. Statist.*, vol. 46, no. 6A, pp. 2711-2746, Dec. 2018.
[3] K. Chen, P. Delicado, and H.-G. Müller, “Modelling function-valued stochastic processes, with applications to fertility dynamics,” *J. Roy. Statist. Soc. Ser. B. Stat. Methodol.*, vol. 79, no. 1, pp. 177-196, Feb. 2016.
[4] X. Dai and H.-G. Müller, “Principal component analysis for functional data on Riemannian manifolds and spheres,” *Ann. Statist.*, vol. 46, no. 6B, pp. 3334-3361, Dec. 2018.