Hierarchical Clustering on Special Manifolds

Angelos Markos[1*], George Menexes[2]

1. Laboratory of Mathematics & Computer Science, School of Primary Education, Democritus University of Thrace, Greece
2. Laboratory of Agronomy, School of Agriculture, Aristotle University of Thessaloniki, Greece

* Contact author: amarkos@eled.duth.gr

Keywords: Geometric data analysis, Hierarchical clustering, Matrix concordance, Riemannian manifolds

In this work, we address the problem of comparing a set of vectors to other sets of vectors, which naturally corresponds to a clustering problem on spaces of orthogonal linear projections. Such data arise in earth and biological sciences, medicine, computer vision and signal processing. In this context, we review measures for calculating distances between orthonormal matrices and between equivalence classes of matrices that span the same subspace. All distances can be represented with principal angles and their relationships with well established similarity criteria, such as the RV coefficient, are also considered. We adopt two notions of the mean or centroid of subspaces, each associated with a different distance metric: the Karcher mean, which minimizes the sum of squared geodesic distances and a Procrustes mean relying on the embedding of a manifold in the ambient Euclidean space. By exploiting the differential geometry of special Riemannian manifolds, we introduce some hierarchical clustering methods to efficiently group sets of orthonormal matrices. The proposed methods are demonstrated using synthetic and real data.

References

Absil, P.A., Mahony, R. and Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press.

Chikuse, Y. (2003). *Statistics on special manifolds*, Lecture Notes in Statistics, vol. 174, Springer, New York.

Edelman, A., Arias, T., & Smith, S. (1998). The Geometry of Algorithms with Orthogonality Constraints. *SIAM Journal on Matrix Analysis and Applications*, **20**(2), 303–353.

Golub, G.H. & Van Loan, C.F. (1996). *Matrix Computations*, Johns Hopkins University Press, Baltimore.

Robert, P. & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Applied Statistics*, **25**, 257–265.