Master Internship at Inria Sophia Antipolis

Epione & Cronos teams
ERC G-Statistics

Interpolation and extrapolation of correlation matrices for dynamic functional connectivity

Project Description:

Symmetric positive definite (SPD) matrices are ubiquitous in signal processing and are a central object in medical image processing for instance in diffusion tensor imaging (DTI). The difficulties encountered with the usual Euclidean metric led to develop alternatives based on Riemannian metrics such as the affine-invariant [Pennec et al., 2006] or the log-Euclidean ones [Arsigny et al, 2007]. Furthermore, many image processing algorithms were shown to be generalizable to this context using geodesics of the Riemannian structure instead of Euclidean straight lines. In other domains such as Brain-Computer Interfaces (BCI), the affine-invariant metric on SPD matrices was also shown to produce radically better results for clustering signals [Barachant et al. 2012].

However, in a number of practical applications, the scaling of each individual variable of a multivariate distribution is considered as a nuisance parameter. For instance, the scale of individual signals in resting-state functional magnetic resonance imaging (rs-fMRI) is not physiologically meaningful. Thus, most of the current methods renormalize the signals to work with correlation matrices, that are intrinsically invariant to the scaling of individual signals, instead of covariance matrices. Correlation matrices are more generally used to describe the brain connectivity in anatomical and functional neuroimaging, at both the sensor and brain level. When dealing with functional imaging data such as resting-state function fMRI, these matrices can be computed using a sliding window approach, leading to a timeseries of correlation matrices. This sequence contains valuable insight on the dynamics of brain networks that are however difficult to extract. Given a series of correlation matrices, we are interested in:

1. predicting the future correlation matrices given the previous ones;
2. denoising correlation matrices by assuming temporal smoothness;
3. classifying matrices to identify “state changes” of the temporal signal and the underlying brain networks.

These 3 tasks all require the *correct* interpolation and extrapolation of correlation matrices. Using the usual Euclidean structure on correlation coefficients raises problems because the valid correlation matrices belong to a strange berlingot-shape domain included in the cube $[-1,1]^n$. Thus, extrapolating in any direction reaches the boundary very shortly. Moreover, even if this domain is convex and thus the linear interpolation is well defined between any two correlation matrices, this linear Euclidean interpolation do not seem to describe particularly well intermediate data in practical cases.

Thus, there is a need to explore computationally more convenient and more meaningful Riemannian metrics on correlation matrices where geodesics would be better interpolating and extrapolating data. For that purpose, several Riemannian metrics were recently introduced, similarly to the case of the now classical SPD matrices. On the space of full-rank correlation matrices, the first idea is to take the quotient of the affine -invariant metric by the diagonal scaling group action [David and Gu 2019]. However, it was shown that the curvature of this quotient-affine metric is of non-constant sign and unbounded from above, which may pose important numerical problems with this geometry. In [Thanwerdas & Pennec 2022], we introduce three computationally more convenient Hadamard or log-Euclidean metrics, along with their basic geometric operations.

The goal of this internship is to implement the above metrics on correlation matrices and to test the quality of their respective geodesic regression on real resting-state fMRI correlation time-series data, both in interpolation and extrapolation regimes. The implementation will be done in the open-source geomstats python package (https://github.com/geomstats/geomstats) [Miolane et al. 2020]. For each metric, implementing the basic Riemannian operations (essentially the exponential, i.e. geodesic shooting, and the logarithm, i.e. the boundary value geodesic problem) is sufficient to be able to call the generic geodesic regression module. One will then be able to compare the different metrics on the rs-fMRI data and to design a test to select the most proper metric for these data. Thanks to the specific log-Euclidean structure of some of these metrics, the processing framework can be drastically simplified by mapping all the data to a Euclidean space, so that any improvement of the predictive performances with these metrics can be seamlessly put in production quite immediately for application purposes. One can expect here a drastic impact on dynamic functional connectivity analysis.

This Master 2 internship subject can be followed by a PhD with both theoretical and application extensions, including in different application domains.

Hosting groups:

The internship will be advised by X. Pennec (Epione) and Samuel Deslauriers-Gauthier (CRONOS). At Epione, the intern will be more particularly part of the ERC G-Statistics group. The EPIONE and CRONOS (successor of ATHENA) teams (Inria Center of Univ. Cote d'Azur) are located in the tech Park of Sophia Antipolis and in Nice, in the French Riviera. The internship will also be done in close collaboration with Yann Thanwerdas at Ecole CentraleSupelec.

Required competences:

Competences in signal processing and statistics are required as well as a god knowledge of mathematics and in particular differential geometry (Master 2 level). Solid programming and IT skills are necessary (Python, bash scripting, version control systems), along with strong communication abilities.

Contacts:

Xavier.pennec@inria.fr

References:

- [Arsigny et al 2007] V. Arsigny, P. Fillard, X. Pennec, N. Ayache. **Log-Euclidean metrics for fast and simple calculus on diffusion tensors**. Magnetic Resonnance in Medicine 56(2): 411-421, 2006. https://doi.org/10.1002/mrm.20965
- [Barachant et al. 2012] A Barachant, S Bonnet, M Congedo, C Jutten. **Multi-class Brain Computer Interface Classification by Riemannian Geometry**. IEEE Trans. on Biomedical Engineering, 59 (4), 920-928, 2012. https://ieeexplore.ieee.org/abstract/document/6046114
- [David and Gu 2019] P. David and W. Gu. **A Riemannian structure for correlation matrices**. Operators and Matrices, 13(3):607–627, 2019.
- [Miolane et al 202] N. Miolane et al. Geomstats: **A Python Package for Riemannian Geometry in Machine Learning**. J. of Machine Learning Research (JMLR) 21(223):1−9, 2020. https://www.jmlr.org/papers/v21/19-027.html
- [Pennec et al., 2006] X. Pennec, P. Fillard, and N. Ayache. **A Riemannian Framework for Tensor Computing**. Int. J. of Computer Vision, 66(1):41-66, January 2006. https://doi.org/10.1109/TMI.2007.907286
- [Thanwerdas & Pennec 2022] Y. Thanwerdas, Xavier Pennec. **Theoretically and computationally convenient geometries on full-rank correlation matrices**. SIAM Journal on Matrix Analysis and Applications, 43(4):1851-1872, 2022. https://doi.org/10.1137/22M1471729