3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

# Ph.D. research topic

---

- Title of the proposed topic: **Geometry, stratification and applications of structured correlation matrices**
- Research axis of the 3IA: Core elements of AI
- **Supervisor (name, affiliation, email): Xavier Pennec (Inria, Epione)**
- Potential co-supervisor: Samuel Deslauriers-Gauthier (Inria, Cronos)
- The laboratory and/or research group: Epione, Centre Inria d'Université Côte d'Azur

---

**Apply by sending an email directly to the supervisor.**
**The application will include:**

- Letter of recommendation of the supervisor indicated above
- Curriculum vitæ.
- Motivation Letter.
- Academic transcripts of a master's degree(s) or equivalent.
- At least, one letter of recommendation.
- Internship report, if possible.

---

- Description of the topic:

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 the development of 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 time-series of correlation matrices, that are intrinsically invariant to the scaling of individual signals, instead of covariance matrices. Likewise, in biology, the expression of specific RNA expression levels may be multiplicatively biased by amplification in single-cell transcriptomics so that the correlations between the expression of different genes or groups of genes may be more meaningful than their covariance. Finally, Gaussian graphical models are often used to model interactions between random variables that are sparsely conditionally correlated. The graph of conditional dependences is encoded by the non-zero elements of the precision matrix, the inverse of the covariance matrix of the Gaussian process. Accounting for biases in the individual variance of each variable leads to focus on partial correlation matrices with a structured set of zeros. In this process, fully correlated variables lead to rank deficiency of the matrices, so that it is important to be able to work with and compare low rank partial correlation matrices with different ranks and structures in a single space.

Using the usual Euclidean structure on correlation coefficients in not convenient because the valid correlation matrices belong to a convex berlingot-shape domain included in the hypercube $[-1,1]^n$. Thus, extrapolating in any direction reaches the boundary very shortly and any learning system relying on this Euclidean structure will yield unstable results. Moreover, even if this domain is convex and thus the linear interpolation is well defined between any two correlation matrices, this linear Euclidean interpolation does not describe the intermediate data particularly well in practice. This is of particular importance in machine learning, e.g. in variational auto-encoders, where meaningful interpolation of the latent space is paramount. Thus, there is a need to explore computationally more convenient and more meaningful geometric structures on correlation matrices where geodesics would better interpolate and extrapolate the data with high potential impact in AI. For that purpose, several Riemannian metrics were recently introduced on the space of full-rank correlation matrices. A first idea is to take the quotient of the affine-invariant metric by the positive diagonal scaling group action [David and Gu 2019]. In [Thanwerdas & Pennec 2022], three computationally more convenient Hadamard or log-Euclidean metrics were introduced, along with their basic geometric operations.

The goal of this PhD topic is to investigate how well these geometric structures impact the application problems in neuroimaging, biology and machine learning. Since the currently proposed geometric structures cannot deal with low rank or sparse (partial) correlation matrices, a secondary theoretical goal will be to design new mathematically and computationally convenient structures to handle the stratifications induced by the rank and sparsity pattern. To fully achieve this, a significant step will be to understand how to generalise the theory and implementation of geometric statistics from smooth Riemannian manifolds to stratified Riemannian spaces. With such tools, we expect to demonstrate radically new possibilities of modelling and learning, notably with Gaussian graphical networks

## Hosting groups:

The PhD will be advised by X. Pennec (Epione) with the ERC G-Statistics group. The PhD will require close collaborations with Samuel Deslauriers-Gauthier (CRONOS) for the neuroimaging part and with Yann Thanwerdas at Ecole CentraleSupélec for the mathematical part. The EPIONE and CRONOS teams (Inria Center of Univ. Cote d'Azur) are located in the tech Park of Sophia Antipolis and in Nice, in the French Riviera.

## Required competences:

Competences in signal processing and statistics are required as well as a deep 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 Resonance 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