**2022-05183 - PhD Position F/M Rethinking statistical methods with Flags spaces**

**Type de contrat :** CDD  
**Niveau de diplôme exigé :** Bac + 5 ou équivalent  
**Fonction :** Doctorant  
**Niveau d'expérience souhaité :** Jeune diplômé

**A propos du centre ou de la direction fonctionnelle**

The Inria Sophia Antipolis - Méditerranée center counts 34 research teams as well as 7 support departments. The center's staff (about 500 people including 320 Inria employees) is made up of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrative staff. 1/3 of the staff are civil servants, the others are contractual agents. The majority of the center's research teams are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Four teams are based in Montpellier and two teams are hosted in Bologna in Italy and Athens. The Center is a founding member of Université Côte d'Azur and partner of the I-site MUSE supported by the University of Montpellier.

**Contexte et atouts du poste**

This PhD proposal is part of the ERC G-Statistics advance grant # 786854 (2018-2023) from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program.

**Mission confiée**

Principal Component Analysis (PCA) is the ubiquitous tool to obtain low dimensional representation of many types of data in linear spaces. PCA is often seen as an iterative method where one component is added at each step. We can also see it as the search for a sequence of embedded linear subspaces that best approximate the data. Geometrically, sequences of properly embedded linear spaces of increasing dimension are points of a flag manifold. We have recently shown that PCA is optimizing a criterion on the flag manifold: the accumulated unexplained variance (AUV). On this basis, new robust variants like the minimal accumulated unexplained p-variance have been proposed [Pennec AOS 2018, GSI 2017].

In Riemannian manifolds, PCA can be generalized thanks to the analysis of the covariance matrix of the data in the tangent space at the Fréchet mean (Tangent PCA). This is often sufficient when data which are sufficiently centered around a central value (unimodal or Gaussian-like data), but generally fails for multimodal or distributions with a large variability with respect to the curvature. Instead of maximizing the explained variance, methods minimizing the unexplained variance were proposed: Principal Geodesic Analysis (PGA) and Geodesic PCA (GPCA) minimize the distance...
to a Geodesic Subspaces (GS) spanned by the geodesics going through a point with tangent vector is a restricted linear subspace of the tangent space.

A new type of subspaces in manifolds was introduced in [Pennec AOS 2018]: Barycentric Subspaces (BS) are implicitly defined as the locus of weighted means of k+1 reference points, and they can naturally be nested, which allow the construction of inductive forward or backward nested subspaces approximating data points. This results into a particularly appealing generalization of PCA on manifolds, that is called Barycentric Subspaces Analysis (BSA). In practice, a hierarchy of embedded barycentric subspaces is defined by an ordered series of points in the manifold, and data may be characterized by their barycentric coordinates inside the submanifold plus an orthogonal residual from the data points to their projection in the BS. Such a hierarchy of properly embedded barycentric subspaces may be seen as a natural generalization to manifolds of the flag manifolds of embedded linear subspaces.

The goal of this PhD is to explore, study and test extensions and applications of classical statistical methods like PCA from these two new points of view: flag manifolds and barycentric subspace analysis. We expect to further extend the procedure to new criteria interpolating between PCA and k-means. We also aim at reformulating nonlinear iterative least squares algorithms (NiPALS) like Partial least squares (PLS), or independent component analysis (ICA) [Nichimori et al. 2006] in purely geometric terms as optimization methods on flag spaces. Such a corpus would bring under a unique geometric formulation four of the mostly used data statistical modeling methods. Moreover, these formulations should be seamlessly generalizable to manifold-valued data thanks to the barycentric subspace approach.

Starting with simple data living in Euclidean spaces, a first step will be to implement geodesic computational methods based on the geometry of the flag manifolds [Ye et al 2019] in the geomstats python package. This will enable the use of geometric optimization methods on flags like natural Gauss-Newton and derived schemes such as levenberg-Marquardt gradient descent methods to find the optimum of the PCA criterion and its extensions. It will also be interesting to investigate neural networks optimizations of the PCA loss function and its newly developed extensions. We will turn to BSA and modify the criterion to interpolate between PCA and k-means, before looking at extensions of PLS and ICA.

From the practical point of view, applications will be investigated in computational anatomy with statistics on diffeomorphisms to analyse the shape of the brain. An example application to the analysis of 3D cardiac image sequences through non-linear image registration has shown that the optima reference points were actually very meaningful transition points between the cardiac phases in the sequence. Moreover, the barycentric coordinates were powerful signatures discriminating different clinical conditions [Rohe et al., 2016]. We will investigate how this method can be extended to inter-subject registration of brain images from the ADNI database (more than 1000 MRI of the brain with different clinical conditions). We expect reference images and barycentric weight to be indicative of the clinical condition.

Other applications involving geometric data may include simulated conformations of molecules in collaboration with I. Dryden (U. Nottingham, UK). The simulation of the motion of macromolecules produces massive amounts of conformations, whose statistical study is expected to uncover the main stable modes and transitions. Small motions of a macromolecule are random fluctuations due to thermal noise but their combination sometimes leads to large conformation transitions that are of the upper interest.

**Principales activités**

- Study the state of the art in this domain as well as related ones
- Develop a deep understanding of how flag spaces interacts with statistical tools.
- Design, prototype and evaluate solutions for the inference of flags and their generalizations from observations.
- Publish scientific results in the relevant scientific conferences and journals

**Compétences**

- Excellence in mathematics, including advanced geometry and statistics
- Knowledge of programming in python
- Excellent communication and writing skills.
- Good command of English.

**Avantages**

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage
- Supplementary social protection

**Rémunération**

Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 1982€ per month (year 1 & 2) and 2085€ per month (year 3)
Informations générales

- **Thème/Domaine** : Approches stochastiques
  Statistiques (Big data) (BAP E)
- **Ville** : Sophia Antipolis
- **Centre Inria** : CRI Sophia Antipolis - Méditerranée
- **Date de prise de fonction souhaitée** : 2022-10-01
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 2022-09-01

Contacts

- **Equipe Inria** : EPIONE
- **Directeur de thèse** :
  Pennec Xavier / Xavier.Pennec@inria.fr

A propos d'Inria

Inria est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnels d’appui à la recherche et à l’innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L’institut s’efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l’économie.

Consignes pour postuler

**Sécurité défense** :
Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

**Politique de recrutement** :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

**Attention** : Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d’autres canaux n’est pas garanti.