PHD Thesis Proposal

Laboratory: Inria Sophia-Antipolis Méditerranée, Epione team (https://team.inria.fr/epione/)
Advisor: Xavier Pennec (Xavier.pennec@inria.fr, http://www-sop.inria.fr/members/Xavier.Pennec/)

Statistical Dimension Reduction in Non-Linear Manifolds for Brain Shape Analysis, Connectomics & Brain-Computer Interfaces

In many applications, data belong to non-linear sub-manifolds of a high dimensional space. The natural invariance properties of the space in which the data live often encode informative priors that turn out to be key features to improve the results of analyses. This is the case in Computational Anatomy (CA), Brain Computer Interfaces (BCI) and Brain Connectomics where data naturally belong to shape spaces, Lie groups and symmetric positive definite matrices. Since these data live in very high dimension, (this could even be infinite dimensional spaces for diffeomorphisms), any correlation analysis or interpretation requires first reducing the dimension of the space. However, a difficulty is not to lose the original structure, i.e. to find a submanifold in a manifold. Moreover, the assumption that data live in a submanifold of fixed dimension is often too strong: we rather aim at finding a consistent series of nested subspace that better and better approximate the original data. This is a significantly harder problem than classical manifold learning in Euclidean or Hilbert spaces.

The research approach investigated in this PhD will explore the power of the concept of sequences of properly nested affine subspaces (flags) in manifolds. In linear spaces, such sequences belong to flag manifolds and it has been shown that Principal Component Analysis can be rephrased as the optimization of the accumulated unexplained variance criterion in that space (Pennec 2016). The generalization to flags of affine subspaces of Riemannian manifolds, called barycentric subspace analysis, is raising an increasing interest in the geometric statistics community. Exploring other criteria on flags spaces will certainly allow more robust and more adaptive subspace approximation algorithms that will provide novel strategies for big data analysis. In order to demonstrate the generality of the methods, this PhD will investigate three applications of this framework to the brain. First, we expect to improve the statistical modelling of the variability of the anatomical shape (computational anatomy) of the brain; Second, anatomical and functional connectivity properties coming from diffusion or functional MRI are often encoded through symmetric positive definite (SPD) matrices. Here, the use of principled methods respecting the natural structure of that space should improve the consistency current results in brain connectomics. Thirdly, the natural structure of SPD matrices was also shown to enhance the classification of mental tasks from EEG signals for brain-computer interfaces (BCI). Consistent and hierarchical dimension reduction methods are likely to bring further improvements.

Computational Brain Anatomy

At the interface of geometry, statistics, image analysis and medicine, computational anatomy aims at analysing and modelling the biological variability of the organs shapes at the population level and discovering morphological differences between normal and pathological groups. Important applications include the spatial normalization of subjects in neuroscience needed in all group studies and personalized atlases in medical image analysis. The main difficulty to analyse distribution of shapes or deformations is that they belong to non-linear spaces like Riemannian manifolds, quotient spaces or affine connection spaces. An important effort has been made to compute the mean in these types of manifolds (Battacharya et al. 2003, Pennec 2006, Pennec et al. 2012). The goal of this PhD is investigate higher dimensional approximations (subspaces) of geometric data living in manifolds, either globally (extension of PCA or PLS) or locally (extension of statistical manifold learning). Multi-atlas methods are currently among the best brain segmentation methods. Defining a series of nested subspaces parameterized by these multiple brain atlases could further improve the results by providing globally consistent graph of transformations (registration transitivity). Several very large databases of brain images and segmentations are accessible (ADNI, UKBiobank, with more than 10k subject) and the expertise on brain image processing is readily available within the Epione team.

Connectomics: from anatomical to functional brain networks

The increasing availability of diffusion (dMRI) and functional (fMRI) Magnetic Resonance Imaging has led to a number of studies of brain networks based on connectivity matrices resulting from the integration of local connectivity measures extracted from dMRI or the correlation matrices of time signals at different regions/voxels in fMRI sequences. Given populations of networks of control and/ or disease subjects, typically represented as symmetric positive definite (SPD) matrices in very high dimension, the goal is to classify them in a supervised or unsupervised way and to model the typical differences between subgroups in order to establish potentially new biomarkers of brain states and diseases. In such an analysis, the use of a natural invariant metric on SPD matrices has shown to improve the classification results [Ng et al 2015]. In this Phd, the goal will be to develop sequences of natural nested subspaces in SPD matrices that best approximate the data and their classification. A special effort will be taken to introduce sparsity constrains in the process in order to compete with graphical lasso techniques. This axis will be investigated in collaboration with the Athena team at Inria Sophia-Antipolis and the Parietal team (B. Thirion, G. Varoquaux) at Inria Saclay.

Brain Computer Interfaces (BCI)

Electro-encephalograpm (EEG) based BCI are interfaces where the application-specific information is extracted from the resulting signal [Clerc et al., 2016] through different summary statistics, including for instance the temporal correlation of time signals. The goal is to classify segments of EEG signals in this feature space to match their corresponding mental tasks. EEG data have an extremely low Signal-to-Noise Ratio, while being subject to an important amount of cross-session and cross-subject variability. In two of the classical methods, the feature space considered is an SPD matrix and recent researches have shown that the classification task is more robust to cross-session / cross-subject variability under the framework of Riemannian geometry [Barachant et al., 2012]. As shown in recent works by Athena, while Riemannian distance-based classifiers show promising results, the high dimensionality of the manifold is still an issue [Gayraud et al., 2016]. The goal is to explore dimension reduction methods in order to solve this problem. The research in this axis will be done in collaboration with the Athena team at Inria Sophia-Antipolis (M. Clerc, N. Gayraud).

Geometric learning: Non-linear subspace approximation in manifolds

Principal Component Analysis (PCA) is a ubiquitous tool to obtain low dimensional representation of the data in linear spaces. In Riemannian manifolds, the classical tangent PCA in the tangent space at the Fréchet mean generally fails for multimodal or distributions with a large variability. Minimizing the unexplained variance with respect to a subspace was considered in Principal Geodesic Analysis (PGA) and Geodesic PCA (GPCA) but the geodesic subspaces were still defined locally at one point. A new type of natural subspaces called affine spans in manifolds were recently introduced in [Pennec, 2016]. They are implicitly defined as the locus of weighted means of k+1 reference points, which can be seen as references, archetypes or centroid of clusters. These points can be very distantly located on the manifold and affine spans are consequently more adapted to data with a large variability. Affine spans can naturally be nested to construct sequences of nested subspaces (flags of affine spans) that minimize the sum of unexplained variance of all the subspaces of the sequence (Barycentric Subspaces Analysis). PCA in Euclidean spaces was shown to actually optimize this criterion. Many extensions are possible, for instance to other statistical methods such as PLS or ICA. The PhD will aim at further developing the theory and at applying it to the previous driving applications in brain computational anatomy, brain connectomics and BCI.

References

- [Barachan, Bonnet, Congedo & Jutten 2012] Multiclass brain computer interface classification by Riemannian geometry. IEEE Transactions on Biomedical Engineering, 59(4):920-928.
- [Clerc, Bougrain & Lotte 2016] Brain-Computer Interfaces 1: Methods and Perspectives. John Wiley & Sons.
- [Gayraud, Foy & Clerc 2016] A separability marker based on high-dimensional statistics for classification confidence assessment. In IEEE Int. Conf. on Systems, Man, and Cybernetics 2016.
- [Bhattacharya and Patrangenaru. 2003] Large sample theory of intrinsic and extrinsic sample means on manifolds. I, Ann. Statist. 31 (2003), 1-29.
- [Ng, Varoquaux, Poline Greicius & Thirion 2015] Transport on Riemannian Manifold for Connectivity-based Brain Decoding. IEEE Transactions on Medical Imaging, 2016, 35(1):208-216
- [Pennec 2016] Barycentric subspace analysis on manifolds. Annals of Statistics, to appear, 2018. arXiv:1607.02833.
- [Pennec 2006] Intrinsic Statistics on Riemannian Manifolds. J. of mathematical imaging and vision 25(1):127-154.
- [Pennec & Arsigny 2012] Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups. In Matrix Information Geometry, pages 123-166. Springer.