2018-00564 - Post-doctorant (e) : Statistical geometric inference and multi-persistence in Topological Data Analysis

Level of qualifications required: PhD or equivalent
Fonction: Post-Doctoral Research Visit

About the research centre or Inria department

Located at the heart of the main national research and higher education cluster, member of the Université Paris Saclay, a major actor in the French Investments for the Future Programme (Idex, LabEx, IRT, Equipex) and partner of the main establishments present on the plateau, the centre is particularly active in three major areas: data and knowledge; safety, security and reliability; modelling, simulation and optimisation (with priority given to energy).

The 450 researchers and engineers from Inria and its partners who work in the research centre's 31 teams, the 100 research support staff members, the high-level equipment at their disposal (image walls, high-performance computing clusters, sensor networks), and the privileged relationships with prestigious industrial partners, all make Inria Saclay Île-de-France a key research centre in the local landscape and one that is oriented towards Europe and the world.

Context

The recent years have seen all domains of science, economy and even everyday life overwhelmed by massive amounts of data. Bringing scientists and users to the most relevant, often unexpected, features and giving them the tools to discover and extract the best knowledge out of their data are fundamental challenges for our modern society. Taking a closer look at data sometimes reveals that data points turn out to concentrate around, possibly low-dimensional, shapes that can be modeled as Riemannian sub-manifold or more general stratified sets.

Identifying, extracting and exploiting the topological and geometric invariants of these shapes is an old-standing problem and research direction in computational geometry (surface and low-dimensional manifold reconstruction) in deterministic settings where data are not random and not corrupted by noise and outliers. In a more general and statistical setting, this problem also appeared to be of fundamental importance to lay-down the mathematical foundations of Geometric and Topological Data Analysis and is currently attracting more and more attention among the statistical community.

There exist various statistical and machine learning methods that intend to uncover the geometric structure of data, such as clustering, manifold learning and nonlinear dimensionality reduction, principal curves, sets estimation, to name a few. Most of them assume the underlying structure to have a very simple geometry - diffeomorphic to a disc or isometric to a domain of a Euclidean space. Furthermore, the only topological information they seek for is connectivity. On another hand, with the emergence of distance based approaches and persistent homology, geometric inference and computational topology have recently known an important development. New mathematically well-founded theories gave birth to the field of Topological Data Analysis (TDA) which is knowing an increasing interest both in academy and industry.

Many important results have been established on the inference of topological and global geometric invariants during the last few years, but the inference of more local (e.g. curvatures) and metric (e.g. geodesic distances) quantities are still largely unexplored from a statistical and data analysis perspective. There is a real need to develop new approaches and algorithms with strong statistical guarantees to infer such quantities in order to propose new well-founded methods in Topological Data Analysis.

On another hand, statistical aspects of persistent homology – a pillar of TDA – have attracted a lot of attention during the last couple of years leading to spectacular and promising new results. However, statistical persistent homology is still in its infancy, in particular regarding some of its variants like zig-zag persistence or multi-persistence.
Assignment

The first aim of this post-doc is to develop new algorithms for geometric inference with statistical guarantees under various geometric and noise models, with a particular focus on the estimation of metric quantities such as geodesic distances from randomly sampled data that play a fundamental role in further Topological Data Analysis or machine learning algorithms. Beyond their statistical properties, the computational efficiency of the proposed estimators and algorithms will also be carefully studied.

The second aim of this post-doc is to study the statistical behavior of distributions of persistence diagrams, in particular in the setting of multi-persistence which is knowing an increasing interest among the mathematical and TDA community. Multi-persistent homology appears naturally in Topological Data Analysis when multivariate functions and/or multi-parameters filtrations on data are considered. Unlike classical persistent homology, multi-persistence does not admit finite descriptors as classical persistence diagrams are leading to severe theoretical and practical issues for its effective use in TDA. To address this problem, an approach consisting in generating distributions of persistence diagrams from a multi-persistence filtration and studying its statistical properties will be adopted.

Main activities

The first line of work consists in studying the problem of the estimation of the geodesic distance on sub-manifolds of $\mathbb{R}^d$ from a minimax perspective. Although there already exist some algorithms to address this problem that are based on shortest path computation on neighboring graphs built on top of the data, it is easy to show that they are far from having optimal rate of convergence as $n$ goes to infinity. Determining this optimal rate for different statistical models will be part of the work.

The second line of work will consist in studying the statistical behavior of random diagrams obtain from random uni-dimensional persistence modules derived from a given multi-dimensional persistence module. The case of the sublevel set filtrations of a multi-variate function composed with random projections will be first considered. Summaries, such as expectation, of such distributions of persistence diagrams will be studied as new topological descriptors of data.

Beyond the mathematical results, the designed estimators and algorithms will be implemented and could be integrated to the GUDHI library and its statistical TDA toolbox developed by the DataShape team. Real applications of the proposed methods could also be considered.

Skills

The successful candidate will have a strong background in mathematics and statistics (PhD in Mathematics or Statistics). Some knowledge in differential/Riemannian geometry is also expected as well as strong programming skills in C++/Python.

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration

Monthly gross salary : 2.653 euros

General Information

- **Theme/Domain**: Algorithmics, Computer Algebra and Cryptology
- **Scientific computing** (BAP E)
- **Town/city**: PALAISEAU
- **Inria Center**: CRI Saclay - Île-de-France
- **Starting date**: 2018-11-01
- **Duration of contract**: 1 year, 4 months
- **Deadline to apply**: 2018-04-10
Contacts

- Inria Team: DATASHAPE
- Recruiter: Chazal Frederic / frederic.chazal@inria.fr

Conditions for application

Applicants have to provide the following documents to be considered at the selection procedure:

- CV
- publication list and 2 representative publications (that you judge representative of your work)
- motivation letter (explaining why the selected topic, how do you think you can fit the topic, what are your motivations for the topic compared to your previous work, etc)
- 2 recommendation letters
- perspective of professional insertion after the post-doc

Defence Security:
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy:
As part of its diversity policy, all Inria positions are accessible to people with disabilities.

Warning: you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.