Activity Report 2019

Project-Team DATASHAPE

Understanding the shape of data

RESEARCH CENTERS
Saclay - Île-de-France
Sophia Antipolis - Méditerranée

THEME
Algorithmics, Computer Algebra and Cryptology

Table of contents

Project-Team DATASHAPE

Creation of the Team: 2016 January 01, updated into Project-Team: 2016 January 01

Keywords:

Computer Science and Digital Science:

A3. - Data and knowledge
A3.4. - Machine learning and statistics
A7.1. - Algorithms
A8. - Mathematics of computing
A8.1. - Discrete mathematics, combinatorics
A8.3. - Geometry, Topology
A9. - Artificial intelligence

Other Research Topics and Application Domains:

B1. - Life sciences
B2. - Health
B5. - Industry of the future
B9. - Society and Knowledge
B9.5. - Sciences

1. Team, Visitors, External Collaborators

Research Scientists

Frédéric Chazal [Team leader, Inria, Senior Researcher, Saclay - Île-de-France, HDR]
Jean-Daniel Boissonnat [Inria, Senior Researcher, Sophia Antipolis - Méditerranée, HDR]
David Cohen-Steiner [Inria, Researcher, Sophia Antipolis - Méditerranée]
Kunal Dutta [Inria, Starting Research Position, until Jan 2019, Sophia Antipolis - Méditerranée]
Marc Glisse [Inria, Researcher, Saclay - Île-de-France]
Miroslav Kramar [Inria, Advanced Research Position, until Aug 2019, Saclay - Île-de-France]
Clément Maria [Inria, Researcher, Sophia Antipolis - Méditerranée]
Steve Oudot [Inria, Researcher, Saclay - Île-de-France, HDR]
Mathijs Wintraecken [Inria, Starting Research Position, until Jan 2019, Sophia Antipolis - Méditerranée]

Faculty Members

Clément Levrard [Univ Denis Diderot, Associate Professor, until Aug 2019, Saclay - Île-de-France]
Maksims Ovsjanikovs [École polytechnique, Professor, Saclay - Île-de-France]

Post-Doctoral Fellows

Hariprasad Kannan [Inria, Post-Doctoral Fellow, Saclay - Île-de-France]
Jisu Kim [Inria, Post-Doctoral Fellow, Saclay - Île-de-France]
Martin Royer [Inria, Post-Doctoral Fellow, Saclay - Île-de-France]

PhD Students

Bertrand Beaufils [Sysnav, PhD Student, granted by CIFRE, Saclay - Île-de-France]
Nicolas Berkouk [Inria, PhD Student, Saclay - Île-de-France]
Denys Bulavka [Inria, PhD Student, until Apr 2019, Sophia Antipolis - Méditerranée]
Alex Delalande [Inria, PhD Student, from Dec 2019, Saclay - Île-de-France]
Vincent Divol [Univ Paris-Sud, PhD Student, Saclay - Île-de-France]
Siargey Kachanovich [Inria, PhD Student, from Apr 2019 until May 2019, Sophia Antipolis - Méditerranée]

Théo Lacombe [École polytechnique, PhD Student, Saclay - Île-de-France]
Etienne Lasalle [Univ Paris-Sud, PhD Student, from Sep 2019, Saclay - Île-de-France]
Siddharth Pritam [Inria, PhD Student, Sophia Antipolis - Méditerranée]
Louis Pujol [Univ Paris-Sud, PhD Student, from Nov 2019, Saclay - Île-de-France]
Owen Rouille [Inria, PhD Student, Sophia Antipolis - Méditerranée]
Raphael Tinarrage [École Normale Supérieure de Cachan, PhD Student, Saclay - Île-de-France]

Technical staff

Thomas Bonis [Inria, Engineer, Saclay - Île-de-France]
Vincent Rouvreau [Inria, Engineer, Saclay - Île-de-France]

Interns and Apprentices

Alex Delalande [Inria, from May 2019 until Nov 2019, Saclay - Île-de-France]
Vadim Lebovici [École Normale Supérieure de Paris, from Apr 2019 until Aug 2019, Saclay - Île-de-France]

Administrative Assistants

Florence Barbara [Inria, Administrative Assistant, until May 2019, Sophia Antipolis - Méditerranée]
Bahar Carabetta [Inria, Administrative Assistant, Saclay - Île-de-France]
Sophie Honnorat [Inria, Administrative Assistant, Sophia Antipolis - Méditerranée]

Visiting Scientists

Yoann Barszezak [École Normale Supérieure de Cachan, from Mar 2019 until Jul 2019, Saclay - Île-de-France]
Arijit Ghosh [Indian Statistical Institute, from Aug 2019 until Sep 2019, Sophia Antipolis - Méditerranée]
Yuichi Ike [Fujitsu Laboratories, from May 2019 until Jul 2019, Saclay - Île-de-France]
Jacob Leygonie [Oxford, from Jul 2019 until Sep 2019, Saclay - Île-de-France]
Elchanan Solomon [Brown, until Jan 2019, Saclay - Île-de-France]

External Collaborators

Clément Levrard [Univ Denis Diderot, from Sep 2019, Saclay - Île-de-France]
Bertrand Michel [Univ Pierre et Marie Curie, Saclay - Île-de-France]
Louis Pujol [Univ Paris-Sud, Oct 2019, Saclay - Île-de-France]
Georg Wilding [University of Groningen, from Sep 2019 until Nov 2019, Sophia Antipolis - Méditerranée]
Claire Brecheteau [Ecole Centrale Nantes, Saclay - Île-de-France]

2. Overall Objectives

2.1. Overall Objectives

DataShape is a research project in Topological Data Analysis (TDA), a recent field whose aim is to uncover, understand and exploit the topological and geometric structure underlying complex and possibly high dimensional data. The DATASHAPE project gathers a unique variety of expertise that allows it to embrace the mathematical, statistical, algorithmic and applied aspects of the field in a common framework ranging from fundamental theoretical studies to experimental research and software development.

The expected output of DATASHAPE is two-fold. First, we intend to set-up and develop the mathematical, statistical and algorithmic foundations of Topological and Geometric Data Analysis. Second, we intend to develop the Gudhi platform in order to provide an efficient state-of-the-art toolbox for the understanding of the topology and geometry of data.

3. Research Program

3.1. Algorithmic aspects of topological and geometric data analysis

TDA requires to construct and manipulate appropriate representations of complex and high dimensional shapes. A major difficulty comes from the fact that the complexity of data structures and algorithms used to approximate shapes rapidly grows as the dimensionality increases, which makes them intractable in high dimensions. We focus our research on simplicial complexes which offer a convenient representation of general shapes and generalize graphs and triangulations. Our work includes the study of simplicial complexes with good approximation properties and the design of compact data structures to represent them.

In low dimensions, effective shape reconstruction techniques exist that can provide precise geometric approximations very efficiently and under reasonable sampling conditions. Extending those techniques to higher dimensions as is required in the context of TDA is problematic since almost all methods in low dimensions rely on the computation of a subdivision of the ambient space. A direct extension of those methods would immediately lead to algorithms whose complexities depend exponentially on the ambient dimension, which is prohibitive in most applications. A first direction to by-pass the curse of dimensionality is to develop algorithms whose complexities depend on the intrinsic dimension of the data (which most of the time is small although unknown) rather than on the dimension of the ambient space. Another direction is to resort to cruder approximations that only captures the homotopy type or the homology of the sampled shape. The recent theory of persistent homology provides a powerful and robust tool to study the homology of sampled spaces in a stable way.

3.2. Statistical aspects of topological and geometric data analysis

The wide variety of larger and larger available data - often corrupted by noise and outliers - requires to consider the statistical properties of their topological and geometric features and to propose new relevant statistical models for their study.

There exist various statistical and machine learning methods intending to uncover the geometric structure of data. Beyond manifold learning and dimensionality reduction approaches that generally do not allow to assert the relevance of the inferred topological and geometric features and are not well-suited for the analysis of complex topological structures, set estimation methods intend to estimate, from random samples, a set around which the data is concentrated. In these methods, that include support and manifold estimation, principal curves/manifolds and their various generalizations to name a few, the estimation problems are usually considered under losses, such as Hausdorff distance or symmetric difference, that are not sensitive to the topology of the estimated sets, preventing these tools to directly infer topological or geometric information.

Regarding purely topological features, the statistical estimation of homology or homotopy type of compact subsets of Euclidean spaces, has only been considered recently, most of the time under the quite restrictive assumption that the data are randomly sampled from smooth manifolds.

In a more general setting, with the emergence of new geometric inference tools based on the study of distance functions and algebraic topology tools such as persistent homology, computational topology has recently seen an important development offering a new set of methods to infer relevant topological and geometric features of data sampled in general metric spaces. The use of these tools remains widely heuristic and until recently there were only a few preliminary results establishing connections between geometric inference, persistent homology and statistics. However, this direction has attracted a lot of attention over the last three years. In particular, stability properties and new representations of persistent homology information have led to very promising results to which the DATASHAPE members have significantly contributed. These preliminary results open many perspectives and research directions that need to be explored.

Our goal is to build on our first statistical results in TDA to develop the mathematical foundations of Statistical Topological and Geometric Data Analysis. Combined with the other objectives, our ultimate goal is to provide a well-founded and effective statistical toolbox for the understanding of topology and geometry of data.

3.3. Topological approach for multimodal data processing

Due to their geometric nature, multimodal data (images, video, 3D shapes, etc.) are of particular interest for the techniques we develop. Our goal is to establish a rigorous framework in which data having different representations can all be processed, mapped and exploited jointly. This requires adapting our tools and sometimes developing entirely new or specialized approaches.

The choice of multimedia data is motivated primarily by the fact that the amount of such data is steadily growing (with e.g. video streaming accounting for nearly two thirds of peak North-American Internet traffic, and almost half a billion images being posted on social networks each day), while at the same time it poses significant challenges in designing informative notions of (dis)-similarity as standard metrics (e.g. Euclidean distances between points) are not relevant.

3.4. Experimental research and software development

We develop a high quality open source software platform called GUDHI which is becoming a reference in geometric and topological data analysis in high dimensions. The goal is not to provide code tailored to the numerous potential applications but rather to provide the central data structures and algorithms that underlie applications in geometric and topological data analysis.

The development of the GUDHI platform also serves to benchmark and optimize new algorithmic solutions resulting from our theoretical work. Such development necessitates a whole line of research on software architecture and interface design, heuristics and fine-tuning optimization, robustness and arithmetic issues, and visualization. We aim at providing a full programming environment following the same recipes that made up the success story of the CGAL library, the reference library in computational geometry.

Some of the algorithms implemented on the platform will also be interfaced to other software platform, such as the R software [1] for statistical computing, and languages such as Python in order to make them usable in combination with other data analysis and machine learning tools. A first attempt in this direction has been done with the creation of an R package called TDA in collaboration with the group of Larry Wasserman at Carnegie Mellon University (Inria Associated team CATS) that already includes some functionalities of the GUDHI library and implements some joint results between our team and the CMU team. A similar interface with the Python language is also considered a priority. To go even further towards helping users, we will provide utilities that perform the most common tasks without requiring any programming at all.

4. New Software and Platforms

4.1. GUDHI

Geometric Understanding in Higher Dimensions

KEYWORDS: Computational geometry - Topology

SCIENTIFIC DESCRIPTION: The current release of the GUDHI library includes: – Data structures to represent, construct and manipulate simplicial and cubical complexes. – Algorithms to compute simplicial complexes from point cloud data. – Algorithms to compute persistent homology and multi-field persistent homology. – Simplification methods via implicit representations.

FUNCTIONAL DESCRIPTION: The GUDHI open source library will provide the central data structures and algorithms that underly applications in geometry understanding in higher dimensions. It is intended to both help the development of new algorithmic solutions inside and outside the project, and to facilitate the transfer of results in applied fields.

[1] https://www.r-project.org/

NEWS OF THE YEAR: - Cover complex - Representation of persistence diagrams - Cech complex - weighted periodic 3d alpha-complex - sparse Rips complex - debian / docker / conda-forge packages

- Participants: Clément Maria, François Godi, David Salinas, Jean-Daniel Boissonnat, Marc Glisse, Mariette Yvinec, Pawel Dlotko, Siargey Kachanovich, Vincent Rouvreau, Mathieu Carrière and Bertrand Michel
- Contact: Jean-Daniel Boissonnat
- URL: https://gudhi.inria.fr/

4.2. CGAL module: interval arithmetics

KEYWORD: Arithmetic

FUNCTIONAL DESCRIPTION: This package of CGAL (Computational Geometry Algorithms Library http://www.cgal.org) provides an efficient number type for intervals of double and the corresponding arithmetic operations. It is used in the evaluation of geometric predicates for a first quick computation, which either provides the result with guarantees, or rarely answers that more precision is needed.

RELEASE FUNCTIONAL DESCRIPTION: Partial rewrite to take advantage of SIMD instructions on recent x86 processors.

- Contact: Marc Glisse
- URL: https://www.cgal.org/

4.3. CGAL module: interface to Boost.Multiprecision

KEYWORD: Arithmetic

FUNCTIONAL DESCRIPTION: This package of CGAL (Computational Geometry Algorithms Library http://www.cgal.org) makes it possible to use some number types from Boost.Multiprecision in CGAL.

- Author: Marc Glisse
- Contact: Marc Glisse
- URL: https://www.cgal.org/

4.4. Module CGAL: New dD Geometry Kernel

KEYWORD: Computational geometry

FUNCTIONAL DESCRIPTION: This package of CGAL (Computational Geometry Algorithms Library http://www.cgal.org) provides the basic geometric types (point, vector, etc) and operations (orientation test, etc) used by geometric algorithms in arbitrary dimension. It uses filters for efficient exact predicates.

RELEASE FUNCTIONAL DESCRIPTION: New kernel with lazy exact constructions.

- Author: Marc Glisse
- Contact: Marc Glisse
- URL: http://www.cgal.org/

5. New Results

5.1. Algorithmic aspects of topological and geometric data analysis

5.1.1. Sampling and Meshing Submanifolds

Participants: Jean-Daniel Boissonnat, Siargey Kachanovich.

In collaboration with Mathijs Wintraecken (IST Autria).

This work [41], [11] presents a rather simple tracing algorithm to sample and mesh an m-dimensional submanifold of $\mathbb{R}^d$ for arbitrary m and d. We extend the work of Dobkin et al. to submanifolds of arbitrary dimension and codimension. The algorithm is practical and has been thoroughly investigated from both theoretical and experimental perspectives. The paper provides a full description and analysis of the data structure and of the tracing algorithm. The main contributions are : 1. We unify and complement the knowledge about Coxeter and Freudenthal-Kuhn triangulations. 2. We introduce an elegant and compact data structure to store Coxeter or Freudenthal-Kuhn triangulations and describe output sensitive algorithms to compute faces and cofaces or any simplex in the triangulation. 3. We present a manifold tracing algorithm based on the above data structure. We provide a detailled complexity analysis along with experimental results that show that the algorithm can handle cases that are far ahead of the state-of-the-art.

5.1.2. Topological correctness of PL-approximations of isomanifolds

Participant: Jean-Daniel Boissonnat.

In collaboration with Mathijs Wintraecken (IST Autria).

Isomanifolds are the generalization of isosurfaces to arbitrary dimension and codimension, i.e. manifolds defined as the zero set of some multivariate multivalued function $f : \mathbb{R}^d \to \mathbb{R}^{d-n}$. A natural (and efficient) way to approximate an isomanifold is to consider its Piecewise-Linear (PL) approximation based on a triangulation $\mathcal{T}$ of the ambient space $\mathbb{R}^d$. In this paper [43], we give conditions under which the PL-approximation of an isomanifold is topologically equivalent to the isomanifold. The conditions are easy to satisfy in the sense that they can always be met by taking a sufficiently fine triangulation $\mathcal{T}$. This contrasts with previous results on the triangulation of manifolds where, in arbitrary dimensions, delicate perturbations are needed to guarantee topological correctness, which leads to strong limitations in practice. We further give a bound on the Fréchet distance between the original isomanifold and its PL-approximation. Finally we show analogous results for the PL-approximation of an isomanifold with boundary.

5.1.3. Dimensionality Reduction for k-Distance Applied to Persistent Homology

Participants: Jean-Daniel Boissonnat, Kunal Dutta.

In collaboration with Shreya Arya (Duke University)

Given a set P of n points and a constant k, we are interested in computing the persistent homology of the Čech filtration of P for the k-distance, and investigate the effectiveness of dimensionality reduction for this problem, answering an open question of Sheehy [*Proc. SoCG, 2014*] [38]. We first show using the Johnson-Lindenstrauss lemma, that the persistent homology can be preserved up to a $(1 \pm \epsilon)$ factor while reducing dimensionality to $O(k \log n/\varepsilon^2)$. Our main result shows that the target dimension can be improved to $O(\log n/\varepsilon^2)$ under a reasonable and naturally occuring condition. The proof involves a multi-dimensional variant of the Hanson-Wright inequality for subgaussian quadratic forms and works when the random matrices are used for the Johnson-Lindenstrauss mapping are subgaussian. This includes the Gaussian matrices of Indyk-Motwani, the sparse random matrices of Achlioptas and the Ailon-Chazelle fast Johnson-Lindenstrauss transform. To provide evidence that our condition encompasses quite general situations, we show that it is satisfied when the points are independently distributed (i) in $\mathbb{R}^D$ under a subgaussian distribution, or (ii) on a spherical shell in $\mathbb{R}^D$ with a minimum angular separation, using Gershgorin's theorem. Our results also show that the JL-mapping preserves up to a $(1 \pm \epsilon)$ factor, the Rips and Delaunay filtrations for the k-distance, as well as the Čech filtration for the approximate k-distance of Buchet et al.

5.1.4. Edge Collapse and Persistence of Flag Complexes

Participants: Jean-Daniel Boissonnat, Siddharth Pritam.

In this article [42], we extend the notions of dominated vertex and strong collapse of a simplicial complex as introduced by J. Barmak and E. Miniam adn build on the initial success of [30]. We say that a simplex (of any dimension) is dominated if its link is a simplicial cone. Domination of edges appear to be very powerful and we study it in the case of flag complexes in more detail. We show that edge collapse (removal of dominated edges) in a flag complex can be performed using only the 1-skeleton of the complex. Furthermore, the residual complex is a flag complex as well. Next we show that, similar to the case of strong collapses, we can use edge collapses to reduce a flag filtration $\mathcal{F}$ to a smaller flag filtration $\mathcal{F}^c$ with the same persistence. Here again, we only use the 1-skeletons of the complexes. The resulting method to compute $\mathcal{F}^c$ is simple and extremely efficient and, when used as a preprocessing for Persistence Computation, leads to gains of several orders of magnitude wrt the state-of-the-art methods (including our previous approach using strong collapse). The method is exact, irrespective of dimension, and improves performance of Persistence Computation even in low dimensions. This is demonstrated by numerous experiments on publicly available data.

5.1.5. DTM-based Filtrations

Participants: Frédéric Chazal, Marc Glisse, Raphael Tinarrage.

In collaboration with Anai, Hirokazu and Ike, Yuichi and Inakoshi, Hiroya and Umeda, Yuhei (Fujitsu Labs).

Despite strong stability properties, the persistent homology of filtrations classically used in Topological Data Analysis, such as, e.g. the Čech or Vietoris-Rips filtrations, are very sensitive to the presence of outliers in the data from which they are computed. In [15], we introduce and study a new family of filtrations, the DTM-filtrations, built on top of point clouds in the Euclidean space which are more robust to noise and outliers. The approach adopted in this work relies on the notion of distance-to-measure functions, and extends some previous work on the approximation of such functions.

5.1.6. Recovering the homology of immersed manifolds

Participant: Raphael Tinarrage.

Given a sample of an abstract manifold immersed in some Euclidean space, in [57], we describe a way to recover the singular homology of the original manifold. It consists in estimating its tangent bundle -seen as subset of another Euclidean space- in a measure theoretic point of view, and in applying measure-based filtrations for persistent homology. The construction we propose is consistent and stable, and does not involve the knowledge of the dimension of the manifold.

5.1.7. Regular triangulations as lexicographic optimal chains

Participant: David Cohen-Steiner.

In collaboration with André Lieutier and Julien Vuillamy (Dassault Systèmes).

We introduce [46] a total order on n-simplices in the n-Euclidean space for which the support of the lexicographic-minimal chain with the convex hull boundary as boundary constraint is precisely the n-dimensional Delaunay triangulation, or in a more general setting, the regular triangulation of a set of weighted points. This new characterization of regular and Delaunay triangulations is motivated by its possible generalization to submanifold triangulations as well as the recent development of polynomial-time triangulation algorithms taking advantage of this order.

5.1.8. Discrete Morse Theory for Computing Zigzag Persistence

Participant: Clément Maria.

In collaboration with Hannah Schreiber (Graz University of Technology, Austria)

We introduce a framework to simplify zigzag filtrations of general complexes using discrete Morse theory, in order to accelerate the computation of zigzag persistence. Zigzag persistence is a powerful algebraic generalization of persistent homology. However, its computation is much slower in practice, and the usual optimization techniques cannot be used to compute it. Our approach is different in that it preprocesses the filtration before computation. Using discrete Morse theory, we get a much smaller zigzag filtration with same persistence. The new filtration contains general complexes. We introduce new update procedures to modify on the fly the algebraic data (the zigzag persistence matrix) under the new combinatorial changes induced by the Morse reduction. Our approach is significantly faster in practice [35].

5.1.9. Computing Persistent Homology with Various Coefficient Fields in a Single Pass

Participants: Jean-Daniel Boissonnat, Clément Maria.

This article [18] introduces an algorithm to compute the persistent homology of a filtered complex with various coefficient fields in a single matrix reduction. The algorithm is output-sensitive in the total number of distinct persistent homological features in the diagrams for the different coefficient fields. This computation allows us to infer the prime divisors of the torsion coefficients of the integral homology groups of the topological space at any scale, hence furnishing a more informative description of topology than persistence in a single coefficient field. We provide theoretical complexity analysis as well as detailed experimental results. The code is part of the Gudhi software library.

5.1.10. Exact computation of the matching distance on 2-parameter persistence modules

Participant: Steve Oudot.

In collaboration with Michael Kerber (T.U. Graz) and Michael Lesnick (SUNY).

The matching distance is a pseudometric on multi-parameter persistence modules, defined in terms of the weighted bottleneck distance on the restriction of the modules to affine lines. It is known that this distance is stable in a reasonable sense, and can be efficiently approximated, which makes it a promising tool for practical applications. In [31] we show that in the 2-parameter setting, the matching distance can be computed exactly in polynomial time. Our approach subdivides the space of affine lines into regions, via a line arrangement. In each region, the matching distance restricts to a simple analytic function, whose maximum is easily computed. As a byproduct, our analysis establishes that the matching distance is a rational number, if the bigrades of the input modules are rational.

5.1.11. Decomposition of exact pfd persistence bimodules

Participant: Steve Oudot.

In collaboration with Jérémy Cochoy (Symphonia).

In [24] we identify a certain class of persistence modules indexed over $\mathbb{R}^2$ that are decomposable into direct sums of indecomposable summands called blocks. The conditions on the modules are that they are both pointwise finite-dimensional (pfd) and exact. Our proof follows the same scheme as the one for pfd persistence modules indexed over $\mathbb{R}$, yet it departs from it at key stages due to the product order not being a total order on $\mathbb{R}^2$, which leaves some important gaps open. These gaps are filled in using more direct arguments. Our work is motivated primarily by the study of interlevel-sets persistence, although the proposed results reach beyond that setting.

5.1.12. Level-sets persistence and sheaf theory

Participants: Nicolas Berkouk, Steve Oudot.

In collaboration with Grégory Ginot (Paris 13).

In [39] we provide an explicit connection between level-sets persistence and derived sheaf theory over the real line. In particular we construct a functor from 2-parameter persistence modules to sheaves over R, as well as a functor in the other direction. We also observe that the 2-parameter persistence modules arising from the level sets of Morse functions carry extra structure that we call a Mayer-Vietoris system. We prove classification, barcode decomposition, and stability theorems for these Mayer-Vietoris systems, and we show that the aforementioned functors establish a pseudo-isometric equivalence of categories between derived constructible sheaves with the convolution or (derived) bottleneck distance and the interleaving distance of strictly pointwise finite-dimensional Mayer-Vietoris systems. Ultimately, our results provide a functorial equivalence between level-sets persistence and derived pushforward for continuous real-valued functions.

5.1.13. Intrinsic Interleaving Distance for Merge Trees

Participant: Steve Oudot.

In collaboration with Ellen Gasparovic (Union College), Elizabeth Munch (Michigan State), Katharine Turner (Australian National University), Bei Wang (Utah), and Yusu Wang (Ohio-State).

Merge trees are a type of graph-based topological summary that tracks the evolution of connected components in the sublevel sets of scalar functions. They enjoy widespread applications in data analysis and scientific visualization. In [49] we consider the problem of comparing two merge trees via the notion of interleaving distance in the metric space setting. We investigate various theoretical properties of such a metric. In particular, we show that the interleaving distance is intrinsic on the space of labeled merge trees and provide an algorithm to construct metric 1-centers for collections of labeled merge trees. We further prove that the intrinsic property of the interleaving distance also holds for the space of unlabeled merge trees. Our results are a first step toward performing statistics on graph-based topological summaries.

5.2. Statistical aspects of topological and geometric data analysis

5.2.1. Estimating the Reach of a Manifold

Participants: Frédéric Chazal, Jisu Kim, Bertrand Michel.

In collaboration with E. Aamari (Univ. Paris-Diderot), A. Rinaldo, L. Wasserman (Carnegie Mellon University).

In [13], various problems in manifold estimation make use of a quantity called the reach, denoted by τ_M, which is a measure of the regularity of the manifold. This paper is the first investigation into the problem of how to estimate the reach. First, we study the geometry of the reach through an approximation perspective. We derive new geometric results on the reach for submanifolds without boundary. An estimator $\widehat{\tau}$ of τ_M is proposed in an oracle framework where tangent spaces are known, and bounds assessing its efficiency are derived. In the case of i.i.d. random point cloud X_n, $\widehat{\tau}(X_n)$ is showed to achieve uniform expected loss bounds over a C^3-like model. Finally, we obtain upper and lower bounds on the minimax rate for estimating the reach.

5.2.2. A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature

Participant: Claire Brecheteau.

In [20], we introduce the notion of DTM-signature, a measure on $\mathbb{R}$ that can be associated to any metric-measure space. This signature is based on the function distance to a measure (DTM) introduced in 2009 by Chazal, Cohen-Steiner and Mérigot. It leads to a pseudo-metric between metric-measure spaces, that is bounded above by the Gromov-Wasserstein distance. This pseudo-metric is used to build a statistical test of isomorphism between two metric-measure spaces, from the observation of two N-samples.

The test is based on subsampling methods and comes with theoretical guarantees. It is proven to be of the correct level asymptotically. Also, when the measures are supported on compact subsets of $\mathbb{R}^d$, rates of convergence are derived for the $L1$-Wasserstein distance between the distribution of the test statistic and its subsampling approximation. These rates depend on some parameter $\rho > 1$. In addition, we prove that the power is bounded above by $\exp(-CN1/\rho)$, with C proportional to the square of the aforementioned pseudo-metric between the metric-measure spaces. Under some geometrical assumptions, we also derive lower bounds for this pseudo-metric.

An algorithm is proposed for the implementation of this statistical test, and its performance is compared to the performance of other methods through numerical experiments.

5.2.3. On the choice of weight functions for linear representations of persistence diagrams

Participant: Vincent Divol.

In collaboration with Wolfgang Polonik (UC Davis).

Persistence diagrams are efficient descriptors of the topology of a point cloud. As they do not naturally belong to a Hilbert space, standard statistical methods cannot be directly applied to them. Instead, feature maps (or representations) are commonly used for the analysis. A large class of feature maps, which we call linear, depends on some weight functions, the choice of which is a critical issue. An important criterion to choose a weight function is to ensure stability of the feature maps with respect to Wasserstein distances on diagrams. In [21], we improve known results on the stability of such maps, and extend it to general weight functions. We also address the choice of the weight function by considering an asymptotic setting; assume that $\mathbb{X}_n$ is an i.i.d. sample from a density on $[0,1]^d$. For the Č ech and Rips filtrations, we characterize the weight functions for which the corresponding feature maps converge as n approaches infinity, and by doing so, we prove laws of large numbers for the total persistences of such diagrams. Those two approaches (stability and convergence) lead to the same simple heuristic for tuning weight functions: if the data lies near a d-dimensional manifold, then a sensible choice of weight function is the persistence to the power α with $\alpha \geq d$.

5.2.4. Understanding the Topology and the Geometry of the Persistence Diagram Space via Optimal Partial Transport

Participants: Vincent Divol, Théo Lacombe.

Despite the obvious similarities between the metrics used in topological data analysis and those of optimal transport, an optimal-transport based formalism to study persistence diagrams and similar topological descriptors has yet to come. In [48], by considering the space of persistence diagrams as a measure space, and by observing that its metrics can be expressed as solutions of optimal partial transport problems, we introduce a generalization of persistence diagrams, namely Radon measures supported on the upper half plane. Such measures naturally appear in topological data analysis when considering continuous representations of persistence diagrams (e.g. persistence surfaces) but also as limits for laws of large numbers on persistence diagrams or as expectations of probability distributions on the persistence diagrams space. We study the topological properties of this new space, which will also hold for the closed subspace of persistence diagrams. New results include a characterization of convergence with respect to transport metrics, the existence of Fréchet means for any distribution of diagrams, and an exhaustive description of continuous linear representations of persistence diagrams. We also showcase the usefulness of this framework to study random persistence diagrams by providing several statistical results made meaningful thanks to this new formalism.

5.3. Topological approach for multimodal data processing

5.3.1. A General Neural Network Architecture for Persistence Diagrams and Graph Classification

Participants: Frédéric Chazal, Théo Lacombe, Martin Royer.

In collaboration with Mathieu Carrière (Colombia Univ.) and Umeda Yuhei and Ike Yiuchi (Fujitsu Labs).

Persistence diagrams, the most common descriptors of Topological Data Analysis, encode topological properties of data and have already proved pivotal in many different applications of data science. However, since the (metric) space of persistence diagrams is not Hilbert, they end up being difficult inputs for most Machine Learning techniques. To address this concern, several vectorization methods have been put forward that embed persistence diagrams into either finite-dimensional Euclidean space or (implicit) infinite dimensional Hilbert space with kernels. In [44], we focus on persistence diagrams built on top of graphs. Relying on extended persistence theory and the so-called heat kernel signature, we show how graphs can be encoded by (extended) persistence diagrams in a provably stable way. We then propose a general and versatile framework for learning vectorizations of persistence diagrams, which encompasses most of the vectorization techniques used in the literature. We finally showcase the experimental strength of our setup by achieving competitive scores on classification tasks on real-life graph datasets.

5.3.2. Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks

Participant: Frédéric Chazal.

In collaboration with Umeda Yuhei and Meryll Dindin (Fujitsu Labs).

In [47], we present an innovative and generic deep learning approach to monitor heart conditions from ECG signals.We focus our attention on both the detection and classification of abnormal heartbeats, known as arrhythmia. We strongly insist on generalization throughout the construction of a deep-learning model that turns out to be effective for new unseen patient. The novelty of our approach relies on the use of topological data analysis as basis of our multichannel architecture, to diminish the bias due to individual differences. We show that our structure reaches the performances of the state-of-the-art methods regarding arrhythmia detection and classification.

5.3.3. ATOL: Automatic Topologically-Oriented Learning

Participants: Frédéric Chazal, Martin Royer.

In collaboration with Umeda Yuhei and Ike Yiuchi (Fujitsu Labs).

There are abundant cases for using Topological Data Analysis (TDA) in a learning context, but robust topological information commonly comes in the form of a set of persistence diagrams, objects that by nature are uneasy to affix to a generic machine learning framework. In [56], we introduce a vectorisation method for diagrams that allows to collect information from topological descriptors into a format fit for machine learning tools. Based on a few observations, the method is learned and tailored to discriminate the various important plane regions a diagram is set into. With this tool one can automatically augment any sort of machine learning problem with access to a TDA method, enhance performances, construct features reflecting underlying changes in topological behaviour. The proposed methodology comes with only high level tuning parameters such as the encoding budget for topological features. We provide an open-access, ready-to-use implementation and notebook. We showcase the strengths and versatility of our approach on a number of applications. From emulous and modern graph collections to a highly topological synthetic dynamical orbits data, we prove that the method matches or beats the state-of-the-art in encoding persistence diagrams to solve hard problems. We then apply our method in the context of an industrial, difficult time-series regression problem and show the approach to be relevant.

5.3.4. Inverse Problems in Topological Persistence: a Survey

Participant: Steve Oudot.

In collaboration with Elchanan Solomon (Duke).

In [27] we review the literature on inverse problems in topological persistence theory.The first half of the survey is concerned with the question of surjectivity, i.e. the existence of rightinverses, and the second half focuses on injectivity, i.e. left inverses. Throughout, we highlightthe tools and theorems that underlie these advances, and direct the reader's attention to openproblems, both theoretical and applied.

5.3.5. Intrinsic Topological Transforms via the Distance Kernel Embedding

Participants: Clément Maria, Steve Oudot.

In collaboration with Elchanan Solomon (Duke).

Topological transforms are parametrized families of topological invariants, which, by analogy with transforms in signal processing, are much more discriminative than single measurements. The first two topological transforms to be defined were the Persistent Homology Transform and Euler Characteristic Transform, both of which apply to shapes embedded in Euclidean space. The contribution of this work [54] is to define topological transforms that depend only on the intrinsic geometry of a shape, and hence are invariant to the choice of embedding. To that end, given an abstract metric measure space, we define an integral operator whose eigenfunctions are used to compute sublevel set persistent homology. We demonstrate that this operator, which we call the distance kernel operator, enjoys desirable stability properties, and that its spectrum and eigenfunctions concisely encode the large-scale geometry of our metric measure space. We then define a number of topological transforms using the eigenfunctions of this operator, and observe that these transforms inherit many of the stability and injectivity properties of the distance kernel operator.

5.3.6. A Framework for Differential Calculus on Persistence Barcodes

Participant: Steve Oudot.

In collaboration with Jacob Leygonie and Ulrike Tillmann (Oxford).

In [52], we define notions of differentiability for maps from and to the space of persistence barcodes. Inspired by the theory of diffeological spaces, the proposed framework uses lifts to the space of ordered barcodes, from which derivatives can be computed. The two derived notions of differentiability (respectively from and to the space of barcodes) combine together naturally to produce a chain rule that enables the use of gradient descent for objective functions factoring through the space of barcodes. We illustrate the versatility of this framework by showing how it can be used to analyze the smoothness of various parametrized families of filtrations arising in topological data analysis.

5.4. Experimental research and software development

5.4.1. Robust Stride Detector from Ankle-Mounted Inertial Sensors for Pedestrian Navigation and Activity Recognition with Machine Learning Approaches

Participants: Bertrand Beaufils, Frédéric Chazal, Bertrand Michel.

In collaboration with Marc Grelet (Sysnav).

In [16], a stride detector algorithm combined with a technique inspired by zero velocity update (ZUPT) is proposed to reconstruct the trajectory of a pedestrian from an ankle-mounted inertial device. This innovative approach is based on sensor alignment and machine learning. It is able to detect 100% of both normal walking strides and more than 97% of atypical strides such as small steps, side steps, and backward walking that existing methods can hardly detect. This approach is also more robust in critical situations, when for example the wearer is sitting and moving the ankle or when the wearer is bicycling (less than two false detected strides per hour on average). As a consequence, the algorithm proposed for trajectory reconstruction achieves much better performances than existing methods for daily life contexts, in particular in narrow areas such as in a house. The computed stride trajectory contains essential information for recognizing the activity (atypical stride, walking, running, and stairs). For this task, we adopt a machine learning approach based on descriptors of these trajectories, which is shown to be robust to a large of variety of gaits. We tested our algorithm on recordings of healthy adults and children, achieving more than 99% success. The algorithm also achieved more than 97by children suffering from movement disorders. Compared to most algorithms in the literature, this original method does not use a fixed-size sliding window but infers this last in an adaptive way

5.4.2. Robust pedestrian trajectory reconstruction from inertial sensor

Participants: Bertrand Beaufils, Frédéric Chazal, Bertrand Michel.

In collaboration with Marc Grelet (Sysnav).

In [28], a strides detection algorithm combined with a technique inspired by Zero Velocity Update (ZUPT) is proposed using inertial sensors worn on the ankle. This innovative approach based on a sensors alignment and machine learning can detect both normal walking strides and atypical strides such as small steps, side steps and backward walking that existing methods struggle to detect. As a consequence, the trajectory reconstruction achieves better performances in daily life contexts for example, where a lot of these kinds of strides are performed in narrow areas such as in a house. It is also robust in critical situations, when for example the wearer is sitting and moving the ankle or bicycling, while most algorithms in the literature would wrongly detect strides and produce error in the trajectory reconstruction by generating movements.Our algorithm is evaluated on more than 7800 strides from seven different subjects performing several activities. We validated the trajectory reconstruction during motion capture sessions by analyzing the stride length. Finally, we tested the algorithm in a challenging situation by plotting the computed trajectory on the building map of an 5 hours and 30 minutes office worker recording.

5.5. Algorithmic and Combinatorial Aspects of Low Dimensional Topology

5.5.1. Treewidth, crushing and hyperbolic volume

Participant: Clément Maria.

In collaboration with Jessica S. Purcell (Monash University, Australia)

The treewidth of a 3-manifold triangulation plays an important role in algorithmic 3-manifold theory, and so it is useful to find bounds on the tree-width in terms of other properties of the manifold. In [26], we prove that there exists a universal constant c such that any closed hyperbolic 3-manifold admits a triangulation of tree-width at most the product of c and the volume. The converse is not true: we show there exists a sequence of hyperbolic 3-manifolds of bounded tree-width but volume approaching infinity. Along the way, we prove that crushing a normal surface in a triangulation does not increase the carving-width, and hence crushing any number of normal surfaces in a triangulation affects tree-width by at most a constant multiple.

5.5.2. Parameterized complexity of quantum knot invariants

Participant: Clément Maria.

In [53], we give a general fixed parameter tractable algorithm to compute quantum invariants of links presented by diagrams, whose complexity is singly exponential in the carving-width (or the tree-width) of the diagram. In particular, we get a $O(N^{3/2\mathrm{cw}}\mathrm{poly}(n))$ time algorithm to compute any Reshetikhin-Turaev invariant-derived from a simple Lie algebra g of a link presented by a planar diagram with n crossings and carving-width cw, and whose components are coloured with g-modules of dimension at most N. For example, this includes the Nth-coloured Jones polynomial and the Nth-coloured HOMFLYPT polynomial.

5.6. Miscellaneous

5.6.1. Material Coherence from Trajectories via Burau Eigenanalysis of Braids

Participant: David Cohen-Steiner.

In collaboration with Melissa Yeung and Mathieu Desbrun (Caltech).

In this paper [58], we provide a numerical tool to study material coherence from a set of 2D Lagrangian trajectories sampling a dynamical system, i.e., from the motion of passive tracers. We show that eigenvectors of the Burau representation of a topological braid derived from the trajectories have levelsets corresponding to components of the Nielsen-Thurston decomposition of the dynamical system. One can thus detect and identify clusters of space-time trajectories corresponding to coherent regions of the dynamical system by solving an eigenvalue problem. Unlike previous methods, the scalable computational complexity of our braid-based approach allows the analysis of large amounts of trajectories. Studying two-dimensional flows and their induced transport and mixing properties is key to geophysical studies of atmospheric and oceanic processes. However, one often has only sparse tracer trajectories (e.g., positions of buoys in time) to infer the overall flow geometry. Fortunately, topological methods based on the theory of braid groups have recently been proposed to extract structures from such a sparse set of trajectories by measuring their entan-glement. This braid viewpoint offers sound foundations for the definition of coherent structures. Yet, there has been only limited efforts in developing practical tools that can leverage topological properties for the efficient analysis of flow structures: handling a larger number of tra-jectories remains computationally challenging. We contribute a new and simple computational tool to extract Lagrangian structures from sparse trajectories by noting that the eigenstructure of the Burau matrix representation of a braid of particle trajectories can be used to reveal coherent regions of the flows. Detection of clusters of space-time trajectories corresponding to coherent regions of the dynamical system can thus be achieved by solving a simple eigenvalue problem. This paper establishes the theoretical foundations behind this braid eigenanalysis approach, along with numerical validations on various flows.

5.6.2. Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space

Participants: Alex Delalande, Frédéric Chazal.

In collaboration with Quentin Mérigot (Institut de Mathématiques d'Orsay).

In [55], we study an explicit embedding of the set of probability measures into a Hilbert space, defined using optimal transport maps from a reference probability density. This embedding linearizes to some extent the 2-Wasserstein space, and enables the direct use of generic supervised and unsupervised learning algorithms on measure data. Our main result is that the embedding is (bi-)Holder continuous, when the reference density is uniform over a convex set, and can be equivalently phrased as a dimension-independent Hölder-stability results for optimal transport maps.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

- Collaboration with Sysnav, a French SME with world leading expertise in navigation and geopositioning in extreme environments, on TDA, geometric approaches and machine learning for the analysis of movements of pedestrians and patients equipped with inetial sensors (CIFRE PhD of Bertrand Beaufils).
- Research collaboration with Fujitsu on the development of new TDA methods and tools for Machine learning and Artificial Intelligence (started in Dec 2017).
- Research collaboration with MetaFora on the development of new TDA-based and statistical methods for the analysis of cytometric data (started in Nov. 2019).

6.2. Bilateral Grants with Industry

- DATASHAPE and Sysnav have been selected for the ANR/DGA Challenge MALIN (funding: 700 kEuros) on pedestrian motion reconstruction in severe environments (without GPS access).

7. Partnerships and Cooperations

7.1. Regional Initiatives

Mini course on "Sheaf Theory and Topological Data Analysis" taught by Rodrigo Cordoniu (Nice University) at Inria Sophia Antipolis — 8 weeks, 2h per week, Feb 2019 to Apr 2019.

7.2. National Initiatives

7.2.1. ANR

7.2.1.1. ANR ASPAG

Participant: Marc Glisse.

- Acronym : ASPAG.

- Type : ANR blanc.

- Title : Analysis and Probabilistic Simulations of Geometric Algorithms.

- Coordinator : Olivier Devillers (équipe Inria Gamble).

- Duration : 4 years from January 2018 to December 2021.

- Others Partners: Inria Gamble, LPSM, LABRI, Université de Rouen, IECL, Université du Littoral Côte d'Opale, Telecom ParisTech, Université Paris X (Modal'X), LAMA, Université de Poitiers, Université de Bourgogne.

- Abstract:

The analysis and processing of geometric data has become routine in a variety of human activities ranging from computer-aided design in manufacturing to the tracking of animal trajectories in ecology or geographic information systems in GPS navigation devices. Geometric algorithms and probabilistic geometric models are crucial to the treatment of all this geometric data, yet the current available knowledge is in various ways much too limited: many models are far from matching real data, and the analyses are not always relevant in practical contexts. One of the reasons for this state of affairs is that the breadth of expertise required is spread among different scientific communities (computational geometry, analysis of algorithms and stochastic geometry) that historically had very little interaction. The Aspag project brings together experts of these communities to address the problem of geometric data. We will more specifically work on the following three interdependent directions.

(1) Dependent point sets: One of the main issues of most models is the core assumption that the data points are independent and follow the same underlying distribution. Although this may be relevant in some contexts, the independence assumption is too strong for many applications.

(2) Simulation of geometric structures: The phenomena studied in (1) involve intricate random geometric structures subject to new models or constraints. A natural first step would be to build up our understanding and identify plausible conjectures through simulation. Perhaps surprisingly, the tools for an effective simulation of such complex geometric systems still need to be developed.

(3) Understanding geometric algorithms: the analysis of algorithm is an essential step in assessing the strengths and weaknesses of algorithmic principles, and is crucial to guide the choices made when designing a complex data processing pipeline. Any analysis must strike a balance between realism and tractability; the current analyses of many geometric algorithms are notoriously unrealistic. Aside from the purely scientific objectives, one of the main goals of Aspag is to bring the communities closer in the long term. As a consequence, the funding of the project is crucial to ensure that the members of the consortium will be able to interact on a very regular basis, a necessary condition for significant progress on the above challenges.

- See also: https://members.loria.fr/Olivier.Devillers/aspag/

7.3. International Research Visitors

7.3.1. Visits of International Scientists

- Arijit Ghosh, Indian Statistical Institute, Kolkata, India (September 2019)
- Ramsay Dyer Berkeley Publishing (September 2019)
- Mathijs Wintraecken, IST Austria (September and October 2019)

7.3.1.1. Internships

- Alex Delalande, Centrale-Supelec, (May-October 2019).

7.3.1.2. Research Stays Abroad

- Martin Royer, Fujitsu Laboratories, Tokyo, 2 months.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events: Organisation

- F. Chazal was co-organiser of the Geometric Data Analysis conference at the Stevanovich Center for Financial Mathematics, Chicago, May 2019.
- F. Chazal was co-organiser of the conference New Horizons in Computational Geometry and Topology, Sophia-Antipolis, September 2019.

8.1.2. Scientific Events: Selection

8.1.2.1. Chair of Conference Program Committees

- Steve Oudot was chair of the program committee for the Young Researchers Forum at the International Symposium on Computational Geometry (SoCG) 2019.

8.1.2.2. Member of the Conference Program Committees

- Jean-Daniel Boissonnat was a member of the Organizing Committee of the International Conference on Curves and Surfaces 2019.
- Clément Maria was a member of the Program Committee of the International Symposium on Computational Geometry (SoCG), Portland, USA, June 2019.
- David Cohen-Steiner was a member of the Program Committees of SGP'19 and SMI'19.

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

Jean-Daniel Boissonnat is a member of the Editorial Board of *Journal of the ACM*, *Discrete and Computational Geometry*, *International Journal on Computational Geometry and Applications*.

Frédéric Chazal is a member of the Editorial Board of *SIAM Journal on Imaging Sciences*, *Discrete and Computational Geometry (Springer)*, *Graphical Models (Elsevier), and Journal of Applied and Computational Topology (Springer)*.

Frédéric Chazal is a member of the Scientific Board of Journal of Applied and Computational Topology (Springer).

Steve Oudot is a member of the Editorial Board of *Journal of Computational Geometry*.

8.1.4. Invited Talks

Frédéric Chazal, Geometry of Big Data, Institute of Pure and Applied Mathematics at UCLA, Los Angeles, April 2019.

Frédéric Chazal, Workshop on Data Science, Fundaçao Getulio Vargas, Rio de Janeiro, April 2019.

Frédéric Chazal, SUMTOPO 2019, Topology in Data Science, Johannesburg, June 2019.

Frédéric Chazal, New Horizons in Computational Geometry and Topology, Sophia-Antipolis, September 2019.

Frédéric Chazal, Seminar@System X, Saclay, June 2019.

Clément Maria, Monash University, Australia.

Clément Maria, New Horizons in Computational Geometry and Topology, Sophia-Antipolis, September 2019.

Clément Maria, Applied Algebraic Topology Research Network Seminar, October 2019.

Clément Maria, Dagstuhl Seminar 19352 "Computation in Low-Dimensional Geometry and Topology", August 2019.

Clément Maria, 2ème Demi-journée Pole Calcul - Topologie et Géométrie du Calcul - LIS Marseille, June 2019.

Steve Oudot, Workshop on computational applications of quiver representations: TDA and QPA, Bielefeld University, April 2019.

Steve Oudot, Minisymposium on Algebraic Geometry in Topological Data Analysis, SIAM Conference on Applied Algebraic Geometry, Bern, July 2019.

Steve Oudot, Summer Conference on Topology and its Applications, Johannesburg, July 2019.

Steve Oudot, Spires workshop: from theory to applications of TDA, Oxford, September 2019.

8.1.5. Leadership within the Scientific Community

Frédéric Chazal is co-responsible, with S. Arlot (Paris-Sud Univ.), of the “programme Maths-STIC” of the Labex Fondation Mathématique Jacques Hadamard (FMJH).

Frédéric Chazal is a member of the “Comité de pilotage” of the SIGMA group at SMAI.

Steve Oudot is co-head with Luca Castelli Aleardi (École polytechnique) of the GT GeoAlgo within the GdR-IM.

8.1.6. Research Administration

Marc Glisse is president of the CDT at Inria Saclay.

Steve Oudot is vice-president of the Commission Scientifique at Inria Saclay.

Clément Maria is a member of the CDT at Inria Sophia Antipolis-Méditerranée.

Clément Maria, responsable RAweb pour Datashape.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Master: Frédéric Chazal and Quentin Mérigot, Analyse Topologique des Données, 30h eq-TD, Université Paris-Sud, France.

Master: Marc Glisse and Clément Maria, Computational Geometry Learning, 36h eq-TD, M2, MPRI, France.

Master: Frédéric Cazals and Frédéric Chazal, Geometric Methods for Data Analysis, 30h eq-TD, M1, École Centrale Paris, France.

Master: Frédéric Chazal and Julien Tierny, Topological Data Analysis, 38h eq-TD, M2, Mathématiques, Vision, Apprentissage (MVA), ENS Paris-Saclay, France.

Course on TDA at the Statistics with Geometry and Topology Conference 2019, 4hCM, Toulouse, France.

Master: Steve Oudot, Topological data analysis, 45h eq-TD, M1, École polytechnique, France.

Master: Steve Oudot, Data Analysis: geometry and topology in arbitrary dimensions, 24h eq-TD, M2, graduate program in Artificial Intelligence & Advanced Visual Computing, École polytechnique, France.

Undergrad-Master: Steve Oudot, Algorithms for data analysis in C++, 22.5h eq-TD, L3/M1, École polytechnique, France.

8.2.2. Supervision

PhD : Alba Chiara de Vitis, Kernel Methods for High Dimensional Data Analysis [12]. Université Côte d’Azur. May 28, 2019. Jean-Daniel Boissonnat and David Cohen-Steiner.

PhD : Sergei Kachanovich, Meshing submanifolds using Coxeter triangulations [11]. Université Côte d’Azur. October 23, 2019. Jean-Daniel Boissonnat.

PhD: Yitchzak Solomon, Inverse problems in topological data analysis, Brown University. Defended in May 2019. Jeff Brock (Yale University) and Steve Oudot.

PhD in progress: Raphaël Tinarrage, Persistence and stability of nerves in measured metric spaces for Topological Data Analysis, started September 1st, 2017, Frédéric Chazal and Marc Glisse.

PhD in progress: Louis Pujol, Partitionnement de données cytométriques, started Novermber 1st, 2019, Pascal Massart and Marc Glisse.

PhD in progress: Bertrand Beaufils, Méthodes topologiques et apprentissage statistique pour l’actimétrie du piéton à partir de données de mouvement, Frédéric Chazal and Bertrand Michel (Ecole Centrale de Nantes).

PhD in progress: Vincent Divol, statistical aspects of TDA, started September 1st, 2017, Frédéric Chazal and Pascal Massart (LMO).

PhD in progress: Etienne Lasalle, TDA for graph data, started September 1st, 2019, Frédéric Chazal and Pascal Massart (LMO).

PhD in progress: Alex Delalande, Measure embedding with Optimal Transport and applications in Machine Learning, started December 1st, 2019, Frédéric Chazal and Quentin Mérigot (LMO).

PhD in progress: Owen Rouillé, started September 2018, co-advised by C. Maria and J-D. Boissonnat.

PhD in progress: Nicolas Berkouk, Categorification of topological graph structures, started November 1st, 2016, Steve Oudot.

PhD in progress: Théo Lacombe, Statistics for persistence diagrams using optimal transport, started October 1st, 2017. Marco Cuturi (ENSAE & Google Brain) and Steve Oudot.

8.2.3. Juries

Clément Maria was a member of the jury attributing the Gilles Kahn PhD award, from the SIF and the Academy of Science, Nov. 2019.

Steve Oudot was reviewer and committee member for the PhD defence of Hannah Schreiber, October 2019.

8.3. Popularization

8.3.1. Interventions

- Frédéric Chazal, Un exemple de dispositif dédié aux études cliniques. Seminar on AI, Haute Autorité de Santé, September 2019.

9. Bibliography

Major publications by the team in recent years

[1] D. ATTALI, U. BAUER, O. DEVILLERS, M. GLISSE, A. LIEUTIER. *Homological Reconstruction and Simplification in R3*, in "Computational Geometry", 2014 [*DOI :* 10.1016/J.COMGEO.2014.08.010], https://hal.archives-ouvertes.fr/hal-01132440

[2] J.-D. BOISSONNAT, R. DYER, A. GHOSH. *Delaunay Triangulation of Manifolds*, in "Foundations of Computational Mathematics", 2017, vol. 45, 38 p. [*DOI :* 10.1007/S10208-017-9344-1], https://hal.inria.fr/hal-01509888

[3] J.-D. BOISSONNAT, R. DYER, A. GHOSH, S. Y. OUDOT. *Only distances are required to reconstruct submanifolds*, in "Computational Geometry", 2017, vol. 66, pp. 32 - 67 [*DOI :* 10.1016/J.COMGEO.2017.08.001], https://hal.inria.fr/hal-01583086

[4] J.-D. BOISSONNAT, K. C. SRIKANTA, S. TAVENAS. *Building Efficient and Compact Data Structures for Simplicial Complexe*, in "Algorithmica", September 2016 [*DOI :* 10.1007/S00453-016-0207-Y], https://hal.inria.fr/hal-01364648

[5] F. CHAZAL, D. COHEN-STEINER, A. LIEUTIER. *A Sampling Theory for Compact Sets in Euclidean Space*, in "Discrete Comput. Geom.", 2009, vol. 41, n^{o} 3, pp. 461–479, http://dx.doi.org/10.1007/s00454-009-9144-8

[6] F. CHAZAL, D. COHEN-STEINER, Q. MÉRIGOT. *Geometric Inference for Measures based on Distance Functions*, in "Foundations of Computational Mathematics", 2011, vol. 11, n^o 6, pp. 733-751, RR-6930 [*DOI :* 10.1007/s10208-011-9098-0], http://hal.inria.fr/inria-00383685

[7] F. CHAZAL, S. Y. OUDOT, M. GLISSE, V. DE SILVA. *The Structure and Stability of Persistence Modules*, SpringerBriefs in Mathematics, Springer Verlag, 2016, VII, 116 p. , https://hal.inria.fr/hal-01330678

[8] L. J. GUIBAS, S. Y. OUDOT, P. SKRABA, F. CHAZAL. *Persistence-Based Clustering in Riemannian Manifolds*, in "Journal of the ACM", November 2013, vol. 60, n^o 6, 38 p. , http://hal.archives-ouvertes.fr/hal-00923563

[9] M. MANDAD, D. COHEN-STEINER, L. KOBBELT, P. ALLIEZ, M. DESBRUN. *Variance-Minimizing Transport Plans for Inter-surface Mapping*, in "ACM Transactions on Graphics", 2017, vol. 36, 14 p. [*DOI :* 10.1145/3072959.3073671], https://hal.inria.fr/hal-01519006

[10] S. Y. OUDOT. *Persistence Theory: From Quiver Representations to Data Analysis*, Mathematical Surveys and Monographs, American Mathematical Society, 2015, n^o 209, 218 p. , https://hal.inria.fr/hal-01247501

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. KACHANOVICH. *Meshing submanifolds using Coxeter triangulations*, Université Côte d'Azur, October 2019, https://hal.inria.fr/tel-02419148

[12] A. C. DE VITIS. *Kernel Methods for High Dimensional Data Analysis*, Université Côte d'Azur, May 2019, https://hal.inria.fr/tel-02419727

Articles in International Peer-Reviewed Journals

[13] E. AAMARI, J. KIM, F. CHAZAL, B. MICHEL, A. RINALDO, L. WASSERMAN. *Estimating the Reach of a Manifold*, in "Electronic journal of statistics ", April 2019, https://arxiv.org/abs/1705.04565 [*DOI :* 10.1214/19-EJS1551], https://hal.archives-ouvertes.fr/hal-01521955

[14] E. AAMARI, C. LEVRARD. *Non-Asymptotic Rates for Manifold, Tangent Space, and Curvature Estimation*, in "Annals of Statistics", 2019, vol. 47, n^o 1, https://arxiv.org/abs/1705.00989 [*DOI :* 10.1214/18-AOS1685], https://hal.archives-ouvertes.fr/hal-01516032

[15] H. ANAI, F. CHAZAL, M. GLISSE, Y. IKE, H. INAKOSHI, R. TINARRAGE, Y. UMEDA. *DTM-based Filtrations*, in "Abel Symposia", 2019, https://arxiv.org/abs/1811.04757 , forthcoming, https://hal.archives-ouvertes.fr/hal-01919562

[16] B. BEAUFILS, F. CHAZAL, M. GRELET, B. MICHEL. *Robust Stride Detector from Ankle-Mounted Inertial Sensors for Pedestrian Navigation and Activity Recognition with Machine Learning Approaches*, in "Sensors", October 2019, vol. 19, n^o 20, 4491 p. [*DOI :* 10.3390/s19204491], https://hal.inria.fr/hal-02365094

[17] J.-D. BOISSONNAT, A. LIEUTIER, M. WINTRAECKEN. *The reach, metric distortion, geodesic convexity and the variation of tangent spaces*, in "Journal of Applied and Computational Topology", June 2019 [*DOI :* 10.1007/s41468-019-00029-8], https://hal.inria.fr/hal-02266258

[18] J.-D. BOISSONNAT, C. MARIA. *Computing Persistent Homology with Various Coefficient Fields in a Single Pass*, in "Journal of Applied and Computational Topology", September 2019, vol. 3, n^{o} 1-2, 16 p. , https://arxiv.org/abs/2001.02960 [*DOI :* 10.1007/S41468-019-00025-Y], https://hal.inria.fr/hal-00922572

[19] J.-D. BOISSONNAT, M. ROUXEL-LABBÉ, M. WINTRAECKEN. *Anisotropic triangulations via discrete Riemannian Voronoi diagrams*, in "SIAM Journal on Computing", 2019, https://hal.inria.fr/hal-02419460

[20] C. BRÉCHETEAU. *A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature*, in "Electronic journal of statistics ", 2019, MSC: Primary 62G10; secondary 62G09, forthcoming [*DOI :* 10.1214/154957804100000000], https://hal.inria.fr/hal-01426331

[21] V. DIVOL, W. POLONIK. *On the choice of weight functions for linear representations of persistence diagrams*, in "Journal of Applied and Computational Topology", August 2019, https://arxiv.org/abs/1807.03678 [*DOI :* 10.1007/S41468-019-00032-Z], https://hal.inria.fr/hal-01833660

[22] K. DUTTA, A. GHOSH, B. JARTOUX, N. MUSTAFA. *Shallow packings, semialgebraic set systems, Macbeath regions, and polynomial partitioning*, in "Discrete and Computational Geometry", 2019, vol. 61, n^{o} 4, pp. 756-777 [*DOI :* 10.1007/S00454-019-00075-0], https://hal.archives-ouvertes.fr/hal-02316975

[23] S. HARKER, M. KRAMÁR, R. LEVANGER, K. MISCHAIKOW. *A Comparison Framework for Interleaved Persistence Modules*, in "Journal of Applied and Computational Topology", 2019 [*DOI :* 10.1007/S41468-019-00026-X], https://hal.inria.fr/hal-01757092

[24] C. JÉRÉMY, S. Y. OUDOT. *Decomposition of exact pfd persistence bimodules*, in "Discrete and Computational Geometry", December 2019, https://arxiv.org/abs/1605.09726 , https://hal.inria.fr/hal-01359312

[25] J. KIM, A. RINALDO, L. WASSERMAN. *Minimax Rates for Estimating the Dimension of a Manifold*, in "Journal of Computational Geometry", 2019, vol. 10, n^{o} 1, https://arxiv.org/abs/1605.01011 - 54 pages, 11 figures [*DOI :* 10.20382/JOCG.V10I1A3], https://hal.inria.fr/hal-02425684

[26] C. MARIA, J. PURCELL. *Treewidth, crushing and hyperbolic volume*, in "Algebraic and Geometric Topology", 2019, vol. 19, n^{o} 5, pp. 2625-2652, https://arxiv.org/abs/1805.02357 [*DOI :* 10.2140/AGT.2019.19.2625], https://hal.archives-ouvertes.fr/hal-02429712

[27] S. Y. OUDOT, E. SOLOMON. *Inverse Problems in Topological Persistence: a Survey*, in "Abel Symposia", 2019, https://arxiv.org/abs/1810.10813 , forthcoming, https://hal.inria.fr/hal-01966676

International Conferences with Proceedings

[28] B. BEAUFILS, F. CHAZAL, M. GRELET, B. MICHEL. *Robust pedestrian trajectory reconstruction from inertial sensor*, in "IPIN 2019 - 10th International Conference on Indoor Positioning and Indoor Navigation", Pisa, Italy, 2019, https://hal.inria.fr/hal-02271580

[29] J.-D. BOISSONNAT, O. DEVILLERS, K. DUTTA, M. GLISSE. *Randomized incremental construction of Delaunay triangulations of nice point sets*, in "ESA 2019 - 27th Annual European Symposium on Algorithms", Munich, Germany, September 2019 [*DOI :* 10.4230/LIPICS.ESA.2019.22], https://hal.inria.fr/hal-02185566

[30] J.-D. BOISSONNAT, S. PRITAM. *Computing Persistent Homology of Flag Complexes via Strong Collapses*, in "SoCG 2019 - International Symposium on Computational geometry", Portland, United States, April 2019 [*DOI :* 10.4230/LIPIcs.SoCG.2019.55], https://hal.inria.fr/hal-02078311

[31] M. KERBER, M. LESNICK, S. Y. OUDOT. *Exact computation of the matching distance on 2-parameter persistence modules*, in "International Symposium on Computational Geometry", Portland, Oregon, United States, 2019, https://arxiv.org/abs/1812.09085 , https://hal.inria.fr/hal-01966666

[32] J. KIM, J. SHIN, A. RINALDO, L. WASSERMAN. *Uniform Convergence of the Kernel Density Estimator Adaptive to Intrinsic Volume Dimension*, in "Thirty-sixth International Conference on Machine Learning (ICML 2019)", Long Beach, United States, Proceedings of Thirty-sixth International Conference on Machine Learning (ICML 2019), 2019, vol. 97, https://arxiv.org/abs/1810.05935 - 51 pages, https://hal.inria.fr/hal-02404295

Conferences without Proceedings

[33] H. ANAI, F. CHAZAL, M. GLISSE, Y. IKE, H. INAKOSHI, R. TINARRAGE, Y. UMEDA. *DTM-based Filtrations*, in "SoCG 2019 - 35th International Symposium on Computational Geometry", Portland, United States, June 2019, https://arxiv.org/abs/1811.04757 - This version is an extended abstract. The complete version of the paper, including proofs and additional comments, can be found at https://arxiv.org/abs/1811.04757 [*DOI :* 10.4230/LIPIcs.SoCG.2019.58], https://hal.archives-ouvertes.fr/hal-02093445

Scientific Books (or Scientific Book chapters)

[34] H. KANNAN, N. KOMODAKIS, N. PARAGIOS. *Tighter continuous relaxations for MAP inference in discrete MRFs: A survey*, in "Handbook of Numerical Analysis: Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 2, Volume 20, pages 351 - 400, Elsevier, https://www.sciencedirect.com/science/article/pii/S1570865919300092", Elsevier, October 2019 [*DOI :* 10.1016/BS.HNA.2019.06.001], https://hal.inria.fr/hal-02432873

[35] C. MARIA, H. SCHREIBER. *Discrete Morse Theory for Computing Zigzag Persistence*, in "Algorithms and Data Structures - 16th International Symposium, WADS 2019, Edmonton, AB, Canada, August 5-7, 2019, Proceedings", Z. FRIGGSTAD, J.-R. SACK, M. R. SALAVATIPOUR (editors), Springer, July 2019, pp. 538-552, https://arxiv.org/abs/1807.05172 [*DOI :* 10.1007/978-3-030-24766-9_39], https://hal.inria.fr/hal-01971682

Research Reports

[36] J.-D. BOISSONNAT, O. DEVILLERS, K. DUTTA, M. GLISSE. *Randomized incremental construction of Delaunay triangulations of nice point sets*, Inria, 2019, https://hal.inria.fr/hal-01950119

[37] O. DEVILLERS, P. DUCHON, M. GLISSE, X. GOAOC. *On Order Types of Random Point Sets*, Inria, 2019, https://arxiv.org/abs/1812.08525 , https://hal.inria.fr/hal-01962093

Other Publications

[38] S. ARYA, J.-D. BOISSONNAT, K. DUTTA. *Dimensionality Reduction for k-Distance Applied to Persistent Homology*, December 2019, working paper or preprint, https://hal.inria.fr/hal-01950051

[39] N. BERKOUK, G. GINOT, S. Y. OUDOT. *Level-sets persistence and sheaf theory*, December 2019, https://arxiv.org/abs/1907.09759 - working paper or preprint, https://hal.inria.fr/hal-02425597

[40] J.-D. BOISSONNAT, R. DYER, A. GHOSH, A. LIEUTIER, M. WINTRAECKEN. *Local conditions for triangulating submanifolds of Euclidean space*, July 2019, working paper or preprint, https://hal.inria.fr/hal-02267620

[41] J.-D. BOISSONNAT, S. KACHANOVICH, M. WINTRAECKEN. *Sampling and Meshing Submanifolds in High Dimension*, November 2019, working paper or preprint, https://hal.inria.fr/hal-02386169

[42] J.-D. BOISSONNAT, S. PRITAM. *Edge Collapse and Persistence of Flag Complexes*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02395227

[43] J.-D. BOISSONNAT, M. WINTRAECKEN. *The topological correctness of PL-approximations of isomanifolds*, November 2019, working paper or preprint, https://hal.inria.fr/hal-02386193

[44] M. CARRIERE, F. CHAZAL, Y. IKE, T. LACOMBE, M. ROYER, Y. UMEDA. *A General Neural Network Architecture for Persistence Diagrams and Graph Classification*, April 2019, working paper or preprint, https://hal.inria.fr/hal-02105788

[45] D. COHEN-STEINER, A. C. DE VITIS. *Spectral Properties of Radial Kernels and Clustering in High Dimensions*, January 2020, working paper or preprint, https://hal.inria.fr/hal-01969956

[46] D. COHEN-STEINER, A. LIEUTIER, J. VUILLAMY. *Regular triangulations as lexicographic optimal chains*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02391285

[47] M. DINDIN, Y. UMEDA, F. CHAZAL. *Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks*, June 2019, https://arxiv.org/abs/1906.05795 - 7 pages, 4 figures, https://hal.inria.fr/hal-02155849

[48] V. DIVOL, T. LACOMBE. *Understanding the Topology and the Geometry of the Persistence Diagram Space via Optimal Partial Transport*, April 2019, https://arxiv.org/abs/1901.03048 - working paper or preprint, https://hal.inria.fr/hal-01970466

[49] E. GASPAROVIC, E. MUNCH, S. Y. OUDOT, K. TURNER, B. WANG, Y. WANG. *Intrinsic Interleaving Distance for Merge Trees*, December 2019, https://arxiv.org/abs/1908.00063 - working paper or preprint, https://hal.inria.fr/hal-02425600

[50] J. KIM, J. SHIN, F. CHAZAL, A. RINALDO, L. WASSERMAN. *Homotopy Reconstruction via the Cech Complex and the Rips Complex*, December 2019, https://arxiv.org/abs/1903.06955 - 41 pages, 4 figures, https://hal.inria.fr/hal-02425686

[51] M. KRAMÁR, L. KOVALCINOVA, K. MISCHAIKOW, L. KONDIC. *Quantitative Measure of Memory Loss in Complex Spatio-Temporal Systems*, March 2019, working paper or preprint, https://hal.inria.fr/hal-02065027

[52] J. LEYGONIE, S. Y. OUDOT, U. TILLMANN. *A Framework for Differential Calculus on Persistence Barcodes*, October 2019, https://arxiv.org/abs/1910.00960 - 30 pages, 5 figures [*DOI :* 10.00960], https://hal.archives-ouvertes.fr/hal-02304300

[53] C. MARIA. *Parameterized complexity of quantum knot invariants*, January 2020, https://arxiv.org/abs/1910.00477 - working paper or preprint [*DOI :* 10.00477], https://hal.archives-ouvertes.fr/hal-02429767

[54] C. MARIA, S. Y. OUDOT, E. SOLOMON. *Intrinsic Topological Transforms via the Distance Kernel Embedding*, December 2019, https://arxiv.org/abs/1912.02225 - 50 pages, https://hal.inria.fr/hal-02425598

[55] Q. MÉRIGOT, A. DELALANDE, F. CHAZAL. *Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space*, October 2019, https://arxiv.org/abs/1910.05954 - 21 pages [*DOI :* 10.05954], https://hal.archives-ouvertes.fr/hal-02315998

[56] M. ROYER, F. CHAZAL, Y. IKE, Y. UMEDA. *ATOL: Automatic Topologically-Oriented Learning*, August 2019, https://arxiv.org/abs/1909.13472 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02296513

[57] R. TINARRAGE. *Recovering the homology of immersed manifolds*, December 2019, https://arxiv.org/abs/1912.03033 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02396261

[58] M. YEUNG, D. COHEN-STEINER, M. DESBRUN. *Material Coherence from Trajectories via Burau Eigenanalysis of Braids*, September 2019, working paper or preprint, https://hal.inria.fr/hal-02295987