Activity Report 2017

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
B9.4. - Sciences

1. Personnel

Research Scientists

Jean-Daniel Boissonnat [Inria, Senior Researcher, Sophia Antipolis - Méditerranée, HDR]
Frédéric Chazal [Team leader, Inria, Senior Researcher, Saclay - Île-de-France, HDR]
David Cohen-Steiner [Inria, Researcher, Sophia Antipolis - Méditerranée]
Kunal Dutta [Inria, Starting Research Position, from Oct 2017, Sophia Antipolis - Méditerranée]
Marc Glisse [Inria, Researcher, Saclay - Île-de-France]
Miroslav Kramar [Inria, Advanced Research Position, from Apr 2017, Saclay - Île-de-France]
Steve Oudot [Inria, Researcher, Saclay - Île-de-France, HDR]
Mathijs Wintraecken [Inria, Starting Research Position, from Nov 2017, Sophia Antipolis - Méditerranée]

Faculty Members

Guilherme Dias Da Fonseca [Univ d'Auvergne Clermont-Ferrand, Associate Professor, from Sep 2017, Sophia Antipolis - Méditerranée]
Maksims Ovsjanikovs [Ecole Polytechnique, Professor, Saclay - Île-de-France]

Post-Doctoral Fellows

Kunal Dutta [Inria, until Sep 2017, Sophia Antipolis - Méditerranée]
Mathijs Wintraecken [Inria, until Oct 2017, Sophia Antipolis - Méditerranée]

PhD Students

Eddie Aamari [Région, until Aug 2017, Saclay - Île-de-France]
Bertrand Beaufils [Sysnav, industrial grant, Saclay - Île-de-France]
Maxime Bellec [Sysnav, industrial grant, from Oct 2017, Saclay - Île-de-France]
Nicolas Berkouk [Inria, Saclay - Île-de-France]
Claire Brecheteau [Univ Paris-Sud, Saclay - Île-de-France]
Mathieu Carrière [Inria, Saclay - Île-de-France]
Jeremy Sylvain Cochoy [Univ Paris-Sud, Saclay - Île-de-France]

Alba Chiara de Vitis [Inria, until Jul 2017, Sophia Antipolis - Méditerranée]
Francois Godi [Univ de Nice - Sophia Antipolis, Sophia Antipolis - Méditerranée]
Siargey Kachanovich [Inria, Sophia Antipolis - Méditerranée]
Theo Lacombe [Ecole Polytechnique, from Sep 2017, Saclay - Île-de-France]
Siddharth Pritam [Inria, Sophia Antipolis - Méditerranée]
Raphael Tinarrage [Ecole Normale Supérieure Cachan, from Sep 2017, Saclay - Île-de-France]

Technical staff

Clément Jamin [Inria, until Nov 2017, Sophia Antipolis - Méditerranée]
Vincent Rouvreau [Inria, Engineer SED 40%, Saclay - Île-de-France]

Interns

Louis Cohen [Inria, from Mar 2017 until Jul 2017, Saclay - Île-de-France]
Vincent Divol [Ecole Normale Supérieure Cachan, from Mar 2017, Saclay - Île-de-France]
Raphael Tinarrage [Ecole Normale Supérieure Cachan, from Apr 2017 until Aug 2017, Saclay - Île-de-France]
Divyansh Pareek [IIT Bombay, from May 2017 until Jul 2017, Sophia Antipolis - Méditerranée]

Administrative Assistants

Christine Aklouche [Inria, until Aug 2017, Saclay - Île-de-France]
Florence Barbara [Inria, Assistant, Sophia Antipolis - Méditerranée]

Visiting Scientists

Magnus Bakke Botnan [TU München, Mar 2017, Saclay - Île-de-France]
Arijit Ghosh [Indian Statistical Institute, June 2017 and November 2017, Sophia Antipolis - Méditerranée]
Ilaria Giulini [Univ Pierre et Marie Curie, Saclay - Île-de-France]
Bruno Jartoux [ESIEE, Jul 2017, Saclay - Île-de-France]
Manish Krishan Lal [IIT Madras, from May 2017 until Jul 2017, Sophia Antipolis - Méditerranée]
Theo Lacombe [Ecole Polytechnique, from Apr 2017 until Aug 2017, Saclay - Île-de-France]
Bertrand Michel [Univ Pierre et Marie Curie, Saclay - Île-de-France]
Mariette Yvinec [Retraite, Sophia Antipolis - Méditerranée, HDR]

External Collaborators

Eddie Aamari [UC San Diego, from Sep 2017, Saclay - Île-de-France]
Pawel Dlotko [University of Swansea, from Mar 2017, Saclay - Île-de-France]
Ramsay Dyer [Mathematical Sciences Publishers, Sophia Antipolis - Méditerranée]
Clément Maria [University of Brisbane, until Nov 2017, Sophia Antipolis - Méditerranée]
Konstantin Mischaikow [Rutgers University, Nov 2017, 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 underly 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. Application Domains

4.1. Main application domains

Our work is mostly of a fundamental mathematical and algorithmic nature but fin ds applications in a variety of application in data analysis, more precisely in Topological Data Analysis (TDA). Although TDA is a quite recent field, it al ready founds applications in material science, biology, sensor networks, 3D shap es analysis and processing, to name a few.

More specifically, DATASHAPEhas recently started to work on the analysis of t rajectories obtained from inertial sensors (starting PhD thesis of Bertrand Bea ufils) and is exploring some possible new applications in material science.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Chairs

Jean-Daniel Boissonnat was elected a professor at the Collège de France, on the Chair Informatics and Computational Sciences for the academic year 2016-2017.

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

6. New Software and Platforms

6.1. 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.

RELEASE FUNCTIONAL DESCRIPTION: Major new features in 2017: - python interface - bottleneck distance - tangential complex - relaxed witness complex

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

6.2. dD Triangulations

CGAL module: Triangulations in any dimension

KEYWORDS: 3D modeling - Triangulation - Delaunay triangulation - Voronoi diagram - Regular triangulation

FUNCTIONAL DESCRIPTION: This package of CGAL (Computational Geometry Algorithms Library http://www.cgal.org) allows to compute triangulations, Delaunay triangulations and regular triangulations in any dimension. Those triangulations are built incrementally and can be modified by insertion or removal of vertices.

RELEASE FUNCTIONAL DESCRIPTION: Version 4.11 adds the regular triangulations to the package.

- Participants: Clément Jamin, Olivier Devillers and Samuel Hornus
- Contact: Samuel Hornus
- URL: http://www.cgal.org

7. New Results

7.1. Algorithmic aspects of topological and geometric data analysis

7.1.1. Variance Minimizing Transport Plans for Inter-surface Mapping

Participant: David Cohen-Steiner.

In collaboration with Manish Mandad, Leik Kobbelt (RWTH Aachen), Pierre Alliez (Inria), and Mathieu Desbrun (Caltech).

We introduce an effcient computational method for generating dense and low distortion maps between two arbitrary surfaces of same genus. Instead of relying on semantic correspondences or surface parameterization, we directly optimize a variance-minimizing transport plan between two input surfaces that defines an as-conformal-as-possible inter-surface map satisfying a user-prescribed bound on area distortion. The transport plan is computed via two alternating convex optimizations, and is shown to minimize a generalized Dirichlet energy of both the map and its inverse. Computational efficiency is achieved through a coarse-tone approach in diffusion geometry, with Sinkhorn iterations modified to enforce bounded area distortion. The resulting inter-surface mapping algorithm applies to arbitrary shapes robustly, with little to no user interaction.

7.1.2. Approximating the spectrum of a graph

Participant: David Cohen-Steiner.

In collaboration with Weihao Kong, Gregory Valiant (Stanford), and Christian Sohler (TU Dortmund).

The spectrum of a network or graph $G=(V,E)$ with adjacency matrix A consists of the eigenvalues of the normalized Laplacian $L=I-D^{-1/2}AD^{-1/2}$. This set of eigenvalues encapsulates many aspects of the structure of the graph, including the extent to which the graph posses community structures at multiple scales. We study the problem of approximating the spectrum $\lambda=(\lambda_1,...,\lambda_{|V|})$, $0\le\lambda 1,\le...,\le\lambda|V|\le 2$ of G in the regime where the graph is too large to explicitly calculate the spectrum. We present a sublinear time algorithm that, given the ability to query a random node in the graph and select a random neighbor of a given node, computes a succinct representation of an approximation $\widetilde{\lambda}$ such that $\|\widetilde{\lambda}-\lambda\|_1\le\varepsilon|V|$. Our algorithm has query complexity and running time $\exp(O(1/\varepsilon))$, independent of the size of the graph, $|V|$. We demonstrate the practical viability of our algorithm on 15 different real-world graphs from the Stanford Large Network Dataset Collection, including social networks, academic collaboration graphs, and road networks. For the smallest of these graphs, we are able to validate the accuracy of our algorithm by explicitly calculating the true spectrum; for the larger graphs, such a calculation is computationally prohibitive. In addition we study the implications of our algorithm to property testing in the bounded degree graph model.

7.1.3. Anisotropic triangulations via discrete Riemannian Voronoi diagrams

Participants: Jean-Daniel Boissonnat, Mathijs Wintraecken.

In collaboration with mael Rouxel-Labbé (GeometryFactory).

The construction of anisotropic triangulations is desirable for various applications, such as the numerical solving of partial differential equations and the representation of surfaces in graphics. To solve this notoriously difficult problem in a practical way, we introduce the discrete Riemannian Voronoi diagram, a discrete structure that approximates the Riemannian Voronoi diagram. This structure has been implemented and was shown to lead to good triangulations in $\mathbb{R}^2$ and on surfaces embedded in $\mathbb{R}^3$ as detailed in our experimental companion paper.

In [23], [32], [34], we study theoretical aspects of our structure. Given a finite set of points $\mathcal{P}$ in a domain Ω equipped with a Riemannian metric, we compare the discrete Riemannian Voronoi diagram of $\mathcal{P}$ to its Riemannian Voronoi diagram. Both diagrams have dual structures called the discrete Riemannian Delaunay and the Riemannian Delaunay complex. We provide conditions that guarantee that these dual structures are identical. It then follows from previous results that the discrete Riemannian Delaunay complex can be embedded in Ω under sufficient conditions, leading to an anisotropic triangulation with curved simplices. Furthermore, we show that, under similar conditions, the simplices of this triangulation can be straightened.

7.1.4. Only distances are required to reconstruct submanifolds

Participants: Jean-Daniel Boissonnat, Ramsay Dyer, Steve Oudot.

In collaboration with Arijit Ghosh (Indian Statistical Institute).

In [14], we give the first algorithm that outputs a faithful reconstruction of a submanifold of Euclidean space without maintaining or even constructing complicated data structures such as Voronoi diagrams or Delaunay complexes. Our algorithm uses the witness complex and relies on the stability of *power protection*, a notion introduced in this paper. The complexity of the algorithm depends exponentially on the intrinsic dimension of the manifold, rather than the dimension of ambient space, and linearly on the dimension of the ambient space. Another interesting feature of this work is that no explicit coordinates of the points in the point sample is needed. The algorithm only needs the *distance matrix* as input, i.e., only distance between points in the point sample as input.

7.1.5. An obstruction to Delaunay triangulations in Riemannian manifolds

Participants: Jean-Daniel Boissonnat, Ramsay Dyer.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Nikolay Martynchuk (University of Groningen).

Delaunay has shown that the Delaunay complex of a finite set of points P of Euclidean space $\mathbb{R}^m$ triangulates the convex hull of P, provided that P satisfies a mild genericity property. Voronoi diagrams and Delaunay complexes can be defined for arbitrary Riemannian manifolds. However, Delaunay's genericity assumption no longer guarantees that the Delaunay complex will yield a triangulation; stronger assumptions on P are required. A natural one is to assume that P is sufficiently dense. Although results in this direction have been claimed, we show that sample density alone is insufficient to ensure that the Delaunay complex triangulates a manifold of dimension greater than 2 [13].

7.1.6. Local criteria for triangulation of manifolds

Participants: Jean-Daniel Boissonnat, Ramsay Dyer, Mathijs Wintraecken.

In collaboration with Arijit Ghosh (Indian Statistical Institute).

We present criteria for establishing a triangulation of a manifold [40]. Given a manifold M, a simplicial complex $\mathcal{A}$, and a map H from the underlying space of $\mathcal{A}$ to M, our criteria are presented in local coordinate charts for M, and ensure that H is a homeomorphism. These criteria do not require a differentiable structure, or even an explicit metric on M. No Delaunay property of $\mathcal{A}$ is assumed. The result provides a triangulation guarantee for algorithms that construct a simplicial complex by working in local coordinate patches. Because the criteria are easily checked algorithmically, they are expected to be of general use.

7.1.7. Triangulating stratified manifolds I: a reach comparison theorem

Participants: Jean-Daniel Boissonnat, Mathijs Wintraecken.

In [42], we define the reach for submanifolds of Riemannian manifolds, in a way that is similar to the Euclidean case. Given a d-dimensional submanifold $\mathcal{S}$ of a smooth Riemannian manifold $\mathbb{M}$ and a point $p \in \mathbb{M}$ that is not too far from $\mathcal{S}$ we want to give bounds on local feature size of $\exp_p^{-1}(\mathcal{S})$. Here $\exp_p^{-1}$ is the inverse exponential map, a canonical map from the manifold to the tangent space. Bounds on the local feature size of $\exp_p^{-1}(\mathcal{S})$ can be reduced to giving bounds on the reach of $\exp_p^{-1}(\mathcal{B})$, where $\mathcal{B}$ is a geodesic ball, centred at c with radius equal to the reach of $\mathcal{S}$. Equivalently we can give bounds on the reach of $\exp_p^{-1} \circ \exp_c(B_c)$, where now B_c is a ball in the tangent space $T_c\mathbb{M}$, with the same radius. To establish bounds on the reach of $\exp_p^{-1} \circ \exp_c(B_c)$ we use bounds on the metric and on its derivative in Riemann normal coordinates.

This result is a first step towards answering the important question of how to triangulate stratified manifolds.

7.1.8. The reach, metric distortion, geodesic convexity and the variation of tangent spaces

Participants: Jean-Daniel Boissonnat, Mathijs Wintraecken.

In collaboration with André Lieutier (Dassault Système).

In [41], we discuss three results. The first two concern general sets of positive reach: We first characterize the reach by means of a bound on the metric distortion between the distance in the ambient Euclidean space and the set of positive reach. Secondly, we prove that the intersection of a ball with radius less than the reach with the set is geodesically convex, meaning that the shortest path between any two points in the intersection lies itself in the intersection. For our third result we focus on manifolds with positive reach and give a bound on the angle between tangent spaces at two different points in terms of the distance between the points and the reach.

7.1.9. Delaunay triangulation of a random sample of a good sample has linear size

Participants: Jean-Daniel Boissonnat, Kunal Dutta, Marc Glisse.

In collaboration with Olivier Devillers (Inria Nancy Grand Est).

The *randomized incremental construction* (RIC) for building geometric data structures has been analyzed extensively, from the point of view of worst-case distributions. In many practical situations however, we have to face nicer distributions. A natural question that arises is: do the usual RIC algorithms automatically adapt when the point samples are nicely distributed. We answer positively to this question for the case of the Delaunay triangulation of ϵ-nets.

ϵ-nets are a class of nice distributions in which the point set is such that any ball of radius ϵ contains at least one point of the net and two points of the net are distance at least ϵ apart. The Delaunay triangulations of ϵ-nets are proved to have linear size; unfortunately this is not enough to ensure a good time complexity of the randomized incremental construction of the Delaunay triangulation. In [33], [38], we prove that a uniform random sample of a given size that is taken from an ϵ-net has a linear sized Delaunay triangulation in any dimension. This result allows us to prove that the randomized incremental construction needs an expected linear size and an expected $O(n \log n)$ time.

Further, we also prove similar results in the case of non-Euclidean metrics, when the point distribution satisfies a certain *bounded expansion* property; such metrics can occur, for example, when the points are distributed on a low-dimensional manifold in a high-dimensional ambient space.

7.1.10. Kernelization of the Subset General Position problem in Geometry

Participants: Jean-Daniel Boissonnat, Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Sudeshna Kolay (Eindhoven University of Technology).

In [21], we consider variants of the GEOMETRIC SUBSET GENERAL POSITION problem. In defining this problem, a geometric subsystem is specified, like a subsystem of lines, hyperplanes or spheres. The input of the problem is a set of n points in $\mathbb{R}^d$ and a positive integer k. The objective is to find a subset of at least k input points such that this subset is in general position with respect to the specified subsystem. For example, a set of points is in general position with respect to a subsystem of hyperplanes in $\mathbb{R}^d$ if no $d+1$ points lie on the same hyperplane. In this paper, we study the HYPERPLANE SUBSET GENERAL POSITION problem under two parameterizations. When parameterized by k then we exhibit a polynomial kernelization for the problem. When parameterized by $h = n - k$, or the dual parameter, then we exhibit polynomial kernels which are also tight, under standard complexity theoretic assumptions. We can also conclude similar kernelization results for D-POLYNOMIAL SUBSET GENERAL POSITION, where a vector space of polynomials of degree at most d are specified as the underlying subsystem such that the size of the basis for this vector space is b. The objective is to find a set of at least k input points, or in the dual delete at most $h = n - k$ points, such that no $b + 1$ points lie on the same polynomial. Notice that this is a generalization of many well-studied geometric variants of the SET COVER problem, such as CIRCLE SUBSET GENERAL POSITION. We also study the general projective variants of these problems. These problems are also related to other geometric problems like SUBSET DELAUNAY TRIANGULATION problem.

7.1.11. Tight Kernels for Covering and Hitting: Point Hyperplane Cover and Polynomial Point Hitting Set

Participants: Jean-Daniel Boissonnat, Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Sudeshna Kolay (Eindhoven University of Technology).

The POINT HYPERPLANE COVER problem in $\mathbb{R}^d$ takes as input a set of n points in $\mathbb{R}^d$ and a positive integer k. The objective is to cover all the given points with a set of at most k hyperplanes. The D-POLYNOMIAL POINTS HITTING SET (D-POLYNOMIAL POINTS HS) problem in $\mathbb{R}^d$ takes as input a family $\mathcal{F}$ of D-degree polynomials from a vector space $\mathcal{R}$ in $\mathbb{R}^d$, and determines whether there is a set of at most k points in $\mathbb{R}^d$ that hit all the polynomials in $\mathcal{F}$. In [22], we exhibit tight kernels where k is the parameter for these problems.

7.1.12. Shallow packings, semialgebraic set systems, Macbeath regions, and polynomial partitioning

Participant: Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute) and Bruno Jartoux (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France) and Nabil H. Mustafa (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France).

The packing lemma of Haussler states that given a set system $(X, \mathbb{R})$ with bounded VC dimension, if every pair of sets in $\mathbb{R}$ have large symmetric difference, then $\mathbb{R}$ cannot contain too many sets. Recently it was generalized to the shallow packing lemma, applying to set systems as a function of their shallow-cell complexity. In [29] we present several new results and applications related to packings:

1. an optimal lower bound for shallow packings,
2. improved bounds on Mnets, providing a combinatorial analogue to Macbeath regions in convex geometry,
3. we observe that Mnets provide a general, more powerful framework from which the state-of-the-art unweighted ϵ-net results follow immediately, and
4. simplifying and generalizing one of the main technical tools in Fox *et al.* (*J. of the EMS*, to appear).

7.1.13. A Simple Proof of Optimal Epsilon Nets

Participant: Kunal Dutta.

In collaboration with Nabil H. Mustafa (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France, and Arijit Ghosh (Indian Statistical Institute)).

Showing the existence of ϵ-nets of small size has been the subject of investigation for almost 30 years, starting from the initial breakthrough of Haussler and Welzl (1987). Following a long line of successive improvements, recent results have settled the question of the size of the smallest ϵ-nets for set systems as a function of their so-called shallow-cell complexity.

In [20] we give a short proof of this theorem in the space of a few elementary paragraphs, showing that it follows by combining the ϵ-net bound of Haussler and Welzl (1987) with a variant of Haussler's packing lemma (1991).

This implies all known cases of results on unweighted ϵ-nets studied for the past 30 years, starting from the result of Matoušek, Seidel and Welzl (1990) to that of Clarkson and Varadajan (2007) to that of Varadarajan (2010) and Chan, Grant, Könemann and Sharpe (2012) for the unweighted case, as well as the technical and intricate paper of Aronov, Ezra and Sharir (2010).

7.1.14. On Subgraphs of Bounded Degeneracy in Hypergraphs

Participant: Kunal Dutta.

In collaboration with Arijit Ghosh (Indian Statistical Institute)).

A k-uniform hypergraph is d-degenerate if every induced subgraph has a vertex of degree at most d. In [48], given a k-uniform hypergraph $H = (V(H), E(H))$, we show there exists an induced subgraph of size at least

$$\sum_{v \in V(H)} \min \left\{ 1,\ c_k \left(\frac{d+1}{d_H(v)+1} \right)^{1/(k-1)} \right\},$$

where $c_k = 2^{-\left(1+\frac{1}{k-1}\right)} \left(1 - \frac{1}{k}\right)$ and $d_H(v)$ denotes the degree of vertex v in the hypergraph H. This connects, extends, and generalizes results of Alon-Kahn-Seymour (1987), on d-degenerate sets of graphs, Dutta-Mubayi-Subramanian (2012) on d-degenerate sets of linear hypergraphs, and Srinivasan-Shachnai (2004) on independent sets in hypergraphs to d-degenerate subgraphs of hypergraphs. Our technique also gives optimal lower bounds for a more generalized definition of degeneracy introduced by Zaker (2013). We further give a simple non-probabilistic proof of the Dutta-Mubayi-Subramanian bound for linear k-uniform hypergraphs, which extends the Alon-Kahn-Seymour proof technique to hypergraphs. Finally we provide several applications in discrete geometry, extending results of Payne-Wood (2013) and Cardinal-Tóth-Wood (2016). We also address some natural algorithmic questions. The proof of our main theorem combines the *random permutation* technique of Bopanna-Caro-Wei and Beame and Luby, together with a new *local density* argument which may be of independent interest.

7.2. Statistical aspects of topological and geometric data analysis

7.2.1. The DTM-signature for a geometric comparison of metric-measure spaces from samples

Participant: Claire Brécheteau.

In [43], 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 distance to a measure (DTM) introduced by Chazal, Cohen-Steiner and Mérigot. It leads to a pseudo-metric between metric-measure spaces, upper-bounded by the Gromov-Wasserstein distance. Under some geometric assumptions, we derive lower bounds for this pseudo-metric. Given two N-samples, we also build an asymptotic statistical test based on the DTM-signature, to reject the hypothesis of equality of the two underlying metric measure spaces, up to a measure-preserving isometry. We give strong theoretical justifications for this test and propose an algorithm for its implementation.

7.2.2. Estimating the Reach of a Manifold

Participants: Eddie Aamari, Frédéric Chazal, Bertrand Michel.

In collaboration with J. Kim, A. Rinaldo, L. Wasserman (Carnegie Mellon University)

Various problems of computational geometry and manifold learning encode geometric regularity through the so-called reach, a generalized convexity parameter. The reach τ_M of a submanifold $M \subset \mathbb{R}^D$ is the maximal offset radius on which the projection onto M is well defined. The quantity τ_M renders a certain minimal scale of M, giving bounds on both maximum curvature and possible bottleneck structures. In [35], 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 a framework where tangent spaces are known, and bounds assessing its efficiency are derived. In the case of i.i.d. random point cloud $\mathbb{X}_n$, $\widehat{\tau}(\mathbb{X}_n)$ is showed to achieve uniform expected loss bounds over a $\mathcal{C}^3$-like model. Minimax upper and lower bounds are derived, and we conclude with the extension to a model with unknown tangent spaces.

7.2.3. Robust Topological Inference: Distance To a Measure and Kernel Distance

Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with B. Fasy, F. Lecci, A. Rinaldo, L. Wasserman.

Let P be a distribution with support S. The salient features of S can be quantified with persistent homology, which summarizes topological features of the sublevel sets of the distance function (the distance of any point x to S). Given a sample from P we can infer the persistent homology using an empirical version of the distance function. However, the empirical distance function is highly non-robust to noise and outliers. Even one outlier is deadly. The distance-to-a-measure (DTM) and the kernel distance are smooth functions that provide useful topological information but are robust to noise and outliers. In [17], we derive limiting distributions and confidence sets, and we propose a method for choosing tuning parameters.

7.2.4. Statistical analysis and parameter selection for Mapper

Participants: Steve Oudot, Bertrand Michel, Mathieu Carrière.

In [44] we study the question of the statistical convergence of the 1-dimensional Mapper to its continuous analogue, the Reeb graph. We show that the Mapper is an optimal estimator of the Reeb graph, which gives, as a byproduct, a method to automatically tune its parameters and compute confidence regions on its topological features, such as its loops and flares. This allows to circumvent the issue of testing a large grid of parameters and keeping the most stable ones in the brute-force setting, which is widely used in visualization, clustering and feature selection with the Mapper.

7.2.5. Sliced Wasserstein Kernel for Persistence Diagrams

Participants: Steve Oudot, Mathieu Carrière.

In collaboration with M. Cuturi (ENSAE)

Persistence diagrams (PDs) play a key role in topological data analysis (TDA), in which they are routinely used to describe succinctly complex topological properties of complicated shapes. PDs enjoy strong stability properties and have proven their utility in various learning contexts. They do not, however, live in a space naturally endowed with a Hilbert structure and are usually compared with specific distances, such as the bottleneck distance. To incorporate PDs in a learning pipeline, several kernels have been proposed for PDs with a strong emphasis on the stability of the RKHS distance w.r.t. perturbations of the PDs. In [27], we use the Sliced Wasserstein approximation of the Wasserstein distance to define a new kernel for PDs, which is not only provably stable but also provably discriminative w.r.t. the Wasserstein distance W1∞ between PDs. We also demonstrate its practicality, by developing an approximation technique to reduce kernel computation time, and show that our proposal compares favorably to existing kernels for PDs on several benchmarks.

7.2.6. An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists

Participants: Frédéric Chazal, Bertrand Michel.

Topological Data Analysis (TDA) is a recent and fast growing field providing a set of new topological and geometric tools to infer relevant features for possibly complex data. In [45], we propose a brief introduction, through a few selected recent and state-of-the-art topics, to basic fundamental and practical aspects of TDA for non experts.

7.3. Topological approach for multimodal data processing

7.3.1. On the Stability of Functional Maps and Shape Difference Operators

Participants: Frédéric Chazal, Ruqi Huang, Maks Ovsjanikov.

In this paper, we provide stability guarantees for two frameworks that are based on the notion of functional maps. We consider two types of perturbations in our analysis: one is on the input shapes and the other is on the change in *scale*. In theory, we formulate and justify the robustness that has been observed in practical implementations of those frameworks. Inspired by our theoretical results, we propose a pipeline for constructing shape difference operators on point clouds and show numerically that the results are robust and informative. In particular, we show that both the shape difference operators and the derived areas of highest distortion are stable with respect to changes in shape representation and change of scale. Remarkably, this is in contrast with the well-known instability of the eigenfunctions of the Laplace-Beltrami operator computed on point clouds compared to those obtained on triangle meshes.

7.3.2. Local Equivalence and Intrinsic Metrics Between Reeb Graphs

Participants: Steve Oudot, Mathieu Carrière.

As graphical summaries for topological spaces and maps, Reeb graphs are common objects in the computer graphics or topological data analysis literature. Defining good metrics between these objects has become an important question for applications, where it matters to quantify the extent by which two given Reeb graphs differ. Recent contributions emphasize this aspect, proposing novel distances such as functional distortion or interleaving that are provably more discriminative than the so-called bottleneck distance, being true metrics whereas the latter is only a pseudo-metric. Their main drawback compared to the bottleneck distance is to be comparatively hard (if at all possible) to evaluate. In [28] we take the opposite view on the problem and show that the bottleneck distance is in fact good enough locally, in the sense that it is able to discriminate a Reeb graph from any other Reeb graph in a small enough neighborhood, as efficiently as the other metrics do. This suggests considering the intrinsic metrics induced by these distances, which turn out to be all globally equivalent. This novel viewpoint on the study of Reeb graphs has a potential impact on applications, where one may not only be interested in discriminating between data but also in interpolating between them.

7.3.3. Structure and Stability of the One-Dimensional Mapper

Participants: Steve Oudot, Mathieu Carrière.

Given a continuous function $f : X- > R$ and a cover I of its image by intervals, the Mapper is the nerve of a refinement of the pullback cover $f^{-1}(I)$. Despite its success in applications, little is known about the structure and stability of this construction from a theoretical point of view. As a pixelized version of the Reeb graph of f, it is expected to capture a subset of its features (branches, holes), depending on how the interval cover is positioned with respect to the critical values of the function. Its stability should also depend on this positioning. In [16] we propose a theoretical framework relating the structure of the Mapper to that of the Reeb graph, making it possible to predict which features will be present and which will be absent in the Mapper given the function and the cover, and for each feature, to quantify its degree of (in-)stability. Using this framework, we can derive guarantees on the structure of the Mapper, on its stability, and on its convergence to the Reeb graph as the granularity of the cover I goes to zero.

7.4. Experimental research and software development

7.4.1. Stride detection for pedestrian trajectory reconstruction: a machine learning approach based on geometric patterns

Participants: Frédéric Chazal, Bertrand Michel, Bertrand Beaufils.

In collaboration with M. Grelet (Sysnav)

A strides detection algorithm is proposed using inertial sensors worn on the ankle. This innovative approach based on geometric patterns can detect both normal walking strides and atypical strides such as small steps, side steps and backward walking that existing methods struggle to detect. It is also robust in critical situations, when for example the wearer is sitting and moving the ankle, while most algorithms in the literature would wrongly detect strides.

8. Bilateral Contracts and Grants with Industry

8.1. 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).
- Collaboration with Fujitsu on TDA and Machine learning (started in Dec 2017).

8.2. Bilateral Grants with Industry

- DATASHAPE and Sysnav have been selected for the ANR/DGA Challenge MALIN (funding: 700 kEuros) in September 2017.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

9.1.1.1. ANR TOPDATA

Participants: Jean-Daniel Boissonnat, Frédéric Chazal, David Cohen-Steiner, Mariette Yvinec, Steve Oudot, Marc Glisse.

- Acronym : TopData.

- Type : ANR blanc.

- Title : Topological Data Analysis: Statistical Methods and Inference.

- Coordinator : Frédéric Chazal (DATASHAPE).

- Duration : 4 years from October 2013 to September 2017.

- Others Partners: Département de Mathématiques (Université Paris Sud), Institut de Mathématiques (Université de Bourgogne), LPMA (Université Paris Diderot), LSTA (Université Pierre et Marie Curie).

- Abstract: TopData aims at designing new mathematical frameworks, models and algorithmic tools to infer and analyze the topological and geometric structure of data in different statistical settings. Its goal is to set up the mathematical and algorithmic foundations of Statistical Topological and Geometric Data Analysis and to provide robust and efficient tools to explore, infer and exploit the underlying geometric structure of various data.

Our conviction, at the root of this project, is that there is a real need to combine statistical and topological/geometric approaches in a common framework, in order to face the challenges raised by the inference and the study of topological and geometric properties of the wide variety of larger and larger available data. We are also convinced that these challenges need to be addressed both from the mathematical side and the algorithmic and application sides. Our project brings together in a unique way experts in Statistics, Geometric Inference and Computational Topology and Geometry. Our common objective is to design new theoretical frameworks and algorithmic tools and thus to contribute to the emergence of a new field at the crossroads of these domains. Beyond the purely scientific aspects we hope this project will help to give birth to an active interdisciplinary community. With these goals in mind we intend to promote, disseminate and make our tools available and useful for a broad audience, including people from other fields.

- See also: http://geometrica.saclay.inria.fr/collaborations/TopData/Home.html

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. GUDHI

Title: Algorithmic Foundations of Geometry Understanding in Higher Dimensions

Programm: FP7

Type: ERC

Duration: February 2014 - January 2019

Coordinator: Inria

Inria contact: Jean-Daniel Boissonnat.

'The central goal of this proposal is to settle the algorithmic foundations of geometry understanding in dimensions higher than 3. We coin the term geometry understanding to encompass a collection of tasks including the computer representation and the approximation of geometric structures, and the inference of geometric or topological properties of sampled shapes. The need to understand geometric structures is ubiquitous in science and has become an essential part of scientific computing and data analysis. Geometry understanding is by no means limited to three dimensions. Many applications in physics, biology, and engineering require a keen understanding of the geometry of a variety of higher dimensional spaces to capture concise information from the underlying often highly nonlinear structure of data. Our approach is complementary to manifold learning techniques and aims at developing an effective theory for geometric and topological data analysis. To reach these objectives, the guiding principle will be to foster a symbiotic relationship between theory and practice, and to address fundamental research issues along three parallel advancing fronts. We will simultaneously develop mathematical approaches providing theoretical guarantees, effective algorithms that are amenable to theoretical analysis and rigorous experimental validation, and perennial software development. We will undertake the development of a high-quality open source software platform to implement the most important geometric data structures and algorithms at the heart of geometry understanding in higher dimensions. The platform will be a unique vehicle towards researchers from other fields and will serve as a basis for groundbreaking advances in scientific computing and data analysis.'

9.3. International Initiatives

9.3.1. Inria Associate Teams Not Involved in an Inria International Labs

9.3.1.1. CATS

Title: Computations And Topological Statistics

International Partner (Institution - Laboratory - Researcher):

Carnegie Mellon University (United States) - Department of Statistics - Larry Wasserman

Start year: 2015

See also: http://geometrica.saclay.inria.fr/collaborations/CATS/CATS.html

Topological Data Analysis (TDA) is an emergent field attracting interest from various communities, that has recently known academic and industrial successes. Its aim is to identify and infer geometric and topological features of data to develop new methods and tools for data exploration and data analysis. TDA results mostly rely on deterministic assumptions which are not satisfactory from a statistical viewpoint and which lead to a heuristic use of TDA tools in practice. Bringing together the strong expertise of two groups in Statistics (L. Wasserman's group at CMU) and Computational Topology and Geometry (Inria Geometrica), the main objective of CATS is to set-up the mathematical foundations of Statistical TDA, to design new TDA methods and to develop efficient and easy-to-use software tools for TDA.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

Ramsay Dyer, Mathematical Sciences Publishers, Canada (June and November 2017)

Arijit Ghosh, Indian Statistical Institute, Kolkata (June and november 2017)

Kim Jisu, CMU, Pittsburgh, USA (November 2017).

Wolfgang Polonik, UC Davis, USA (June 2017).

Konstantin Mischaikow, Rutgers University, USA, (November 2017).

Magnus Botnan, TU Munich, Germany (March 2017).

Sara Kalisnik, MPI, Germany (November 2017).

9.4.1.1. Internships

Divyansh Pareek, IIT Bombay (May-July 2017)

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

Vincent Divol, UC Davis (April-June 2017)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Frédéric Chazal co-organized, with M. Meila 5Univ. of Washington) the NIPS 2017 workshop "Synergies in Geometric Data Analysis", December 2017.
- Frédéric Chazal co-organized the workshop "Functoriality in Geometric Data", Schloss Dagstuhl, Germany, January 2017

10.1.2. Journal

10.1.2.1. 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)*.

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

10.1.3. Invited Talks

Jean-Daniel Boissonnat has been invited and gave a talk at NYU-AD on November 19, 2017.

Frédéric Chazal, Foundations of Computational Mathematics (FoCM'17), Computational Topology and Geometry workshop, Barcelona, Spain, July 2017.

Frédéric Chazal, Applied Topology in Bedlewo 2017 Conference, Bedlewo, Poland, June 2017.

Frédéric Chazal, UC Davis Statistical Sciences Symposium, Davis, USA, May 2017.

Frédéric Chazal, Applied and Computational Algebraic Topology, Hausdorff Institute, Bonn,Germany, April 2017.

Frédéric Chazal, The First International Conference on the Mathematics of Data Science, Hong Kong Baptist University, Hong Kong, March 2017.

Frédéric Chazal, CNA/Ki-Net Workshop: Dynamics and Geometry from High Dimensional Data, Carnegie Mellon University, March 2017.

Frédéric Chazal, Colloquium, collaborative research center Discretization in Geometry and Dynamics, Munich, February 7, 2017.

Frédéric Chazal, Statistics/Learning at Paris-Saclay, workshop at IHES, January 19, 2017.

Mathijs Wintraecken gave an invited talk at the SoCG workshop on Algorithms for the Medial Axis in Brisbane, June 2017.

Steve Oudot, BIRS workshop on Topological Data Analysis: developping abstract foundations, Banff, Canada, August 2017.

Steve Oudot, Mini-Symposium on Computational Topology, Brisbane, Australia, July 2017.

Steve Oudot, Dagstuhl seminar on Topology, computation and data analysis, Dagstuhl, Germany, July 2017.

Steve Oudot, Applied Topology Seminar, Brown University, USA, April 2017.

Steve Oudot, TRIPODS wokshop on Geometry and topology for data, ICERM, USA, December 2017.

Steve Oudot, workshop on Mathematical signal processing and data analysis, Bremen University, Germany, September 2017.

Steve Oudot, Conférence de rentrée Maths-Info, ENS Cachan, September 2017.

10.1.4. Leadership within the Scientific Community

Steve Oudot is co-organizing the monthly seminar on combinatorial and computational geometry at Institut Henri Poincaré.

Steve Oudot is co-head of the GT Géométrie Algorithmique since September 2017.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Collège de France : Jean-Daniel Boissonnat, Géométrie algorithmique: données, modèles, programmes, avril-juin 2017.

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

Master: Jean-Daniel Boissonnat and Marc Glisse, Computational Geometry Learning, 36h eq-TD, M2, MPRI, France.

Master: Steve Oudot, Topological Data Analysis, 45h eq-TD, M1, École Polytechnique, France.

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

Master: Frédéric Chazal, Topological Data Analysis, 16h eq-TD, M2, Mathématiques, Vision, Apprentissage (MVA), ENS Paris-Saclay, France.

Winter School on Computational geometry and topology for data analysis, Jean-Daniel Boissonnat, Frédéric Chazal, Sophia-Antipolis, january 2017.

10.2.2. Supervision

PhD: Eddie Aamari, A Statistical Approach of Topological Data Analysis, September 1st, 2017, Frédéric Chazal (co-advised by Pascal Massart).

PhD in progress: Claire Brécheteau, Statistical aspects of distance-like functions , started September 1st, 2015, Frédéric Chazal (co-advised by Pascal Massart).

PhD in progress: Bertrand Beaufils, Méthodes topologiques et apprentissage statistique pour l'actimétrie du piéton à partir de données de mouvement, started November 2016, Frédéric Chazal (co-advised by Bertrand Michel).

PhD: Mathieu Carrière, Topological signatures for geometric data, defended November 21st, 2017, Steve Oudot.

PhD in progress: Jérémy Cochoy, Decomposition and stability of multidimensional persistence modules, started September 1st, 2015, Steve Oudot.

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, Steve Oudot.

PhD in progress: Alba Chiara de Vitis, Concentration of measure and clustering, Jean-Daniel Boissonnat and David Cohen-Steiner.

PhD in progress: Siargey Kachanovich, Manifold reconstruction in higher dimensions, Jean-Daniel Boissonnat.

PhD in progress: François Godi, Data structures and algorithms for topological data analysis and high dimensional geometry, Jean-Daniel Boissonnat.

PhD in progress: Siddharth Pritam, Approximation algorithms in Computational Topology, Jean-Daniel Boissonnat.

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: Vincent Divol, statistical aspects of TDA, started September 1st, 2017, Frédéric Chazal (co-advised by Pascal Massart).

10.2.3. Juries

Frédéric Chazal was a member of the PhD defense committee of Aruni Choudhary, MPI (reviewer) and Aurelien Vasseur (co-advisor).

Jean-Daniel Boissonnat was a member of the PhD defense committee of Eddie Amari.

Steve Oudot was a member (examiner) of the PhD defence committees of Jérémy Dubut and Nicolas Ninin (Cosynus team, Ecole polytechnique).

10.3. Popularization

10.3.1. Inria-Industry Meeting

Marc Glisse, Miro Kramar and Steve Oudot held a booth for half a day.

Marc Glisse played for a small video which is now on the InriaInnovation YouTube channel https://youtu.be/lKNjGk-Z6b4.

11. 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. Boissonnat, R. Dyer, A. Ghosh. *Stability of Delaunay-type structures for manifolds: [extended abstract]*, in "Symposium on Computational Geometry 2012, SoCG '12, Chapel Hill, NC, USA, June 17-20, 2012", 2012, pp. 229–238, http://doi.acm.org/10.1145/2261250.2261284

[3] J.-D. Boissonnat, A. Ghosh. *Manifold reconstruction using tangential Delaunay complexes*, in "Discrete and Computational Geometry", January 2014, vol. 51, n° 1, pp. 221-267 [*DOI :* 10.1007/s00454-013-9557-2], https://hal.inria.fr/hal-00932209

[4] J. BOISSONNAT, C. MARIA. *The Simplex Tree: An Efficient Data Structure for General Simplicial Complexes*, in "Algorithmica", 2014, vol. 70, n^{o} 3, pp. 406–427, http://dx.doi.org/10.1007/s00453-014-9887-3

[5] F. CHAZAL, D. COHEN-STEINER, M. GLISSE, L. J. GUIBAS, S. OUDOT. *Proximity of Persistence Modules and Their Diagrams*, in "Proc. 25th Annual Symposium on Computational Geometry", 2009, pp. 237–246 [*DOI :* 10.1145/1542362.1542407], http://hal.inria.fr/inria-00292566/

[6] 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

[7] 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

[8] F. DE GOES, D. COHEN-STEINER, P. ALLIEZ, M. DESBRUN. *An Optimal Transport Approach to Robust Reconstruction and Simplification of 2D Shapes*, in "Computer Graphics Forum", 2011, vol. 30, n^{o} 5, pp. 1593–1602, Special issue for EUROGRAPHICS Symposium on Geometry Processing, http://hal.inria.fr/hal-00758019

[9] 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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] E. AAMARI. *Convergence Rates for Geometric Inference*, Université Paris-Saclay, September 2017, https://hal.inria.fr/tel-01607782

[11] M. CARRIERE. *On metric and statistical properties of topological descriptors for geometric data*, Université Paris 11, November 2017, https://hal.archives-ouvertes.fr/tel-01659347

Articles in International Peer-Reviewed Journals

[12] 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

[13] J.-D. BOISSONNAT, R. DYER, A. GHOSH, M. NIKOLAY. *An Obstruction to Delaunay Triangulations in Riemannian Manifolds*, in "Discrete and Computational Geometry", 2017 [*DOI :* 10.1145/336154.336221], https://hal.inria.fr/hal-01583073

[14] 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

[15] D. BREMNER, O. DEVILLERS, M. GLISSE, S. LAZARD, G. LIOTTA, T. MCHEDLIDZE, G. MOROZ, S. WHITESIDES, S. WISMATH. *Monotone Simultaneous Paths Embeddings in* $\mathbb{R}^d$, in "Discrete Mathematics and Theoretical Computer Science", January 2018, vol. 20, n^{o} 1, pp. 1-11, https://hal.inria.fr/hal-01529154

[16] M. CARRIERE, S. Y. OUDOT. *Structure and Stability of the One-Dimensional Mapper*, in "Foundations of Computational Mathematics", October 2017, pp. 1–64 [*DOI :* 10.1007/S10208-017-9370-Z], https://hal.archives-ouvertes.fr/hal-01633101

[17] F. CHAZAL, B. FASY, F. LECCI, B. MICHEL, A. RINALDO, L. WASSERMAN. *Robust Topological Inference: Distance To a Measure and Kernel Distance*, in "Journal of Machine Learning Research", 2017, https://arxiv.org/abs/1412.7197v1 , forthcoming, https://hal.inria.fr/hal-01232217

[18] F. CHAZAL, R. HUANG, M. OVSJANIKOV. *On the Stability of Functional Maps and Shape Difference Operators*, in "Computer Graphics Forum", 2017, pp. 1 - 12 [*DOI :* 10.1111/CGF.13238], https://hal.inria.fr/hal-01668186

[19] 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

[20] N. MUSTAFA, K. DUTTA, A. GHOSH. *A Simple Proof of Optimal Epsilon Nets*, in "Combinatorica", 2017, https://hal.archives-ouvertes.fr/hal-01360452

International Conferences with Proceedings

[21] J.-D. BOISSONNAT, K. DUTTA, A. GHOSH, S. KOLAY. *Kernelization of the Subset General Position problem in Geometry*, in "MFCS 2017 - 42nd International Symposium on Mathematical Foundations of Computer Science", Alborg, Denmark, August 2017 [*DOI :* 10.4230/LIPICS.MFCS.2017.25], https://hal.inria.fr/hal-01583101

[22] J.-D. BOISSONNAT, K. DUTTA, A. GHOSH, S. KOLAY. *Tight Kernels for Covering and Hitting: Point Hyperplane Cover and Polynomial Point Hitting Set*, in "LATIN 2018 - 13th Latin American Theoretical INformatics Symposium", Buenos Aires, Argentina, April 2018, https://hal.inria.fr/hal-01669884

[23] J.-D. BOISSONNAT, M. ROUXEL-LABBÉ, M. WINTRAECKEN. *Anisotropic triangulations via discrete Riemannian Voronoi diagrams*, in "Symposium on Computational Geometry SoCG 2017", Brisbane, Australia, July 2017 [*DOI :* 10.4230/LIPICS.SOCG.2017.19], https://hal.inria.fr/hal-01507111

[24] J.-D. BOISSONNAT, K. C. SRIKANTA. *An Efficient Representation for Filtrations of Simplicial Complexes*, in "Symposium on Discrete Algorithms SODA 2017", Barcelona, France, January 2017, https://hal.inria.fr/hal-01416683

[25] F. CHAZAL, B. BEAUFILS, M. GRELET, B. MICHEL. *Stride detection for pedestrian trajectory reconstruction: a machine learning approach based on geometric patterns*, in "IPIN 2017 - 8th International Conference on Indoor Positioning and Indoor Navigation", Sapporo, Japan, IEEE, September 2017, pp. 1-6 [*DOI :* 10.1109/IPIN.2017.8115867], https://hal.inria.fr/hal-01664659

[26] F. CHAZAL, M. OVSJANIKOV, E. CORMAN, M. BRONSTEIN, E. RODOLA, M. BEN-CHEN, L. GUIBAS, A. BRONSTEIN. *Computing and Processing Correspondences with Functional Maps* , in "ACM SIGGRAPH 2017 Courses", Los Angeles, United States, July 2017, pp. 1-62, https://hal.inria.fr/hal-01664767

Conferences without Proceedings

[27] M. CARRIERE, M. CUTURI, S. Y. OUDOT. *Sliced Wasserstein Kernel for Persistence Diagrams*, in "ICML 2017 - Thirty-fourth International Conference on Machine Learning", Sydney, Australia, August 2017, pp. 1-10, https://hal.archives-ouvertes.fr/hal-01633105

[28] M. CARRIERE, S. Y. OUDOT. *Local Equivalence and Intrinsic Metrics Between Reeb Graphs*, in "SoCG 2017 - 33rd International Symposium on Computational Geometry", Brisbane, Australia, July 2017, pp. 1-15 [*DOI :* 10.4230/LIPICS.SOCG.2017.25], https://hal.archives-ouvertes.fr/hal-01633109

[29] K. DUTTA, A. GHOSH, B. JARTOUX, N. MUSTAFA. *Shallow packings, semialgebraic set systems, Macbeath regions and polynomial partitioning*, in "33rd International Symposium on Computational Geometry (SoCG 2017)", Brisbane, Australia, July 2017, https://hal.archives-ouvertes.fr/hal-01360443

Scientific Books (or Scientific Book chapters)

[30] J.-D. BOISSONNAT, F. CHAZAL, M. YVINEC. *Geometric and Topological Inference*, Cambridge University Press, October 2017, To appear in the red series Cambridge Texts in Applied Mathematics, forthcoming, https://hal.inria.fr/hal-01615863

[31] F. CHAZAL, D. COHEN-STEINER, A. LIEUTIER, Q. MÉRIGOT, B. THIBERT. *Inference of curvature using tubular neighborhoods*, in "Lecture Notes in Mathematics", L. NAJMAN, P. ROMON (editors), Modern Approaches to Discrete Curvature, Springer, 2017, vol. 2184, pp. 133-158 [*DOI :* 10.1007/978-3-319-58002-9_4], https://hal.archives-ouvertes.fr/hal-01425558

Research Reports

[32] J.-D. BOISSONNAT, M. ROUXEL-LABBÉ, M. WINTRAECKEN. *Anisotropic triangulations via discrete Riemannian Voronoi diagrams*, Inria Sophia Antipolis, 2017, n^{o} RR-9056, https://hal.inria.fr/hal-01507273

[33] O. DEVILLERS, M. GLISSE. *Delaunay triangulation of a random sample of a good sample has linear size*, Inria Saclay Ile de France ; Inria Nancy - Grand Est, July 2017, n^{o} RR-9082, 6 p. , https://hal.inria.fr/hal-01568030

[34] M. ROUXEL-LABBÉ, M. WINTRAECKEN, J.-D. BOISSONNAT. *Discretized Riemannian Delaunay Triangulations*, Inria Sophia Antipolis - Méditerranée, October 2017, n^{o} RR-9103, 51 p. , https://hal.inria.fr/hal-01612924

Other Publications

[35] E. AAMARI, J. KIM, F. CHAZAL, B. MICHEL, A. RINALDO, L. WASSERMAN. *Estimating the Reach of a Manifold*, May 2017, https://arxiv.org/abs/1705.04565 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01521955

[36] E. AAMARI, C. LEVRARD. *Non-Asymptotic Rates for Manifold, Tangent Space, and Curvature Estimation*, April 2017, https://arxiv.org/abs/1705.00989 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01516032

[37] E. AAMARI, C. LEVRARD. *Stability and Minimax Optimality of Tangential Delaunay Complexes for Manifold Reconstruction*, November 2017, https://arxiv.org/abs/1512.02857 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01245479

[38] J.-D. BOISSONNAT, O. DEVILLERS, K. DUTTA, M. GLISSE. *Delaunay triangulation of a random sample of a good sample has linear size* , December 2017, working paper or preprint, https://hal.inria.fr/hal-01673170

[39] J.-D. BOISSONNAT, K. DUTTA, A. GHOSH, S. KOLAY. *Tight Kernels for Covering with Points and Polynomials*, May 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01518562

[40] J.-D. BOISSONNAT, R. DYER, A. GHOSH, M. WINTRAECKEN. *Local criteria for triangulation of manifolds* *, December 2017, A full version of the paper is available at [5], http://ramsaydyer.com/tmp/loccrit.pdf [*DOI :* 10.4230/LIPICS.CVIT.2016.23], https://hal.inria.fr/hal-01661230

[41] J.-D. BOISSONNAT, A. LIEUTIER, M. WINTRAECKEN. *The reach, metric distortion, geodesic convexity and the variation of tangent spaces*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01661227

[42] J.-D. BOISSONNAT, M. WINTRAECKEN. *Triangulating stratified manifolds I: a reach comparison theorem*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01661233

[43] C. BRÉCHETEAU. *The DTM-signature for a geometric comparison of metric-measure spaces from samples*, January 2017, working paper or preprint, https://hal.inria.fr/hal-01426331

[44] M. CARRIERE, B. MICHEL, S. Y. OUDOT. *Statistical analysis and parameter selection for Mapper*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01633106

[45] F. CHAZAL, B. MICHEL. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists*, October 2017, https://arxiv.org/abs/1710.04019 - working paper or preprint, https://hal.inria.fr/hal-01614384

[46] A. CHOUDHARY, S. KACHANOVICH, M. WINTRAECKEN. *Coxeter triangulations have good quality*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01667404

[47] D. COHEN-STEINER, W. KONG, C. SOHLER, G. VALIANT. *Approximating the Spectrum of a Graph*, December 2017, https://arxiv.org/abs/1712.01725 - working paper or preprint, https://hal.inria.fr/hal-01661199

[48] K. DUTTA, A. GHOSH. *On Subgraphs of Bounded Degeneracy in Hypergraphs*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01669886