Activity Report 2016

Team DATASHAPE

Understanding the Shape of Data

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

Table of contents

Team DATASHAPE

Creation of the Team: 2016 January 01

Keywords:

Computer Science and Digital Science:
3.1.1. - Modeling, representation
3.1.4. - Uncertain data
3.3.3. - Big data analysis
3.4. - Machine learning and statistics
5.5.1. - Geometrical modeling
6.2.4. - Statistical methods
6.2.8. - Computational geometry and meshes
6.3.3. - Data processing
7.2. - Discrete mathematics, combinatorics
7.5. - Geometry, Topology
8.2. - Machine learning

Other Research Topics and Application Domains:
5. - Industry of the future
9.4.1. - Computer science
9.4.2. - Mathematics
9.4.5. - Data science

1. Members

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]
Marc Glisse [Inria, Researcher, Saclay - Île-de-France]
Alfredo Hubard [Inria, Starting Research position, until Aug 2016, granted by FP7 ERC GUDHI project, Sophia Antipolis - Méditerranée]
Steve Oudot [Inria, Researcher, Saclay - Île-de-France, HDR]

Faculty Members

Bertrand Michel [Univ. Paris VI, Associate Professor, Saclay - Île-de-France]
Maksim Ovsjanikov [Ecole Polytechnique, Professor, Saclay - Île-de-France]

Engineers

Clément Jamin [Inria, granted by FP7 ERC GUDHI project, Sophia Antipolis - Méditerranée]
Vincent Rouvreau [Inria, SED 40%, Saclay - Île-de-France]

PhD Students

Eddie Aamari [Région, Saclay - Île-de-France]
Bertrand Beaufils [Sysnav, from Apr 2016, industrial grant, Saclay - Île-de-France]
Nicolas Berkouk [Inria, from Nov 2016, Saclay - Île-de-France]
Thomas Bonis [Inria, Saclay - Île-de-France]
Claire Brécheteau [Univ. Paris XI, Saclay - Île-de-France]

Mathieu Carrière [Inria, granted by FP7 ERC GUDHI project, Saclay - Île-de-France]
Jérémy Sylvain Cochoy [Univ. Paris XI, Saclay - Île-de-France]
Alba Chiara de Vitis [Inria, granted by FP7 ERC GUDHI project, Sophia Antipolis - Méditerranée]
Francois Godi [Univ. Nice, Sophia Antipolis - Méditerranée]
Ruqi Huang [Inria, Saclay - Île-de-France]
Siargey Kachanovich [Inria, granted by FP7 ERC GUDHI project, Sophia Antipolis - Méditerranée]
Siddharth Pritam [Inria, from Apr 2016, granted by FP7 ERC GUDHI project, Sophia Antipolis - Méditerranée]
Mael Rouxel-Labbé [Geometry Factory, granted by CIFRE, Sophia Antipolis - Méditerranée]

Post-Doctoral Fellows
Paweł Dłotko [Inria, granted by FP7 ERC GUDHI project, Saclay - Île-de-France]
Kunal Dutta [Inria, granted by FP7 ERC GUDHI project, Sophia Antipolis - Méditerranée]
Mathijs Wintraecken [Inria, granted by FP7 ERC GUDHI project, Sophia Antipolis - Méditerranée]

Visiting Scientists
Sandip Banerjee [Indian Statistical Institute, Visiting PhD Student, from Mar 2016 until Aug 2016, Sophia Antipolis - Méditerranée]
Sameer Desai [Indian Statistical Institute, Visiting PhD Student, from Oct 2016 until Dec 2016, Sophia Antipolis - Méditerranée]
Arijit Ghosh [Indian Statistical Institute, April 2016 and Nov 2016, Sophia Antipolis - Méditerranée]
Ilaria Giulini [Univ. Paris VI, Saclay - Île-de-France]

Administrative Assistants
Christine Aklouche [Inria, Saclay - Île-de-France]
Florence Barbara [Inria, Sophia Antipolis - Méditerranée]

Others
Emmanuel Caron-Parte [Inria, intern from Université Pierre et Marie Curie, from Apr 2016 until Sep 2016, Saclay - Île-de-France]
Ramsay Dyer [Freelance Editor, External Collaborator, Sophia Antipolis - Méditerranée]
Uday Kusupati [Inria, intern from Indian Institute of Technology, from May 2016 until Jul 2016, Sophia Antipolis - Méditerranée]
Harry Lang [Fullbright fellow, Visiting PhD Student, Saclay - Île-de-France]
Clément Maria [Queensland University, External Collaborator, Sophia Antipolis - Méditerranée]
Laurent Rineau [Geometry Factory, External Collaborator, Sophia Antipolis - Méditerranée]
Zongjun Tan [Inria, intern from École Centrale Paris, from Feb 2016 until Aug 2016, Saclay - Île-de-France]
Jane Tournois [Geometry Factory, External Collaborator, Sophia Antipolis - Méditerranée]
Mariette Yvinec [Retired, Researcher, Sophia Antipolis - Méditerranée, HDR]

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 finds applications in a variety of application in data analysis, more precisely in Topological Data Analysis (TDA). Although TDA is a quite recent field, it already founds applications in material science, biology, sensor networks, 3D shapes analysis and processing, to name a few.

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

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

Jean-Daniel Boissonnat has been 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/

5.1.2. Books

Publication of a book [29], providing a self-contained presentation of the theory of persistence modules over the real line, the objects that are at the heart of the field of TDA.

6. New Software and Platforms

6.1. GUDHI

Geometric Understanding in Higher Dimensions

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 underlie 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.

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

7. New Results

7.1. Algorithmic aspects of topological and geometric data analysis

7.1.1. *An Efficient Representation for Filtrations of Simplicial Complexes*

Participant: Jean-Daniel Boissonnat.

In collaboration with Karthik C.S. (Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Israel)

A filtration over a simplicial complex K is an ordering of the simplices of K such that all prefixes in the ordering are subcomplexes of K. Filtrations are at the core of Persistent Homology, a major tool in Topological Data Analysis. In order to represent the filtration of a simplicial complex, the entire filtration can be appended to any data structure that explicitly stores all the simplices of the complex such as the Hasse diagram or the recently introduced Simplex Tree by Boissonnat and Maria [Algorithmica '14]. However, with the popularity of various computational methods that need to handle simplicial complexes, and with the rapidly increasing size of the complexes, the task of finding a compact data structure that can still support efficient queries is of great interest.

This direction has been recently pursued for the case of maintaining simplicial complexes. For instance, Boissonnat et al. [SoCG '15] considered storing the simplices that are maximal for the inclusion and Attali et al. [IJCGA '12] considered storing the simplices that block the expansion of the complex. Nevertheless, so far there has been no data structure that compactly stores the *filtration* of a simplicial complex, while also allowing the efficient implementation of basic operations on the complex.

In this work [22], we propose a new data structure called the Critical Simplex Diagram (CSD) which is a variant of our work on the Simplex Array List (SAL) introduced in [SoCG '15]. Our data structure allows to store in a compact way the filtration of a simplicial complex, and allows for the efficient implementation of a large range of basic operations. Moreover, we prove that our data structure is essentially optimal with respect to the requisite storage space. Next, we show that the CSD representation admits the following construction algorithms.

- A new *edge-deletion* algorithm for the fast construction of Flag complexes, which only depends on the number of critical simplices and the number of vertices.

- A new *matrix-parsing* algorithm to quickly construct the relaxed strong Delaunay complexes, depending only on the number of witnesses and the dimension of the complex.

7.1.2. *Discretized Riemannian Delaunay triangulations*

Participants: Mael Rouxel-Labbé, Mathijs Wintraecken, Jean-Daniel Boissonnat.

Anisotropic meshes are desirable for various applications, such as the numerical solving of partial differential equations and graphics. In [27], we introduce an algorithm to compute discrete approximations of Riemannian Voronoi diagrams on 2-manifolds. This is not straightforward because geodesics, shortest paths between points, and therefore distances cannot in general be computed exactly. Our implementation employs recent developments in the numerical computation of geodesic distances and is accelerated through the use of an underlying anisotropic graph structure. We give conditions that guarantee that our discrete Riemannian Voronoi diagram is combinatorially equivalent to the Riemannian Voronoi diagram and that its dual is an embedded triangulation, using both approximate geodesics and straight edges. Both the theoretical guarantees on the approximation of the Voronoi diagram and the implementation are new and provide a step towards the practical application of Riemannian Delaunay triangulations.

7.1.3. *Efficient and Robust Persistent Homology for Measures*

Participants: Frédéric Chazal, Steve Oudot.

In collaboration with M. Buchet (Tohoku University), D. Sheehy (Univ. Connecticut).

A new paradigm for point cloud data analysis has emerged recently, where point clouds are no longer treated as mere compact sets but rather as empirical measures. A notion of distance to such measures has been defined and shown to be stable with respect to perturbations of the measure. This distance can easily be computed pointwise in the case of a point cloud, but its sublevel-sets, which carry the geometric information about the measure, remain hard to compute or approximate. This makes it challenging to adapt many powerful techniques based on the Euclidean distance to a point cloud to the more general setting of the distance to a measure on a metric space. We propose an efficient and reliable scheme to approximate the topological structure of the family of sublevel-sets of the distance to a measure. We obtain an algorithm for approximating the persistent homology of the distance to an empirical measure that works in arbitrary metric spaces. Precise quality and complexity guarantees are given with a discussion on the behavior of our approach in practice [17].

7.1.4. *Shallow Packings in Geometry*

Participants: Kunal Dutta, Arijit Ghosh.

A merged paper with Ezra, Esther (School of Mathematics, Georgia Institute of Technology, Atlanta, U.S.A.)

We refine the bound on the packing number, originally shown by Haussler, for shallow geometric set systems. Specifically, let V be a finite set system defined over an n-point set X; we view V as a set of indicator vectors over the n-dimensional unit cube. A $delta$-separated set of V is a subcollection W, such that the Hamming distance between each pair $u, v \in W$ is greater than δ, where $\delta > 0$ is an integer parameter. The δ-packing number is then defined as the cardinality of the largest δ-separated subcollection of V. Haussler showed an asymptotically tight bound of $\Theta((n/delta)^d)$ on the δ-packing number if V has VC-dimension (or primal shatter dimension) d. We refine this bound for the scenario where, for any subset, $X' \subset X$ of size $m \leq n$ and for any parameter $1 \leq k \leq m$, the number of vectors of length at most k in the restriction of V to X' is only

$O(m^{d_1} k^{d-d_1})$, for a fixed integer $d > 0$ and a real parameter $1 \le d_1 \le d$ (this generalizes the standard notion of bounded primal shatter dimension when $d_1 = d$). In this case when V is "k-shallow" (all vector lengths are at most k), we show that its δ-packing number is $O(n^{d_1} k^{d-d_1}/\delta^d)$, matching Haussler's bound for the special cases where $d_1 = d$ or $k = n$. We present two proofs, the first is an extension of Haussler's approach, and the second extends the proof of Chazelle, originally presented as a simplification for Haussler's proof. [21]

- A new *tight upper bound* for shallow-packings in δ-separated set systems of bounded primal shatter dimension.

7.1.5. *On Subgraphs of Bounded Degeneracy in Hypergraphs*

Participants: Kunal Dutta, Arijit Ghosh.

A k-uniform hypergraph has degeneracy bounded by d if every induced subgraph has a vertex of degree at most d. Given a k-uniform hypergraph $H = (V(H), E(H))$, we show there exists an induced subgraph of size at least

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

where $c_k = 2^{-(1+\frac{1}{k-1})} \left(1 - \frac{1}{k}\right)$ and $d_H(v)$ denotes the degree of vertex v in the hypergraph H. This extends and generalizes a result of Alon-Kahn-Seymour (Graphs and Combinatorics, 1987) for graphs, as well as a result of Dutta-Mubayi-Subramanian (SIAM Journal on Discrete Mathematics, 2012) for linear hypergraphs, to general k-uniform hypergraphs. We also generalize the results of Srinivasan and Shachnai (SIAM Journal on Discrete Mathematics, 2004) from independent sets (0-degenerate subgraphs) to d-degenerate subgraphs. 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. Our proof combines the random permutation technique of Bopanna-Caro-Wei (see e.g. The Probabilistic Method, N. Alon and J. H. Spencer; Dutta-Mubayi-Subramanian) and also Beame-Luby (SODA, 1990) together with a new local density argument which may be of independent interest. We also provide some applications in discrete geometry, and address some natural algorithmic questions. [28]

- A new algorithmic *lower bound* for largest d-degenerate subgraphs in k-uniform hypergraphs.

7.1.6. *A Simple Proof of Optimal Epsilon Nets*

Participants: Kunal Dutta, Arijit Ghosh.

In collaboration with Nabil Mustafa (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France.)

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 this paper 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). [40]

- A new *unified proof* for all known bounds on unweighted ε-nets studied in the last 30 years.

7.1.7. *Combinatorics of Set Systems with Small Shallow Cell Complexity: Optimal Bounds via Packings*

Participants: Kunal Dutta, Arijit Ghosh.

In collaboration with Bruno Jartoux and Nabil Mustafa (Université Paris-Est Marne-la-Vallée, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France.)

The packing lemma of Haussler states that given a set system (X, R) with bounded VC dimension, if every pair of sets in R are 'far apart' (i.e., have large symmetric difference), then R cannot contain too many sets. This has turned out to be the technical foundation for many results in geometric discrepancy using the entropy method as well as recent work on set systems with bounded VC dimension. Recently it was generalized to the shallow packing lemma [Dutta-Ezra-Ghosh SoCG 2015, Mustafa DCG 2016], applying to set systems as a function of their shallow cell complexity. In this paper we present several new results and applications related to packings:

1. an optimal lower bound for shallow packings, thus settling the open question in Ezra (SODA 2014) and Dutta et al. (SoCG 2015),
2. improved bounds on Mnets, providing a combinatorial analogue to Macbeath regions in convex geometry (Annals of Mathematics, 1952),
3. simplifying and generalizing the main technical tool in Fox et al. (J. of the EMS, 2016).

Besides using the packing lemma and a combinatorial construction, our proofs combine tools from polynomial partitioning and the probabilistic method. [37]

- A new *optimal lower bound* for shallow packings.
- New *improved bounds* for M-nets - combinatorial analogs of Macbeath regions in convex geometry.

7.1.8. *A new asymmetric correlation inequality for Gaussian measure*
Participants: Kunal Dutta, Arijit Ghosh.

In collaboration with Nabil Mustafa (Université Paris-Est Marne-la-Vallée, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France.)

The Khatri-Šidák lemma says that for any Gaussian measure μ over $\mathbb{R}^n$, given a convex set K and a slab L, both symmetric about the origin, one has $\mu(K \cap L) \geq \mu(K)\mu(L)$. We state and prove a new asymmetric version of the Khatri-Šidák lemma when K is a symmetric convex body and L is a slab (not necessarily symmetric about the barycenter of K). Our result also extends that of Szarek and Werner (1999), in a special case.

- A new *asymmetric* inequality for gaussian measure. [38].

7.2. Statistical aspects of topological and geometric data analysis

7.2.1. *Stability and Minimax Optimality of Tangential Delaunay Complexes for Manifold Reconstruction*
Participant: Eddie Aamari.

In collaboration with C. Levrard (Univ. Paris Diderot).

we consider the problem of optimality in manifold reconstruction. A random sample $\mathbb{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^D$ composed of points lying on a d-dimensional submanifold M, with or without outliers drawn in the ambient space, is observed. Based on the tangential Delaunay complex, we construct an estimator $\widehat{M}$ that is ambient isotopic and Hausdorff-close to M with high probability. $\widehat{M}$ is built from existing algorithms. In a model without outliers, we show that this estimator is asymptotically minimax optimal for the Hausdorff distance over a class of submanifolds with reach condition. Therefore, even with no a priori information on the tangent spaces of M, our estimator based on tangential Delaunay complexes is optimal. This shows that the optimal rate of convergence can be achieved through existing algorithms. A similar result is also derived in a model with outliers. A geometric interpolation result is derived, showing that the tangential Delaunay complex is stable with respect to noise and perturbations of the tangent spaces. In the process, a denoising procedure and a tangent space estimator both based on local principal component analysis (PCA) are studied [32].

7.2.2. *Rates in the Central Limit Theorem and diffusion approximation via Stein's Method*
Participant: Thomas Bonis.

We present a way to apply Stein's method in order to bound the Wasserstein distance between a, possibly discrete, measure and another measure assumed to be the invariant measure of a diffusion operator. We apply this construction to obtain convergence rates, in terms of p-Wasserstein distance for $p \geq 2$, in the Central Limit Theorem in dimension 1 under precise moment conditions. We also establish a similar result for the Wasserstein distance of order 2 in the multidimensional setting. In a second time, we study the convergence of stationary distributions of Markov chains in the context of diffusion approximation, with applications to density estimation from geometric random graphs and to sampling using the Langevin Monte Carlo algorithm [33].

7.2.3. *Rates of Convergence for Robust Geometric Inference*
Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with P. Massart (Univ. Paris Sud et Inria Select team).

Distances to compact sets are widely used in the field of Topological Data Analysis for inferring geometric and topological features from point clouds. In this context, the distance to a probability measure (DTM) has been introduced by Chazal et al. as a robust alternative to the distance a compact set. In practice, the DTM can be estimated by its empirical counterpart, that is the distance to the empirical measure (DTEM). In this paper we give a tight control of the deviation of the DTEM. Our analysis relies on a local analysis of empirical processes. In particular, we show that the rate of convergence of the DTEM directly depends on the regularity at zero of a particular quantile function which contains some local information about the geometry of the support. This quantile function is the relevant quantity to describe precisely how difficult is a geometric inference problem. Several numerical experiments illustrate the convergence of the DTEM and also confirm that our bounds are tight [19].

7.2.4. *Data driven estimation of Laplace-Beltrami operator*
Participants: Frédéric Chazal, Bertrand Michel, Ilaria Giulini.

Approximations of Laplace-Beltrami operators on manifolds through graph Laplacians have become popular tools in data analysis and machine learning. These discretized operators usually depend on bandwidth parameters whose tuning remains a theoretical and practical problem. In this paper, we address this problem for the unnormalized graph Laplacian by establishing an oracle inequality that opens the door to a well-founded data-driven procedure for the bandwidth selection. Our approach relies on recent results by Lacour and Massart on the so-called Lepski's method [26].

7.3. Topological approach for multimodal data processing

7.3.1. *Persistence-based Pooling for Shape Pose Recognition*
Participants: Thomas Bonis, Frédéric Chazal, Steve Oudot, Maksim Ovsjanikov.

We propose a novel pooling approach for shape classification and recognition using the bag-of-words pipeline, based on topological persistence, a recent tool from Topological Data Analysis. Our technique extends the standard max-pooling, which summarizes the distribution of a visual feature with a single number, thereby losing any notion of spatiality. Instead, we propose to use topological persistence, and the derived persistence diagrams, to provide significantly more informative and spatially sensitive characterizations of the feature functions, which can lead to better recognition performance. Unfortunately, despite their conceptual appeal, persistence diagrams are difficult to handle, since they are not naturally represented as vectors in Euclidean space and even the standard metric, the bottleneck distance is not easy to compute. Furthermore, classical distances between diagrams, such as the bottleneck and Wasserstein distances, do not allow to build positive definite kernels that can be used for learning. To handle this issue, we provide a novel way to transform persistence diagrams into vectors, in which comparisons are trivial. Finally, we demonstrate the performance of our construction on the Non-Rigid 3D Human Models SHREC 2014 dataset, where we show that topological pooling can provide significant improvements over the standard pooling methods for the shape pose recognition within the bag-of-words pipeline [23].

7.3.2. *Structure and Stability of the 1-Dimensional Mapper*
Participants: Steve Oudot, Mathieu Carrière.

Given a continuous function $f : X \to \mathbb{R}$ and a cover $\mathcal{I}$ of its image by intervals, the Mapper is the nerve of a refinement of the pullback cover $f^{-1}(\mathcal{I})$. 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. We propose a theoretical framework that relates the structure of the Mapper to the one 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 $\mathcal{I}$ goes to zero [25].

7.3.3. *Decomposition of exact pfd persistence bimodules*
Participants: Steve Oudot, Jérémy Cochoy.

We characterize the class of persistence modules indexed over $\mathbb{R}^2$ that are decomposable into summands whose support have the shape of a *block*—i.e. a horizontal band, a vertical band, an upper-right quadrant, or a lower-left quadrant. Assuming the modules are *pointwise finite-dimensional* (pfd), we show that they are decomposable into block summands if and only if they satisfy a certain local property called *exactness*. Our proof follows the same scheme as the proof of decomposition 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 stability theory for zigzags and interlevel-sets persistence modules, in which block-decomposable bimodules play a key part. Our results allow us to drop some of the conditions under which that theory holds, in particular the Morse-type conditions [39].

7.4. Experimental research and software development

7.4.1. *Topological Microstructure Analysis Using Persistence Landscapes*
Participant: Paweł Dłotko.

In collaboration with T. Wanner (George Mason University).

Phase separation mechanisms can produce a variety of complicated and intricate microstructures, which often can be difficult to characterize in a quantitative way. In recent years, a number of novel topological metrics for microstructures have been proposed, which measure essential connectivity information and are based on techniques from algebraic topology. Such metrics are inherently computable using computational homology, provided the microstructures are discretized using a thresholding process. However, while in many cases the thresholding is straightforward, noise and measurement errors can lead to misleading metric values. In such situations, persistence landscapes have been proposed as a natural topology metric. Common to all of these approaches is the enormous data reduction, which passes from complicated patterns to discrete information. It is therefore natural to wonder what type of information is actually retained by the topology. In the present paper, we demonstrate that averaged persistence landscapes can be used to recover central system information in the Cahn-Hilliard theory of phase separation. More precisely, we show that topological information of evolving microstructures alone suffices to accurately detect both concentration information and the actual decomposition stage of a data snapshot. Considering that persistent homology only measures discrete connectivity information, regardless of the size of the topological features, these results indicate that the system parameters in a phase separation process affect the topology considerably more than anticipated. We believe that the methods discussed in this paper could provide a valuable tool for relating experimental data to model simulations [36].

7.4.2. *Topological analysis of the connectome of digital reconstructions of neural microcircuits*
Participant: Paweł Dłotko.

In collaboration with K. Hess, L. Ran, H. Markram, E. Muller, M. Nolte, M. Reimann, M. Scolamiero, K. Turner (Univ. of Aberden, EPFL, Brain and Mind Institute).

A first draft digital reconstruction and simulation of a microcircuit of neurons in the neocortex of a two-week-old rat was recently published. Since graph-theoretical methods may not be sufficient to understand the immense complexity of the network formed by the neurons and their connections, we explored whether application of methods from algebraic topology can provide a novel and useful perspective on the structural and functional organization of the microcircuit. Structural topological analysis revealed that directed graphs representing the connectivity between neurons are significantly different from random graphs and that there exist an enormous number of simplicial complexes of different dimensions representing all-to-all connections within different sets of neurons, the most extreme motif of neuronal clustering reported so far in the brain. Functional topological analysis based on data from simulations confirmed the interest of a new approach to studying the relationship between the structure of the connectome and its emergent functions. In particular, functional responses to different stimuli can readily be distinguished by topological methods. This study represents the first algebraic topological analysis of connectomics data from neural microcircuits and shows promise for general applications in network science.

7.4.3. *A persistence landscapes toolbox for topological statistics*
Participant: Paweł Dłotko.

In collaboration with P. Bubenik (University of Florida).

Topological data analysis provides a multiscale description of the geometry and topology of quantitative data. The persistence landscape is a topological summary that can be easily combined with tools from statistics and machine learning. We give efficient algorithms for calculating persistence landscapes, their averages, and distances between such averages. We discuss an implementation of these algorithms and some related procedures. These are intended to facilitate the combination of statistics and machine learning with topological data analysis. We present an experiment showing that the low-dimensional persistence landscapes of points sampled from spheres (and boxes) of varying dimensions differ.

7.5. Miscellaneous

7.5.1. *Monotone Simultaneous Paths Embeddings in $\mathbb{R}^d$*
Participant: Marc Glisse.

In collaboration with O. Devillers and S. Lazard (Inria Nancy), David Bremner (University of New Brunswick, Canada), Giuseppe Liotta (University of Perugia, Italy), Tamara Mchedlidze (KIT, Germany), Sue Whitesides (University of Victoria, Canada), Stephen Wismath (University of Lethbridge, Canada).

We study[24] the following problem: Given k paths that share the same vertex set, is there a simultaneous geometric embedding of these paths such that each individual drawing is monotone in some direction? We prove that for any dimension $d \geq 2$, there is a set of $d + 1$ paths that does not admit a monotone simultaneous geometric embedding.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Research contract with GeometryFactory in the context of Mael Rouxel-Labbé's Ph.D. thesis on anisotropic mesh generation [12].

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 starting October 2013.

- 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. ERC GUDHI

Title: Algorithmic Foundations of Geometry Understanding in Higher Dimensions.

Program: FP7.

Type: ERC.

Duration: February 2014 - January 2019.

Coordinator: Inria.

PI: 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 (April and November 2016)

Arijit Ghosh, Indian Statistical Institute, Kolkata (April and November 2016)

Jose Carlos Gomez Larranaga, CIMAT, Guanajuato, Mexico (September 2016)

Kim Jisu, CMU, Pittsburgh, USA (May and December 2016).

Antony Bak, Palantir company, USA (October 2016)

9.4.1.1. Internships

Uday Kusupati, Indian Institute of Technology, Bombay (May-July 2016)

Sandip Banerjee (bourse Charpak), Indian Statistical Institute, Kolkata (March-August 2016)

Sameer Desai, Indian Statistical Institute, Kolkata (October-December 2016)

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

Steve Oudot and Jérémy Cochoy spent 3 months (Sept.-Nov.) at the Institute for Computational and Experimental Research in Mathematics (ICERM) at Brown University. They were invited there for the semester program entitled *Topology in Motion* (see https://icerm.brown.edu/programs/sp-f16/).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Jean-Daniel Boissonnat and Frédéric Chazal co-organized the joint GUDHI-TOPDATA workshop in Porquerolles, October 17-20.

Frédéric Chazal co-organized the SMAI-SIGMA Conference 2016 at Luminy (CIRM) in November.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Maks Ovsjanikov: Paper co-chair of the Symposium on Geometry Processing 2016 (SGP 2016).

10.1.2.2. Member of the Conference Program Committees

Frédéric Chazal: Symposium on Geometry Processing 2016 (SGP 2016).

Steve Oudot: Symposium on Geometry Processing 2016 (SGP 2016).

Marc Glisse: Symposium on Computational Geometry (SoCG 2016).

10.1.3. Journal

10.1.3.1. 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)*.

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

10.1.4. Invited Talks

Frédéric Chazal, ACCAPT conference, Aalborg, Danmark, April 2016.

Frédéric Chazal, Joint Mathematical Meetings, Seattle, USA, January 2016.

Frédéric Chazal, Séminaire Parisien de Géométrie Algorithmique, Paris, October 2016.

Frédéric Chazal, 9th International Conference of theERCIM WG on Computational and Methodological Statistics, December 2016.

Steve Oudot, ACCAPT conference, Aalborg, Danmark, April 2016.

Steve Oudot, Workshop SIGMA 2016, CIRM, France, November 2016.

Steve Oudot, Applied Topology Seminar, Brown University, USA, November 2016.

Steve Oudot, Topology and Neuroscience Seminar, Princeton University, USA, November 2016.

10.1.5. Scientific Expertise

Frédéric Chazal was a member of the ANR committee, CES 40 (Mathematics and Computer Science).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

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.

Doctorat : Frédéric Chazal and Bertrand Michel, An introduction to Topological Data Analysis, 18h eq-TD, Universitad Autonoma de Barcelona, Spain.

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

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

Master : Jean-Daniel Boissonnat, Winter School on Computational geometry and Topology, Inria Sophia Antipolis Méditerrannée, January 2016.

Doctorat : Steve Oudot, École Mathématique en Afrique on *Topologie différentielle, géométrie algébrique et applications*, La Marsa, Tunisia, March-April 2016.

Doctorat : Steve Oudot, Summer School on Mathematical Methods for High-Dimensional Data Analysis, Technical University of Munich, Germany, July 2016.

10.2.2. Supervision

PhD: Thomas Bonis, Statistical Learning Algorithms for Geometric and Topological Data Analysis, December 1st, 2016, Frédéric Chazal.

PhD : Mael Rouxel-Labbé, Génération de maillages anisotropes, december 16, 2016, Jean-Daniel Boissonnat.

PhD: Ruqi Huang, Algorithms for topological inference in metric spaces, December 14, 2016, Frédéric Chazal.

PhD in progress: Eddie Aamari, A Statistical Approach of Topological Data Analysis, started September 1st, 2014, 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 in progress: Mathieu Carrière, Topological signatures for geometric data, started November 1st, 2014, 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: Alba Chiara de Vitis, Concentration of measure and clustering.

PhD in progress: Siargey Kachanovich, Approximate algorithms in higher dimensional geometry.

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

10.2.3. Juries

Frédéric Chazal was a member (and reviewer) of the PhD defense committee of Mariia Fodetenkova (Inria Nancy).

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° 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° 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° 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° 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° 6, 38 p. , http://hal.archives-ouvertes.fr/hal-00923563

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] T. BONIS. *Statistical learning algorithms for topological and geometric data analysis*, Université Paris-Saclay, December 2016, https://hal.archives-ouvertes.fr/tel-01402801

[11] R. HUANG. *Two contributions to geometric data analysis: filamentary structures approximations, and stability properties of functional approaches for shape comparison*, University Paris-Saclay, December 2016, https://hal.archives-ouvertes.fr/tel-01419321

[12] M. ROUXEL-LABBÉ. *Anisotropic Mesh Generation*, Université Côte D'Azur, December 2016, https://hal.archives-ouvertes.fr/tel-01419457

Articles in International Peer-Reviewed Journals

[13] D. ATTALI, O. DEVILLERS, M. GLISSE, S. LAZARD. *Recognizing Shrinkable Complexes Is NP-Complete*, in "Journal of Computational Geometry", 2016, vol. 7, n° 1, pp. 430–443 [*DOI : 10.20382/JOCG.V7I1A18*], https://hal.inria.fr/hal-01384396

[14] J.-D. BOISSONNAT, D. MAZAURIC. *On the complexity of the representation of simplicial complexes by trees*, in "Theoretical Computer Science", February 2016, vol. 617, 17 p. [*DOI : 10.1016/J.TCS.2015.12.034*], https://hal.inria.fr/hal-01259806

[15] 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

[16] P. BUBENIK, D. PAWEL. *A persistence landscapes toolbox for topological statistics*, in "Journal of Symbolic Computation", 2016, https://hal.inria.fr/hal-01258875

[17] M. BUCHET, F. CHAZAL, S. OUDOT, D. SHEEHY. *Efficient and Robust Persistent Homology for Measures*, in "Computational Geometry", 2016, vol. 58 [*DOI : 10.1016/J.COMGEO.2016.07.001*], https://hal.inria.fr/hal-01342385

[18] 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, forthcoming, https://hal.inria.fr/hal-01232217

[19] F. CHAZAL, P. MASSART, B. MICHEL. *Rates of convergence for robust geometric inference*, in "Electronic journal of statistics ", 2016, vol. 10, n° 2, 44 p. , https://hal.inria.fr/hal-01336913

[20] O. DEVILLERS, M. GLISSE, X. GOAOC, R. THOMASSE. *Smoothed complexity of convex hulls by witnesses and collectors*, in "Journal of Computational Geometry", 2016, vol. 7, n° 2, pp. 101-144 [*DOI : 10.20382/JOCG.V7I2A6*], https://hal.inria.fr/hal-01285120

[21] K. DUTTA, E. EZRA, A. GHOSH. *Two proofs for Shallow Packings* , in "Discrete and Computational Geometry", December 2016 [*DOI : 10.1007/S00454-016-9824-0*], https://hal.archives-ouvertes.fr/hal-01360460

International Conferences with Proceedings

[22] 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

[23] T. BONIS, M. OVSJANIKOV, S. OUDOT, F. CHAZAL. *Persistence-based Pooling for Shape Pose Recognition*, in "6th International Workshop on Computational Topology in Image Context (CTIC 2016)", Marseille, France, CTIC 2016 Proceedings of the 6th International Workshop on Computational Topology in Image Context, June 2016, vol. 9667, https://hal.archives-ouvertes.fr/hal-01301937

[24] D. BREMNER, O. DEVILLERS, M. GLISSE, S. LAZARD, G. LIOTTA, T. MCHEDLIDZE, S. WHITESIDES, S. WISMATH. *Monotone Simultaneous Paths Embeddings in $\mathbb{R}^d$*, in "24th International Symposium on Graph Drawing & Network Visualization", Athens, Greece, Proceedings of 24th International Symposium on Graph Drawing & Network Visualization, Springer, September 2016, vol. 9801, https://hal.inria.fr/hal-01366148

[25] M. CARRIÈRE, S. OUDOT. *Structure and Stability of the 1-Dimensional Mapper*, in "Proceedings of the International Symposium on Computational Geometry", Boston, United States, June 2016, Conference version, https://hal.inria.fr/hal-01247511

[26] F. CHAZAL, I. GIULINI, B. MICHEL. *Data driven estimation of Laplace-Beltrami operator*, in "30th Conference on Neural Information Processing Systems (NIPS 2016)", Barcelona, Spain, December 2016, https://hal.inria.fr/hal-01387021

[27] M. ROUXEL-LABBÉ, M. WINTRAECKEN, J.-D. BOISSONNAT. *Discretized Riemannian Delaunay triangulations*, in "Proceedings 25th International Meshing Roundtable (IMR25)", Washington DC, United States, Elsevier, September 2016, https://hal.inria.fr/hal-01367525

Conferences without Proceedings

[28] K. DUTTA, A. GHOSH. *On Subgraphs of Bounded Degeneracy in Hypergraphs*, in "42nd International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2016)", Istanbul, Turkey, June 2016, https://hal.archives-ouvertes.fr/hal-01360425

Scientific Books (or Scientific Book chapters)

[29] 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

[30] F. CHAZAL. *High-Dimensional Topological Data Analysis*, in "3rd Handbook of Discrete and Computational Geometry", CRC Press, 2016, forthcoming, https://hal.inria.fr/hal-01316989

[31] F. CHAZAL, D. COHEN-STEINER, A. LIEUTIER, Q. MERIGOT, B. THIBERT. *Inference of curvature using tubular neighborhoods*, in "Lecture Notes in Mathematics", 2017, forthcoming, https://hal.archives-ouvertes.fr/hal-01425558

Other Publications

[32] E. AAMARI, C. LEVRARD. *Stability and Minimax Optimality of Tangential Delaunay Complexes for Manifold Reconstruction*, June 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01245479

[33] T. BONIS. *Rates in the Central Limit Theorem and diffusion approximation via Stein's Method*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01167372

[34] 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

[35] P. DLOTKO, K. HESS, L. RAN, H. MARKRAM, E. MULLER, M. NOLTE, M. REIMANN, M. SCOLAMIERO, K. TURNER. *Topological analysis of the connectome of digital reconstructions of neural microcircuits*, January 2016, working paper or preprint, https://hal.inria.fr/hal-01261011

[36] P. DLOTKO, T. WANNER. *Topological Microstructure Analysis Using Persistence Landscapes*, January 2016, working paper or preprint, https://hal.inria.fr/hal-01259127

[37] K. DUTTA, A. GHOSH, B. JARTOUX, N. MUSTAFA. *Combinatorics of Set Systems with Small Shallow Cell Complexity: Optimal Bounds via Packings*, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01360443

[38] K. DUTTA, A. GHOSH, N. MUSTAFA. *A new asymmetric correlation inequality for Gaussian measure*, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01360457

[39] C. JÉRÉMY, S. Y. OUDOT. *Decomposition of exact pfd persistence bimodules*, May 2016, working paper or preprint, https://hal.inria.fr/hal-01359312

[40] N. MUSTAFA, K. DUTTA, A. GHOSH. *A Simple Proof of Optimal Epsilon Nets*, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01360452