

RESEARCH CENTRES

Saclay - Île-de-France

Sophia Antipolis - Méditerranée

IN PARTNERSHIP WITH:

CNRS, Université Paris-Saclay

2020

ACTIVITY REPORT

Project-Team

DATASHAPE

## Understanding the shape of data

IN COLLABORATION WITH: Laboratoire de mathématiques d'Orsay de  
l'Université de Paris-Sud (LMO)

### DOMAIN

Algorithmics, Programming, Software  
and Architecture

### THEME

Algorithmics, Computer Algebra and  
Cryptology

# Contents

<b>Project-Team DATASHAPE</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>2</b>
<b>2 Overall objectives</b>	<b>3</b>
<b>3 Research program</b>	<b>4</b>
3.1 Algorithmic aspects and new mathematical directions for topological and geometric data analysis	4
3.2 Statistical aspects of topological and geometric data analysis	4
3.3 Topological and geometric approaches for machine learning	5
3.4 Experimental research and software development	5
<b>4 Application domains</b>	<b>5</b>
<b>5 Social and environmental responsibility</b>	<b>6</b>
5.1 Footprint of research activities	6
<b>6 New software and platforms</b>	<b>6</b>
6.1 New software	6
6.1.1 GUDHI	6
6.1.2 Module CGAL: New dD Geometry Kernel	6
<b>7 New results</b>	<b>7</b>
7.1 Algorithmic aspects and new mathematical directions for topological and geometric data analysis	7
7.1.1 Lexicographic optimal homologous chains and applications to point cloud triangulations	7
7.1.2 Tracing isomanifolds in $R^d$ in time polynomial in $d$	7
7.1.3 A compact data structure for high dimensional Coxeter-Freudenthal-Kuhn triangulations	7
7.1.4 Local characterizations for decomposability of 2-parameter persistence modules	8
7.1.5 On rectangle-decomposable 2-parameter persistence modules	8
7.1.6 Decomposition of exact pfd persistence bimodules	9
7.1.7 Homotopy Reconstruction via the Cech Complex and the Vietoris-Rips Complex	9
7.1.8 Recovering the homology of immersed manifolds	9
7.1.9 Computing persistent Stiefel-Whitney classes of line bundles	10
7.2 Statistical aspects of topological and geometric data analysis	10
7.2.1 Optimal quantization of the mean measure and applications to statistical learning	10
7.2.2 DTM-based Filtrations	10
7.2.3 Understanding the Topology and the Geometry of the Space of Persistence Diagrams via Optimal Partial Transport	11
7.2.4 Minimax adaptive estimation in manifold inference	11
7.2.5 Volume Doubling Condition and a Local Poincaré Inequality on Unweighted Random Geometric Graphs	11
7.3 Topological and geometric approaches for machine learning	12
7.3.1 Inverse Problems in Topological Persistence: a Survey	12
7.3.2 Intrinsic Topological Transforms via the Distance Kernel Embedding	12
7.3.3 PLLay: Efficient Topological Layer based on Persistence Landscapes	12
7.3.4 Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks	13
7.3.5 A note on stochastic subgradient descent for persistence-based functionals: convergence and practical aspects	13
7.3.6 ATOL: Measure Vectorization for Automatic Topologically-Oriented Learning	13
7.3.7 Multiparameter Persistence Image for Topological Machine Learning	14
7.4 Miscellaneous	14

7.4.1	Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space	14
7.4.2	Post hoc confidence bounds on false positives using reference families	14
7.4.3	Compressive Statistical Learning with Random Feature Moments	15
7.4.4	Domain Generalization by Marginal Transfer Learning	15
7.4.5	A polynomial time algorithm to compute quantum invariants of 3-manifolds with bounded first Betti number	16
7.4.6	Variable-width contouring for additive manufacturing	16
7.4.7	Mean curvature motion of point cloud varifolds	16
7.4.8	Covering families of triangles	17
<b>8</b>	<b>Bilateral contracts and grants with industry</b>	<b>17</b>
8.1	Bilateral contracts with industry	17
8.2	Bilateral grants with industry	17
<b>9</b>	<b>Partnerships and cooperations</b>	<b>17</b>
9.1	International initiatives	17
9.1.1	Inria international partners	17
9.2	National initiatives	18
9.2.1	ANR	18
9.2.2	Collaboration with other national research institutes	19
9.3	Regional initiatives	19
<b>10</b>	<b>Dissemination</b>	<b>20</b>
10.1	Promoting scientific activities	20
10.1.1	Scientific events: organisation	20
10.1.2	Scientific events: selection	20
10.1.3	Journal	20
10.1.4	Invited talks	20
10.1.5	Leadership within the scientific community	21
10.1.6	Research administration	21
10.2	Teaching - Supervision - Juries	21
10.2.1	Teaching	21
10.2.2	Supervision	22
10.2.3	Juries	23
10.3	Popularization	23
10.3.1	Interventions	23
<b>11</b>	<b>Scientific production</b>	<b>23</b>
11.1	Major publications	23
11.2	Publications of the year	24

## **Project-Team DATASHAPE**

*Creation of the Team: 2016 January 01, updated into Project-Team: 2017 July 01*

### **Keywords**

#### **Computer sciences and digital sciences**

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

#### **Other research topics and application domains**

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

# 1 Team members, visitors, external collaborators

## Research Scientists

- Frédéric Chazal [Team leader, Inria, Senior Researcher, Saclay - Île-de-France, HDR]
- Jean-Daniel Boissonnat [Inria, Senior Researcher, Sophia Antipolis - Méditerranée, HDR]
- Mathieu Carrière [Inria, Researcher, from Oct 2020, Sophia Antipolis - Méditerranée]
- David Cohen-Steiner [Inria, Researcher, Sophia Antipolis - Méditerranée]
- Marc Glisse [Inria, Researcher, Saclay - Île-de-France]
- Jisu Kim [Inria, Starting Research Position, from Mar 2020, Saclay - Île-de-France]
- Clément Maria [Inria, Researcher, Sophia Antipolis - Méditerranée]
- Steve Oudot [Inria, Researcher, Saclay - Île-de-France, HDR]

## Faculty Members

- Gilles Blanchard [Univ Paris-Saclay, from Nov 2020, Saclay - Île-de-France]
- Blanche Buet [Univ Paris-Saclay, from Nov 2020, Saclay - Île-de-France]
- Pierre Pansu [Univ Paris-Saclay, Professor, from Nov 2020, Saclay - Île-de-France]

## Post-Doctoral Fellows

- Kristof Huszar [Inria, from Oct 2020, Sophia Antipolis - Méditerranée]
- Hariprasad Kannan [Inria, until Jan 2020, Saclay - Île-de-France]
- Jisu Kim [Inria, until Feb 2020, Saclay - Île-de-France]
- Theo Lacombe [Inria, from Oct 2020, Saclay - Île-de-France]
- Siddharth Pritam [Inria, from Oct 2020, Sophia Antipolis - Méditerranée]
- Martin Royer [Inria, until Aug 2020, Saclay - Île-de-France]

## PhD Students

- Bertrand Beaufile [SYSNAV, CIFRE, until Jun 2020, Saclay - Île-de-France]
- Nicolas Berkouk [Inria, until Oct 2020, Saclay - Île-de-France]
- Jeremie Capitaio-Miniconi [Univ Paris-Saclay, from Nov 2020, Saclay - Île-de-France]
- Alex Delalande [Inria, Saclay - Île-de-France]
- Vincent Divol [Univ Paris-Sud, Saclay - Île-de-France]
- Olympio Hacquard [Univ Paris-Saclay, from Sep 2020, Saclay - Île-de-France]
- Theo Lacombe [École polytechnique, until Sep 2020, Saclay - Île-de-France]
- Etienne Lasalle [Univ Paris-Sud, Saclay - Île-de-France]
- Vadim Lebovici [École Normale Supérieure de Paris, from Sep 2020, Saclay - Île-de-France]
- Daniel Perez [Ecole normale supérieure Paris-Saclay, from Nov 2020, Saclay - Île-de-France]

- Siddharth Pritam [Inria, until Sep 2020, Sophia Antipolis - Méditerranée]
- Louis Pujol [Univ Paris-Sud, Saclay - Île-de-France]
- Wojciech Reise [Inria, from Sep 2020, Saclay - Île-de-France]
- Owen Rouille [Inria, Sophia Antipolis - Méditerranée]
- Raphael Tinarrage [École Normale Supérieure de Cachan, Saclay - Île-de-France]
- Christophe Vuong [Telecom ParisTech, from Nov 2020, Saclay - Île-de-France]

### **Technical Staff**

- Thomas Bonis [Inria, Engineer, until Jul 2020, Saclay - Île-de-France]
- Rudresh Mishra [Inria, Engineer, from Dec 2020, Saclay - Île-de-France]
- Vincent Rouvreau [Inria, Engineer, Saclay - Île-de-France]

### **Interns and Apprentices**

- Antoine Commaret [École Normale Supérieure de Paris, from Sep 2020, Saclay - Île-de-France]

### **Administrative Assistants**

- Laurence Fontana [Inria, from Oct 2020, Saclay - Île-de-France]
- Sophie Honnorat [Inria, Sophia Antipolis - Méditerranée]

### **External Collaborators**

- Clément Levrard [Univ Denis Diderot, until Sep 2020, Saclay - Île-de-France]
- Bertrand Michel [Univ Pierre et Marie Curie, Saclay - Île-de-France]

## **2 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 overall objective of the DataShape project is to settle the mathematical, statistical and algorithmic foundations of TDA and to disseminate and promote our results in the data science community.

The approach of DataShape relies on the conviction that it is necessary to combine statistical, topological/geometric and computational approaches in a common framework, in order to face the challenges of TDA. Another conviction of DataShape is that TDA needs to be combined with other data sciences approaches and tools to lead to successful real applications. It is necessary for TDA challenges to be simultaneously addressed from the fundamental and application sides.

The team members have actively contributed to the emergence of TDA during the last few years. The variety of expertise, going from fundamental mathematics to software development, and the strong interactions within our team as well as numerous well established international collaborations make our group one of the best to achieve these goals.

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 pursue the development of the GUDHI platform, initiated by the team members and which is becoming a standard tool in TDA, in order to provide an efficient state-of-the-art toolbox for the understanding of the topology and geometry of data. The ultimate goal of DataShape is to develop and promote TDA as a new family of well-founded methods to uncover and exploit the geometry of data.

This also includes the clarification of the position and complementarity of TDA with respect to other approaches and tools in data science. Our objective is also to provide practically efficient and flexible tools that could be used independently, complementarily or in combination with other classical data analysis and machine learning approaches.

### 3 Research program

#### 3.1 Algorithmic aspects and new mathematical directions for 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 and geometric approaches for machine learning

This objective is driven by the problems raised by the use of topological and geometric approaches in machine learning. The goal is both to use our techniques to better understand the role of topological and geometric structures in machine learning problems and to apply our TDA tools to develop specialized topological approaches to be used in combination with other machine learning methods.

### 3.4 Experimental research and software development

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

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

Some of the algorithms implemented on the platform will also be interfaced to other software platform, such as the R software<sup>1</sup> 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

Our work is mostly of a fundamental mathematical and algorithmic nature but finds a variety of applications in data analysis, e.g., in material science, biology, sensor networks, 3D shape analysis and processing, to name a few.

More specifically, DATASHAPE is working on the analysis of trajectories obtained from inertial sensors (PhD thesis of Bertrand Beaufils with Sysnav) and, more generally on the development of new TDA methods for Machine Learning and Artificial Intelligence for (multivariate) time-dependent data from various kinds of sensors in collaboration with Fujitsu.

DATASHAPE is also working in collaboration with the University of Columbia in New-York, especially with the Rabadan lab, in order to improve bioinformatics methods and analyses for single cell genomic data. For instance, there is a lot of work whose aim is to use TDA tools such as persistent homology and the Mapper algorithm to characterize, quantify and study statistical significance of biological phenomena that occur in large scale single cell data sets. Such biological phenomena include, among others: the cell cycle, functional differentiation of stem cells, and immune system responses (such as the spatial response on the tissue location, and the genomic response with protein expression) to breast cancer.

---

<sup>1</sup><https://www.r-project.org/>



## 5 Social and environmental responsibility

### 5.1 Footprint of research activities

The weekly research seminar of DATASHAPE is now taking place online, and travels for the team members have decreased a lot this year, mainly because of the COVID-19 pandemic.

## 6 New software and platforms

### 6.1 New software

#### 6.1.1 GUDHI

**Name:** Geometric Understanding in Higher Dimensions

**Keywords:** Computational geometry, Topology, Clustering

**Scientific Description:** The Gudhi library is an open source library for Computational Topology and Topological Data Analysis (TDA). It offers state-of-the-art algorithms to construct various types of simplicial complexes, data structures to represent them, and algorithms to compute geometric approximations of shapes and persistent homology.

The GUDHI library offers the following interoperable modules:

. Complexes: + Cubical + Simplicial: Rips, Witness, Alpha and Čech complexes + Cover: Nerve and Graph induced complexes . Data structures and basic operations: + Simplex tree, Skeleton blockers and Toplex map + Construction, update, filtration and simplification . Topological descriptors computation . Manifold reconstruction . Topological descriptors tools: + Bottleneck and Wasserstein distance + Statistical tools + Persistence diagram and barcode

**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.

**News of the Year:** - DTM Rips complex - Edge Collapse - Time delay embedding - Clustering (ToMaTo) - Atol - Persistence representations - Weighted alpha complex - Subsampling - Periodic (weighted or not) 3d Alpha complex - pip packages

**URL:** <https://gudhi.inria.fr/>

**Authors:** Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, Mariette Yvinec, Vincent Rouvreau, Clément Jamin, David Salinas, François Godi, Mathieu Carrière, Pawel Dlotko, Siargey Kachanovich, Siddharth Pritam, Theo Lacombe, Steve Oudot, Bertrand Michel, Frédéric Chazal

**Contacts:** Jean-Daniel Boissonnat, Marc Glisse, Vincent Rouvreau

**Participants:** Clément Maria, François Godi, David Salinas, Jean-Daniel Boissonnat, Marc Glisse, Mariette Yvinec, Pawel Dlotko, Siargey Kachanovich, Vincent Rouvreau, Mathieu Carrière, Bertrand Michel, Clément Jamin, Siddharth Pritam, Theo Lacombe, Frédéric Chazal, Steve Oudot

#### 6.1.2 Module CGAL: New dD Geometry Kernel

**Keyword:** Computational geometry

**Functional Description:** This package of CGAL (Computational Geometry Algorithms Library) provides the basic geometric types (point, vector, etc) and operations (orientation test, etc) used by geometric algorithms in arbitrary dimension. It uses filters for efficient exact predicates.

**Release Contributions:** New predicates for (weighted) alpha complexes, performance improvements.

**URL:** <http://www.cgal.org/>

**Author:** Marc Glisse

**Contact:** Marc Glisse

## 7 New results

### 7.1 Algorithmic aspects and new mathematical directions for topological and geometric data analysis

#### 7.1.1 Lexicographic optimal homologous chains and applications to point cloud triangulations

**Participants** David Cohen-Steiner.

*In collaboration with André Lieutier (Dassault Systèmes) and Julien Vuillamy (Titane team, Inria Sophia-Antipolis).*

This work [30] considers a particular case of the Optimal Homologous Chain Problem (OHCP), where optimality is meant as a minimal lexicographic order on chains induced by a total order on simplices. The matrix reduction algorithm used for persistent homology is used to derive polynomial algorithms solving this problem instance, whereas OHCP is NP-hard in the general case. The complexity is further improved to a quasilinear algorithm by leveraging a dual graph minimum cut formulation when the simplicial complex is a strongly connected pseudomanifold. We then show how this particular instance of the problem is relevant, by providing an application in the context of point cloud triangulation

#### 7.1.2 Tracing isomanifolds in $\mathbb{R}^d$ in time polynomial in $d$

**Participants** Jean-Daniel Boissonnat, Siargey Kachanovich.

*In collaboration with Mathijs Wintraecken (IST Austria).*

Isomanifolds are the generalization of isosurfaces to arbitrary dimension and codimension, i.e. submanifolds of  $\mathbb{R}^d$  defined as the zero set of some multivariate multivalued smooth function  $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d-n}$ , where  $n$  is the intrinsic dimension of the manifold. A natural way to approximate a smooth isomanifold  $M$  is to consider its Piecewise-Linear (PL) approximation  $\hat{M}$  based on a triangulation  $\mathcal{T}$  of the ambient space  $\mathbb{R}^d$ . In [36], we describe a simple algorithm to trace isomanifolds from a given starting point. The algorithm works for arbitrary dimensions  $n$  and  $d$ , and any precision  $D$ . Our main result is that, when  $f$  (or  $M$ ) has bounded complexity, the complexity of the algorithm is polynomial in  $d$  and  $\delta = 1/D$  (and unavoidably exponential in  $n$ ). Since it is known that for  $\delta = \Omega(d^{2.5})$ ,  $\hat{M}$  is  $O(D^2)$ -close and isotopic to  $M$ , our algorithm produces a faithful PL-approximation of isomanifolds of bounded complexity in time polynomial in  $d$ . Combining this algorithm with dimensionality reduction techniques, the dependency on  $d$  in the size of  $\hat{M}$  can be completely removed with high probability. We also show that the algorithm can handle isomanifolds with boundary and, more generally, isostratifolds. The algorithm has been implemented and experimental results are reported, showing that it is practical and can handle cases that are far ahead of the state-of-the-art.

#### 7.1.3 A compact data structure for high dimensional Coxeter-Freudenthal-Kuhn triangulations

**Participants** Jean-Daniel Boissonnat, Sargey Kachanovich.

*In collaboration with Mathijs Wintraecken (IST Austria).*

In [45], we consider a family of highly regular triangulations of  $\mathbb{R}^d$  that can be stored and queried efficiently in high dimensions. This family consists of Freudenthal-Kuhn triangulations and their images through affine mappings, among which are the celebrated Coxeter triangulations of type  $\tilde{A}_d$ . Those triangulations have major advantages over grids in applications in high dimensions like interpolation of functions and manifold sampling and meshing. We introduce an elegant and very compact data structure to implicitly store the full facial structure of such triangulations. This data structure allows to locate a point and to retrieve the faces or the cofaces of a simplex of any dimension in an output sensitive way. The data structure has been implemented and experimental results are presented.

#### 7.1.4 Local characterizations for decomposability of 2-parameter persistence modules

**Participants** Steve Oudot, Vadim Lebovici.

*In collaboration with Magnus Botnan (Vrije Universiteit Amsterdam)*

In this work [48] we investigate the existence of sufficient local conditions under which representations of a given poset will be guaranteed to decompose as direct sums of indecomposables from a given class. Our indecomposables of interest belong to the so-called interval modules, which by definition are indicator representations of intervals in the poset. In contexts where the poset is the product of two totally ordered sets (which corresponds to the setting of 2-parameter persistence in topological data analysis), we show that the whole class of interval modules itself does not admit such a local characterization, even when locality is understood in a broad sense. By contrast, we show that the subclass of rectangle modules does admit such a local characterization, and furthermore that it is, in some precise sense, the largest subclass to do so.

#### 7.1.5 On rectangle-decomposable 2-parameter persistence modules

**Participants** Steve Oudot, Vadim Lebovici.

*In collaboration with Magnus Botnan (Vrije Universiteit Amsterdam)*

This work [28] addresses two questions: (a) can we identify a sensible class of 2-parameter persistence modules on which the rank invariant is complete? (b) can we determine efficiently whether a given 2-parameter persistence module belongs to this class? We provide positive answers to both questions, and our class of interest is that of rectangle-decomposable modules. Our contributions include: on the one hand, a proof that the rank invariant is complete on rectangle-decomposable modules, together with an inclusion-exclusion formula for counting the multiplicities of the summands; on the other hand, algorithms to check whether a module induced in homology by a bifiltration is rectangle-decomposable, and to decompose it in the affirmative, with a better complexity than state-of-the-art decomposition methods for general 2-parameter persistence modules. Our algorithms are backed up by a new structure theorem, whereby a 2-parameter persistence module is rectangle-decomposable if, and only if, its restrictions to squares are. This local characterization is key to the efficiency of our algorithms, and it generalizes previous conditions derived for the smaller class of block-decomposable modules. It also

admits an algebraic formulation that turns out to be a weaker version of the one for block-decomposability. By contrast, we show that general interval-decomposability does not admit such a local characterization, even when locality is understood in a broad sense. Our analysis focuses on the case of modules indexed over finite grids.

### 7.1.6 Decomposition of exact pfd persistence bimodules

**Participants** Jérémy Cochoy, Steve Oudot.

In this work [22] 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 on  $\mathbb{R}^2$  not being a total order, 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.

### 7.1.7 Homotopy Reconstruction via the Čech Complex and the Vietoris-Rips Complex

**Participants** Frédéric Chazal, Jisu Kim.

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

In this work [32], we derive conditions under which the reconstruction of a target space is topologically correct via the Čech complex or the Vietoris-Rips complex obtained from possibly noisy point cloud data. We provide two novel theoretical results. First, we describe sufficient conditions under which any non-empty intersection of finitely many Euclidean balls intersected with a positive reach set is contractible, so that the Nerve theorem applies for the restricted Čech complex. Second, we demonstrate the homotopy equivalence of a positive  $\mu$ -reach set and its offsets. Applying these results to the restricted Čech complex and using the interleaving relations with the Čech complex (or the Vietoris-Rips complex), we formulate conditions guaranteeing that the target space is homotopy equivalent to the Čech complex (or the Vietoris-Rips complex), in terms of the  $\mu$ -reach. Our results sharpen existing results.

### 7.1.8 Recovering the homology of immersed manifolds

**Participants** Raphaël Tinarrage.

Given a sample of an abstract manifold immersed in some Euclidean space, we describe [68] a way to recover the singular homology of the original manifold. It consists in estimating its tangent bundle—seen as subset of another Euclidean space—in a measure theoretic point of view, and in applying measure-based filtrations for persistent homology. The construction we propose is consistent and stable, and does not involve the knowledge of the dimension of the manifold. In order to obtain quantitative results, we introduce the normal reach, which is a notion of reach suitable for an immersed manifold.

### 7.1.9 Computing persistent Stiefel-Whitney classes of line bundles

**Participants** Raphaël Tinarrage.

We propose [67] a definition of persistent Stiefel-Whitney classes of vector bundle filtrations. It relies on seeing vector bundles as subsets of some Euclidean spaces. The usual Čech filtration of such a subset can be endowed with a vector bundle structure, that we call a Čech bundle filtration. We show that this construction is stable and consistent. When the dataset is a finite sample of a line bundle, we implement an effective algorithm to compute its persistent Stiefel-Whitney classes. In order to use simplicial approximation techniques in practice, we develop a notion of weak simplicial approximation. As a theoretical example, we give an in-depth study of the normal bundle of the circle, which reduces to understanding the persistent cohomology of the torus knot (1,2).

## 7.2 Statistical aspects of topological and geometric data analysis

### 7.2.1 Optimal quantization of the mean measure and applications to statistical learning

**Participants** Frédéric Chazal, Martin Royer.

*In collaboration with Clément Levrard (Université Paris-Diderot)*

This work [51] addresses the case where data come as point sets, or more generally as discrete measures. Our motivation is twofold: first we intend to approximate with a compactly supported measure the mean of the measure generating process, that coincides with the intensity measure in the point process framework, or with the expected persistence diagram in the framework of persistence-based topological data analysis. To this aim we provide two algorithms that we prove almost minimax optimal. Second we build from the estimator of the mean measure a vectorization map, that sends every measure into a finite-dimensional Euclidean space, and investigate its properties through a clustering-oriented lens. In a nutshell, we show that in a mixture of measure generating process, our technique yields a representation in  $\mathbb{R}^k$ , for  $k \in \mathbb{N}^*$  that guarantees a good clustering of the data points with high probability. Interestingly, our results apply in the framework of persistence-based shape classification via the ATOL procedure. At last, we assess the effectiveness of our approach on simulated and real datasets, encompassing text classification and large-scale graph classification.

### 7.2.2 DTM-based Filtrations

**Participants** Frédéric Chazal, Marc Glisse, Raphael Tinarrage.

*In collaboration with H. Anai, H. Inakoshi and Y. Umeda (Fujitsu, Japan)*

Despite strong stability properties, the persistent homology of filtrations classically used in Topological Data Analysis, such as, e.g. the Čech or Vietoris-Rips filtrations, are very sensitive to the presence of outliers in the data from which they are computed. In this work [12], we introduce and study a new family of filtrations, the DTM-filtrations, built on top of point clouds in the Euclidean space which are more robust to noise and outliers. The approach adopted in this work relies on the notion of distance-to-measure functions and extends some previous work on the approximation of such functions.

### 7.2.3 Understanding the Topology and the Geometry of the Space of Persistence Diagrams via Optimal Partial Transport

**Participants** Vincent Divol, Théo Lacombe.

Despite the obvious similarities between the metrics used in topological data analysis and those of optimal transport, an optimal-transport based formalism to study persistence diagrams and similar topological descriptors has yet to come. In this work [17], by considering the space of persistence diagrams as a space of discrete measures, and by observing that its metrics can be expressed as optimal partial transport problems, we introduce a generalization of persistence diagrams, namely Radon measures supported on the upper half plane. Such measures naturally appear in topological data analysis when considering continuous representations of persistence diagrams (e.g. persistence surfaces) but also as limits for laws of large numbers on persistence diagrams or as expectations of probability distributions on the persistence diagrams space. We explore topological properties of this new space, which will also hold for the closed subspace of persistence diagrams. New results include a characterization of convergence with respect to Wasserstein metrics, a geometric description of barycenters (Fréchet means) for any distribution of diagrams, and an exhaustive description of continuous linear representations of persistence diagrams. We also showcase the strength of this framework to study random persistence diagrams by providing several statistical results made meaningful thanks to this new formalism.

### 7.2.4 Minimax adaptive estimation in manifold inference

**Participants** Vincent Divol.

In this work [57], we focus on the problem of manifold estimation: given a set of observations sampled close to some unknown submanifold  $M$ , one wants to recover information about the geometry of  $M$ . Minimax estimators which have been proposed so far all depend crucially on the a priori knowledge of some parameters quantifying the regularity of  $M$  (such as its reach), whereas those quantities will be unknown in practice. Our contribution to the matter is twofold: first, we introduce a one-parameter family of manifold estimators  $(M_t)$ ,  $t \geq 0$ , and show that for some choice of  $t$  (depending on the regularity parameters), the corresponding estimator is minimax on the class of models of  $\mathcal{C}^2$  manifolds introduced in [Genovese et al., Manifold estimation and singular deconvolution under Hausdorff loss]. Second, we propose a completely data-driven selection procedure for the parameter  $t$ , leading to a minimax adaptive manifold estimator on this class of models. This selection procedure actually allows to recover the sample rate of the set of observations, and can therefore be used as an hyperparameter in other settings, such as tangent space estimation.

### 7.2.5 Volume Doubling Condition and a Local Poincaré Inequality on Unweighted Random Geometric Graphs

**Participants** Gilles Blanchard.

*In collaboration with Franziska Göbel (Institute of Mathematics, University of Potsdam)*

The aim of this work [59] is to establish two fundamental measure-metric properties of particular random geometric graphs. We consider  $\varepsilon$ -neighborhood graphs whose vertices are drawn independently and identically distributed from a common distribution defined on a regular submanifold of  $\mathbb{R}^K$ . We show that a volume doubling condition (VD) and local Poincaré inequality (LPI) hold for the random geometric graph (with high probability, and uniformly over all shortest path distance balls in a certain radius range) under suitable regularity conditions of the underlying submanifold and the sampling distribution.

## 7.3 Topological and geometric approaches for machine learning

### 7.3.1 Inverse Problems in Topological Persistence: a Survey

**Participants** Steve Oudot.

*In collaboration with Elchanan Solomon (Duke University)*

In this survey [23], we review the literature on inverse problems in topological persistence theory. The first half of the survey is concerned with the question of surjectivity, i.e. the existence of right inverses, and the second half focuses on injectivity, i.e. left inverses. Throughout, we highlight the tools and theorems that underlie these advances, and direct the reader's attention to open problems, both theoretical and applied.

### 7.3.2 Intrinsic Topological Transforms via the Distance Kernel Embedding

**Participants** Clément Maria, Steve Oudot.

*In collaboration with Elchanan Solomon (Duke University)*

Topological transforms are parametrized families of topological invariants, which, by analogy with transforms in signal processing, are much more discriminative than single measurements. The first two topological transforms to be defined were the Persistent Homology Transform and Euler Characteristic Transform, both of which apply to shapes embedded in Euclidean space. The contribution of this work [34] is to define topological transforms that depend only on the intrinsic geometry of a shape, and hence are invariant to the choice of embedding. To that end, given an abstract metric measure space, we define an integral operator whose eigenfunctions are used to compute sublevel set persistent homology. We demonstrate that this operator, which we call the distance kernel operator, enjoys desirable stability properties, and that its spectrum and eigenfunctions concisely encode the large-scale geometry of our metric measure space. We then define a number of topological transforms using the eigenfunctions of this operator, and observe that these transforms inherit many of the stability and injectivity properties of the distance kernel operator.

### 7.3.3 PLLay: Efficient Topological Layer based on Persistence Landscapes

**Participants** Frédéric Chazal, Jisu Kim.

*In collaboration with K. Kim, J.S. Kim, L. Wasserman (Carnegie Mellon University) and M. Zaheer (Google Research)*

In this work [33], we propose PLLay, a novel topological layer for general deep learning models based on persistence landscapes, in which we can efficiently exploit the underlying topological features of the input data structure. We show differentiability with respect to layer inputs, for a general persistent homology with arbitrary filtration. Thus, our proposed layer can be placed anywhere in the network and feed critical information on the topological features of input data into subsequent layers to improve the learnability of the networks toward a given task. A task optimal structure of PLLay is learned during training via backpropagation, without requiring any input featurization or data preprocessing. We provide a novel adaptation for the DTM function-based filtration, and show that the proposed layer is robust



against noise and outliers through a stability analysis. We demonstrate the effectiveness of our approach by classification experiments on various datasets.

#### 7.3.4 Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks

**Participants** Frédéric Chazal.

*In collaboration with M. Dindin and Y. Umeda (Fujitsu, Japan)*

This work [31] presents an innovative and generic deep learning approach to monitor heart conditions from ECG signals. We focus our attention on both the detection and classification of abnormal heartbeats, known as arrhythmia. We strongly insist on generalization throughout the construction of a deep learning model that turns out to be effective for new unseen patient. The novelty of our approach relies on the use of topological data analysis as basis of our multichannel architecture, to diminish the bias due to individual differences. We show that our structure reaches the performances of the state-of-the-art methods regarding arrhythmia detection and classification.

#### 7.3.5 A note on stochastic subgradient descent for persistence-based functionals: convergence and practical aspects

**Participants** Mathieu Carrière, Frédéric Chazal, Marc Glisse, Hari Kannan, Théo Lacombe.

*In collaboration with Yiuchi Ike (Fujitsu, Japan)*

Solving optimization tasks based on functions and losses with a topological flavor is a very active and growing field of research in Topological Data Analysis, with plenty of applications in non-convex optimization, statistics and machine learning. All of these methods rely on the fact that most of the topological constructions are actually stratifiable and differentiable almost everywhere. However, the corresponding gradient and associated code is always anchored to a specific application and/or topological construction, and do not come with theoretical guarantees. In this work [50], we study the differentiability of a general functional associated with the most common topological construction, that is, the persistence map, and we prove a convergence result of stochastic subgradient descent for such a functional. This result encompasses all the constructions and applications for topological optimization in the literature, and comes with code that is easy to handle and mix with other non-topological constraints, and that can be used to reproduce the experiments described in the literature.

#### 7.3.6 ATOL: Measure Vectorization for Automatic Topologically-Oriented Learning

**Participants** Frédéric Chazal, Martin Royer.

*In collaboration with Clément Levrard (Université Paris-Diderot), Yiuchi Ike and Yuhei Umeda (Fujitsu, Japan).*

Robust topological information commonly comes in the form of a set of persistence diagrams, finite measures that are in nature uneasy to affix to generic machine learning frameworks. In this work [65], we introduce a fast, learnt, unsupervised vectorization method for measures in Euclidean spaces and use it for reflecting underlying changes in topological behaviour in machine learning contexts. The algorithm is



simple and efficiently discriminates important space regions where meaningful differences to the mean measure arise. It is proven to be able to separate clusters of persistence diagrams. We showcase the strength and robustness of our approach on a number of applications, from emulous and modern graph collections where the method reaches state-of-the-art performance to a geometric synthetic dynamical orbits problem. The proposed methodology comes with a single high level tuning parameter: the total measure encoding budget.

### 7.3.7 Multiparameter Persistence Image for Topological Machine Learning

**Participants** Mathieu Carrière.

*In collaboration with Andrew Blumberg (Université de Columbia, New-York, USA).*

In the last decade, there has been increasing interest in topological data analysis, a new methodology for using geometric structures in data for inference and learning. A central theme in the area is the idea of persistence, which in its most basic form studies how measures of shape change as a scale parameter varies. There are now a number of frameworks that support statistics and machine learning in this context. However, in many applications there are several different parameters one might wish to vary: for example, scale and density. In contrast to the one-parameter setting, techniques for applying statistics and machine learning in the setting of multiparameter persistence are not well understood due to the lack of a concise representation of the results. We introduce a new descriptor for multiparameter persistence, which we call the Multiparameter Persistence Image, that is suitable for machine learning and statistical frameworks, is robust to perturbations in the data, has finer resolution than existing descriptors based on slicing, and can be efficiently computed on data sets of realistic size. Moreover, we demonstrate its efficacy by comparing its performance to other multiparameter descriptors on several classification tasks.

## 7.4 Miscellaneous

### 7.4.1 Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space

**Participants** Frédéric Chazal, Alex Delalande.

*In collaboration with Quentin Mérigot (Laboratoire de Mathématiques d'Orsay, Univ. Paris-Saclay)*

This work [35] studies an explicit embedding of the set of probability measures into a Hilbert space, defined using optimal transport maps from a reference probability density. This embedding linearizes to some extent the 2-Wasserstein space, and enables the direct use of generic supervised and unsupervised learning algorithms on measure data. Our main result is that the embedding is (bi-)Hölder continuous, when the reference density is uniform over a convex set, and can be equivalently phrased as a dimension-independent Hölder-stability results for optimal transport maps.

### 7.4.2 Post hoc confidence bounds on false positives using reference families

**Participants** Gilles Blanchard.

*In collaboration with Étienne Roquain (LPSM, Sorbonne université), Pierre Neuvial (IMT, Toulouse Université)*

In this work [14], we follow a post-hoc, "user-agnostic" approach to false discovery control in a large-scale

multiple testing framework, as introduced by Genovese and Wasserman (2006), Goeman and Solari (2011): the statistical guarantee on the number of correct rejections must hold for any set of candidate items, possibly selected by the user after having seen the data. To this end, we introduce a novel point of view based on a family of reference rejection sets and a suitable criterion, namely the joint-family-wise-error rate over that family (JER for short). First, we establish how to derive post hoc bounds from a given JER control and analyze some general properties of this approach. We then develop procedures for controlling the JER in the case where reference regions are  $p$ -value level sets. These procedures adapt to dependencies and to the unknown quantity of signal (via a step-down principle). We also show interesting connections to confidence envelopes of Meinshausen (2006); Genovese and Wasserman (2006), the closed testing based approach of Goeman and Solari (2011) and to the higher criticism of Donoho and Jin (2004). Our theoretical statements are supported by numerical experiments.

Published in *Annals of Statistics*, 2020.

### 7.4.3 Compressive Statistical Learning with Random Feature Moments

**Participants** Gilles Blanchard.

*In collaboration with Rémi Gribonval (INRIA Lyon), Nicolas Keriven (CNRS, GIPSA, Université Rhône-Alpes), Yan Traonmilin (CNRS, IMB, Université Bordeaux)*

We introduce in this work [20] a general framework –compressive statistical learning– for resource-efficient large-scale learning: the training collection is compressed in one pass into a low-dimensional sketch (a vector of random empirical generalized moments) that captures the information relevant to the considered learning task. A near-minimizer of the risk is computed from the sketch through the solution of a nonlinear least squares problem. We investigate sufficient sketch sizes to control the generalization error of this procedure. The framework is illustrated on compressive PCA, compressive clustering, and compressive Gaussian mixture Modeling with fixed known variance. The latter two are further developed in a companion paper.

Accepted for publication in *Mathematical Statistics and Learning*, 2021.

### 7.4.4 Domain Generalization by Marginal Transfer Learning

**Participants** Gilles Blanchard.

*In collaboration with Aniket Anand Deshmukh (Microsoft Research), Urun Dogan (Microsoft Research), Gyemin Lee (Seoul University for Science and Technology), Clayton Scott (University of Michigan)*

In the problem of domain generalization (DG), there are labeled training data sets from several related prediction problems, and the goal is to make accurate predictions on future unlabeled data sets that are not known to the learner. This problem arises in several applications where data distributions fluctuate because of environmental, technical, or other sources of variation. In the work [42] we introduce a formal framework for DG, and argue that it can be viewed as a kind of supervised learning problem by augmenting the original feature space with the marginal distribution of feature vectors. While our framework has several connections to conventional analysis of supervised learning algorithms, several unique aspects of DG require new methods of analysis. This work lays the learning theoretic foundations of domain generalization, building on our earlier work where the problem of DG was introduced. We present two formal models of data generation, corresponding notions of risk, and distribution-free generalization error analysis. By focusing our attention on kernel methods, we also provide more quantitative results and a universally consistent algorithm. An efficient implementation is provided for this algorithm, which is experimentally compared to a pooling strategy on one synthetic and three real-world data sets.

Published in *Journal of Machine Learning Research*, 2021.

#### 7.4.5 A polynomial time algorithm to compute quantum invariants of 3-manifolds with bounded first Betti number

**Participants** Clément Maria.

*In collaboration with Jonathan Spreer (The University of Sydney, Australia)*

In this article, we introduce a fixed parameter tractable algorithm for computing the Turaev-Viro invariants  $TV_{4,q}$ , using the dimension of the first homology group of the manifold as parameter. This is, to our knowledge, the first parameterised algorithm in computational 3-manifold topology using a topological parameter. The computation of  $TV_{4,q}$  is known to be sharp-P-hard in general; using a topological parameter provides an algorithm polynomial in the size of the input triangulation for the extremely large family of 3-manifolds with first homology group of bounded rank. Our algorithm is easy to implement and running times are comparable with running times to compute integral homology groups for standard libraries of triangulated 3-manifolds. The invariants we can compute this way are powerful: in combination with integral homology and using standard data sets we are able to roughly double the pairs of 3-manifolds we can distinguish. We hope this qualifies  $TV_{4,q}$  to be added to the short list of standard properties (such as orientability, connectedness, Betti numbers, etc.) that can be computed ad-hoc when first investigating an unknown triangulation.

Published in the journal on Foundations of Computational Mathematics (FoCM) 2020.

#### 7.4.6 Variable-width contouring for additive manufacturing

**Participants** Marc Glisse.

*In collaboration with Samuel Hornus, Sylvain Lefebvre, Jonàs Martínez (Inria team MFX), Olivier Devillers, Sylvain Lazard, Monique Teillaud (Inria team Gamble) and Tim Kuipers (Delft University of Technology, Pays-Bas).*

In most layered additive manufacturing processes, a tool solidifies or deposits material while following pre-planned trajectories to form solid beads. Many interesting problems arise in this context, among which one concerns the planning of trajectories for filling a planar shape as densely as possible. This is the problem we tackle in the present work [21]. Recent works have shown that allowing the bead width to vary along the trajectories helps increase the filling density. We present a novel technique that, given a deposition width range, constructs a set of closed beads whose width varies within the prescribed range and fill the input shape. The technique outperforms the state of the art in important metrics: filling density (while still guaranteeing the absence of bead overlap) and trajectories smoothness. We give a detailed geometric description of our algorithm, explore its behavior on example inputs and provide a statistical comparison with the state of the art. We show that it is possible to obtain high quality fabricated layers on commodity FDM printers.

#### 7.4.7 Mean curvature motion of point cloud varifolds

**Participants** Blanche Buet.

*In collaboration with Martin Rumpf (University of Bonn)*

This paper [49] investigates a discretization scheme for mean curvature motion on point cloud varifolds with particular emphasis on singular evolutions. To define the varifold a local covariance analysis is applied to compute an approximate tangent plane for the points in the cloud. The core ingredient of the mean curvature motion model is the regularization of the first variation of the varifold via convolution with kernels with small stencil. Consistency with the evolution velocity for a smooth surface is proven if a sufficiently small stencil and a regular sampling are taken into account. Furthermore, an implicit and a semiimplicit time discretization are derived. The implicit scheme comes with discrete barrier properties known for the smooth, continuous evolution, whereas the semiimplicit still ensures in all our numerical experiments very good approximation properties while being easy to implement. It is shown that the proposed method is robust with respect to noise and recovers the evolution of smooth curves as well as the formation of singularities such as triple points in 2D or minimal cones in 3D.

#### 7.4.8 Covering families of triangles

**Participants** Marc Glisse.

*In collaboration with Olivier Devillers, Ji-Won Park (Inria team Gamble) and Otfried Cheong (KAIST, Corée du sud).*

A cover for a family  $F$  of sets in the plane is a set into which every set in  $F$  can be isometrically moved. We are interested in the convex cover of smallest area for a given family of triangles. Park and Cheong conjectured that any family of triangles of bounded diameter has a smallest convex cover that is itself a triangle. The conjecture is equivalent to the claim that for every convex set  $X$  there is a triangle  $Z$  whose area is not larger than the area of  $X$ , such that  $Z$  covers the family of triangles contained in  $X$ . In this work [52], we prove this claim for the case where a diameter of  $X$  lies on its boundary. We also give a complete characterization of the smallest convex cover for the family of triangles contained in a half-disk, and for the family of triangles contained in a square. In both cases, this cover is a triangle.

## 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 inertial sensors (CIFRE PhD of Bertrand Beauflis).
- Research collaboration with Fujitsu on the development of new TDA methods and tools for Machine learning and Artificial Intelligence (started in Dec 2017).
- Research collaboration with MetaFora on the development of new TDA-based and statistical methods for the analysis of cytometric data (started in Nov. 2019).

### 8.2 Bilateral grants with industry

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

## 9 Partnerships and cooperations

### 9.1 International initiatives

#### 9.1.1 Inria international partners

##### Informal international partners

- TopStat group (L. Wasserman and A. Rinaldo) at Carnegie Mellon: DataShape maintains a long-standing collaboration with this group since several years with several joint publications.

## 9.2 National initiatives

### 9.2.1 ANR

#### ANR ASPAG

**Participants** Marc Glisse.

- Acronym : ASPAG.

- Type : ANR blanc.

- Title : Analysis and Probabilistic Simulations of Geometric Algorithms.

- Coordinator : Olivier Devillers (équipe Inria Gamble).

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

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

- Abstract:

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

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

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

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

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

#### ANR Chair in AI

**Participants** Frédéric Chazal, Marc Glisse, Louis Pujol, Wojciech Riese.

- Acronym : TopAI

- Type : ANR Chair in AI.

- Title : Topological Data Analysis for Machine Learning and AI
- Coordinator : Frédéric Chazal
- Duration : 4 years from September 2020 to August 2024.
- Others Partners: Two industrial partners, the French SME Sysnav and the French start-up MetaFora.
- Abstract:

The TopAI project aims at developing a world-leading research activity on topological and geometric approaches in Machine Learning (ML) and AI with a double academic and industrial/societal objective. First, building on the strong expertise of the candidate and his team in TDA, TopAI aims at designing new mathematically well-founded topological and geometric methods and tools for Data Analysis and ML and to make them available to the data science and AI community through state-of-the-art software tools. Second, thanks to already established close collaborations and the strong involvement of French industrial partners, TopAI aims at exploiting its expertise and tools to address a set of challenging problems with high societal and economic impact in personalized medicine and AI-assisted medical diagnosis.

#### ANR ALGOKNOT

**Participants** Clément Maria.

- Acronym : ALGOKNOT.
- Type : ANR Jeune Chercheuse Jeune Chercheur.
- Title : Algorithmic and Combinatorial Aspects of Knot Theory.
- Coordinator : Clément Maria.
- Duration : 2020 – 2023 (3 years).
- Abstract: The project AlgoKnot aims at strengthening our understanding of the computational and combinatorial complexity of the diverse facets of knot theory, as well as designing efficient algorithms and software to study their interconnections.
- See also: <https://www-sop.inria.fr/members/Clement.Maria/>

#### 9.2.2 Collaboration with other national research institutes

##### SHOM

**Participants** Steve Oudot.

Research collaboration between DataShape and the Service Hydrographique et Océanographique de la Marine (SHOM) on bathymetric data analysis using a combination of TDA and deep learning techniques. This collaboration is funded by the AMI IA *Améliorer la cartographie du littoral*.

##### IFPEN

**Participants** Frédéric Chazal, Marc Glisse, Jisu Kim.

Research collaboration between DataShape and IFPEN on TDA applied to various problems issued from energy transition and sustainable mobility.

### 9.3 Regional initiatives

#### PhD<sup>2</sup> CytoPart

**Participants** Marc Glisse, Louis Pujol.

- Acronym : CytoPart.
- Type : Paris Region PhD<sup>2</sup>.
- Title : Partitionnement de données cytométriques.

The Île-de-France region funds one PhD thesis supervised by Pascal Massart (Inria team Celeste) and Marc Glisse, in collaboration with Metafora biosystems, a company specialized in the analysis of cells through their metabolism. The goal of the project is to improve clustering for this particular type of data.

## 10 Dissemination

### 10.1 Promoting scientific activities

#### 10.1.1 Scientific events: organisation

- F. Chazal was the co-organizer of a Workshop on Topological Data Analysis and beyonds at NeurIPS 2020 (<https://tda-in-ml.github.io/>)

#### 10.1.2 Scientific events: selection

##### Member of the conference program committees

- Marc Glisse was a member of the Program Committee of the International Symposium on Computational Geometry (SoCG), June 2020.
- Gilles Blanchard was an Area Chair for the NeurIPS 2020 conference.

#### 10.1.3 Journal

##### Member of the editorial boards

- Jean-Daniel Boissonnat is a member of the Editorial Board of the Journal of the ACM.
- Jean-Daniel Boissonnat is a member of the Editorial Board of Discrete and Computational Geometry (Springer).
- Frédéric Chazal is a member of the Editorial Board of Discrete and Computational Geometry (Springer).
- Frédéric Chazal is a member of the Editorial Board of Graphical Models (Elsevier).
- Frédéric Chazal is a member of the Scientific Board of Journal of Applied and Computational Topology (Springer), and Editor-in-Chief since January 1st 2021.
- Gilles Blanchard is a member of the Editorial Boards of Bernoulli, Electronic Journal of Statistics, and Annales de l'Institut Henri Poincaré Probability and Statistics.
- Steve Oudot is a member of the Editorial Board of the Journal of Computational Geometry.

#### 10.1.4 Invited talks

- Steve Oudot. Two Decomposition Results for Bipersistence Modules. MFO Workshop on Representation Theory of Quivers and Finite Dimensional Algebras, Oberwolfach, Germany, January 2020.
- Frédéric Chazal. Approches topologiques et géométriques pour l'apprentissage statistique, théorie et pratique, EDF and System X workshop, September 2020.
- Frédéric Chazal. Learning linear representations of persistence diagrams: mathematical aspects and applications. Applied Machine Learning Days at EPFL 2020, January 2020.

- Jean-Daniel Boissonnat. Delaunay triangulation of manifolds. Inaugural conference at the web-seminar series on Applications of Geometry and Topology (GEOTOP-A), January 2020.
- Blanche Buet. Weak and approximate curvatures of a measure: a varifold perspective. Mathematics and Image Analysis MIA'21, January 2021.

### 10.1.5 Leadership within the scientific community

- Frédéric Chazal is co-responsible, with E. Scornett (Ecole Polytechnique), of the “programme Maths-IA” of the Fondation Mathématique Jacques Hadamard (FMJH).
- Frédéric Chazal is a member of the “Comité de pilotage” of the SIGMA group at SMAI.
- Steve Oudot is co-responsible, with L. Castelli-Aleardi, of the GT GeoAlgo within the GdR-IM.

### 10.1.6 Research administration

- Marc Glisse is president of the CDT at Inria Saclay.
- Steve Oudot is president of the Commission Scientifique at Inria Saclay.
- Frédéric Chazal is a member of the Graduate School in Mathematics at Université Paris-Saclay.
- Clément Maria is a member of the CDT at Inria Sophia Antipolis-Méditerranée.
- Blanche Buet is member of Committee on Gender Equality of LMO at Université Paris-Saclay and member of the Laboratory Council of LMO at Université Paris-Saclay. She has also been member of a recruitment committee recruitment committees for a “Maître de conférence” position at IMJ-PRG, Sorbonne Université and a “PRAG” position at LMO, Université Paris-Saclay, both in 2020.

## 10.2 Teaching - Supervision - Juries

### 10.2.1 Teaching

- Master: Frédéric Chazal and Quentin Mérigot, Analyse Topologique des Données, 30h eq-TD, Université Paris-Sud, France.
- Master: Marc Glisse and Clément Maria, Computational Geometry Learning, 36h eq-TD, M2, MPRI, France.
- Master: Frédéric Cazals and Frédéric Chazal, Geometric Methods for Data Analysis, 30h eq-TD, M1, École Centrale Paris, France.
- Master: Frédéric Chazal and Julien Tierny, Topological Data Analysis, 38h eq-TD, M2, Mathématiques, Vision, Apprentissage (MVA), ENS Paris-Saclay, France.
- Master: Steve Oudot, Topological data analysis, 45h eq-TD, M1, École polytechnique, France.
- Master: Steve Oudot, Data Analysis: geometry and topology in arbitrary dimensions, 24h eq-TD, M2, graduate program in Artificial Intelligence & Advanced Visual Computing, École polytechnique, France.
- Master: Gilles Blanchard, Mathematics for Artificial Intelligence 1, 70h eq-TD, IMO, Université Paris-Saclay, France.
- Master: Blanche Buet, TD-Techniques d'Analyse Harmonique , 30h eq-TD, M2 AAG Orsay, Université Paris-Saclay, France.
- Master: Blanche Buet, TD-Distributions et analyse de Fourier, 60h eq-TD, M1, Université Paris-Saclay, France.



- Undergrad-Master: Steve Oudot, Algorithms for data analysis in C++, 22.5h eq-TD, L3/M1, École polytechnique, France.
- Undergrad: Marc Glisse, Mécanismes de la programmation orientée-objet, 40h eq-TD, L3, École Polytechnique, France.

### 10.2.2 Supervision

- PhD: Siddharth Pritam, Collapses and persistent homology, Jean-Daniel Boissonnat (Université Côte d'Azur). Defended in April 2020.
- PhD: Nicolas Berkouk, Persistence and Sheaves : from Theory to Applications, Institut Polytechnique de Paris. Defended in September 2020. Steve Oudot.
- PhD: Théo Lacombe, Statistics for topological descriptors using optimal transport, Institut Polytechnique de Paris. Defended in September 2020. Steve Oudot.
- PhD: Raphaël Tinarrage, Topological inference from measures and vector bundles. Defended in October 2020. Frédéric Chazal and Marc Glisse.
- PhD: Bertrand Beauflis, Méthodes topologiques et apprentissage statistique pour l'actimétrie du piéton à partir de données de mouvement, Frédéric Chazal and Bertrand Michel (Ecole Centrale de Nantes).
- PhD in progress: Vadim Lebovici, Laplace transform for constructible functions. Started September 1st, 2020. Steve Oudot and François Petit (CRESS).
- PhD in progress: Christophe Vuong, Random hypergraphs. Started November 2020. Laurent Decreasefond and Marc Glisse.
- PhD in progress: Louis Pujol, Partitionnement de données cytométriques, started November 1st, 2019, Pascal Massart and Marc Glisse.
- PhD in progress: Vincent Divol, statistical aspects of TDA, started September 1st, 2017, Frédéric Chazal and Pascal Massart (LMO).
- PhD in progress: Etienne Lasalle, TDA for graph data, started September 1st, 2019, Frédéric Chazal and Pascal Massart (LMO).
- PhD in progress: Alex Delalande, Measure embedding with Optimal Transport and applications in Machine Learning, started December 1st, 2019, Frédéric Chazal and Quentin Mérigot (LMO).
- PhD in progress: Wojciech Riese, Geometric inference for curves and trajectories. Applications to speed estimation from magnetic field measurements, started in September 2020, Frédéric Chazal and Bertrand Michel (Ecole Centrale de Nantes).
- PhD in progress: Jérémie Capitaio-Miniconi, Deconvolution for geometric inference, started October 2020, Frédéric Chazal and Elisabeth Gassiat (LMO).
- PhD in progress: Owen Rouillé, Algorithms and Complexity in Geometric Topology, started September 2018. Clément Maria and Jean-Daniel Boissonnat.
- PhD in progress: Oleksandr Zadorozhnyi, Contributions to the theoretical analysis of the algorithms with adversarial and dependent data, started September 2017. Gilles Blanchard and Alexandra Carpentier.
- PhD in progress: El Mehdi Saad, Efficient online methods for variable and model selection, started September 2019. Gilles Blanchard and Sylvain Arlot.
- PhD in progress: Olympio Hacquard, Dimension reduction for persistent homology, started September 2020. Gilles Blanchard and Clément Levrard.
- PhD in progress: Hannah Marienwald, Transfer learning in high dimension. Started September 2019. Gilles Blanchard and Klaus-Robert Müller.

### 10.2.3 Juries

- Clément Maria was a member of the jury attributing the Gilles Kahn PhD award, from the SIF and the Academy of Science, Nov. 2020.
- Steve Oudot was reviewer for the Ph.D. defence of Håvard Bjerkevik, Norwegian University of Science and Technology, June 2020.
- Steve Oudot was a member of the jury for CRCN applications at Inria Nancy – Grand Est, Spring 2020.
- Blanche Buet was a member of the PhD defence of Camille Labourie, Université Paris Saclay, January 2020 ; François Genereau, Université Grenoble Alpes, June 2020 and Raphaël Tinarrage, October 2020, INRIA-Université Paris Saclay.

## 10.3 Popularization

### 10.3.1 Interventions

- Frédéric Chazal. Les données ont elles une forme? Une petite introduction à l'Analyse Topologique des Données. Back-to-school seminar of the Master in Mathematics at Université Paris-Saclay.

## 11 Scientific production

### 11.1 Major publications

- [1] D. Attali, U. Bauer, O. Devillers, M. Glisse and A. Lieutier. 'Homological Reconstruction and Simplification in R<sup>3</sup>'. In: *Computational Geometry* (2014). DOI: [10.1016/j.comgeo.2014.08.010](https://doi.org/10.1016/j.comgeo.2014.08.010). URL: <https://hal.archives-ouvertes.fr/hal-01132440>.
- [2] J.-D. Boissonnat, R. Dyer and A. Ghosh. 'Delaunay Triangulation of Manifolds'. In: *Foundations of Computational Mathematics* 45 (2017), p. 38. DOI: [10.1007/s10208-017-9344-1](https://doi.org/10.1007/s10208-017-9344-1). URL: <https://hal.inria.fr/hal-01509888>.
- [3] J.-D. Boissonnat, R. Dyer, A. Ghosh and S. Y. Oudot. 'Only distances are required to reconstruct submanifolds'. In: *Computational Geometry* 66 (2017), pp. 32–67. DOI: [10.1016/j.comgeo.2017.08.001](https://doi.org/10.1016/j.comgeo.2017.08.001). URL: <https://hal.inria.fr/hal-01583086>.
- [4] J.-D. Boissonnat, K. C. Srikanta and S. Tavenas. 'Building Efficient and Compact Data Structures for Simplicial Complexes'. In: *Algorithmica* (Sept. 2016). DOI: [10.1007/s00453-016-0207-y](https://doi.org/10.1007/s00453-016-0207-y). URL: <https://hal.inria.fr/hal-01364648>.
- [5] B. Buet, G. P. Leonardi and S. Masnou. 'A Varifold Approach to Surface Approximation'. In: *Archive for Rational Mechanics and Analysis* 226.2 (Nov. 2017), pp. 639–694. DOI: [10.1007/s00205-017-1141-0](https://doi.org/10.1007/s00205-017-1141-0). URL: <https://hal.archives-ouvertes.fr/hal-02141325>.
- [6] F. Chazal, D. Cohen-Steiner and A. Lieutier. 'A Sampling Theory for Compact Sets in Euclidean Space'. In: *Discrete Comput. Geom.* 41.3 (2009), pp. 461–479. URL: <http://dx.doi.org/10.1007/s00454-009-9144-8>.
- [7] F. Chazal, D. Cohen-Steiner and Q. Mérigot. 'Geometric Inference for Measures based on Distance Functions'. Anglais. In: *Foundations of Computational Mathematics* 11.6 (2011). RR-6930, pp. 733–751. DOI: [10.1007/s10208-011-9098-0](https://doi.org/10.1007/s10208-011-9098-0). URL: <http://hal.inria.fr/inria-00383685>.
- [8] F. Chazal, S. Y. Oudot, M. Glisse and V. De Silva. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer Verlag, 2016, pp. VII, 116. URL: <https://hal.inria.fr/hal-01330678>.
- [9] L. J. Guibas, S. Y. Oudot, P. Skraba and F. Chazal. 'Persistence-Based Clustering in Riemannian Manifolds'. Anglais. In: *Journal of the ACM* 60.6 (Nov. 2013), p. 38. URL: <http://hal.archives-ouvertes.fr/hal-00923563>.

- [10] M. Mandad, D. Cohen-Steiner, L. Kobbelt, P. Alliez and M. Desbrun. ‘Variance-Minimizing Transport Plans for Inter-surface Mapping’. In: *ACM Transactions on Graphics* 36 (2017), p. 14. DOI: [10.1145/3072959.3073671](https://doi.org/10.1145/3072959.3073671). URL: <https://hal.inria.fr/hal-01519006>.
- [11] S. Y. Oudot. *Persistence Theory: From Quiver Representations to Data Analysis*. Mathematical Surveys and Monographs 209. American Mathematical Society, 2015, p. 218. URL: <https://hal.inria.fr/hal-01247501>.

## 11.2 Publications of the year

### International journals

- [12] H. Anai, F. Chazal, M. Glisse, Y. Ike, H. Inakoshi, R. Tinarrage and Y. Umeda. ‘DTM-based Filtrations’. In: *Abel Symposia*. Topological Data Analysis 15 (2020), pp. 33–66. DOI: [10.1007/978-3-030-43408-3\\_2](https://doi.org/10.1007/978-3-030-43408-3_2). URL: <https://hal.archives-ouvertes.fr/hal-01919562>.
- [13] G. Blanchard and N. Mücke. ‘Kernel regression, minimax rates and effective dimensionality: Beyond the regular case’. In: *Analysis and Applications* 18.04 (July 2020), pp. 683–696. DOI: [10.1142/S0219530519500258](https://doi.org/10.1142/S0219530519500258). URL: <https://hal.archives-ouvertes.fr/hal-03082307>.
- [14] G. Blanchard, P. Neuvial and E. Roquain. ‘Post hoc confidence bounds on false positives using reference families’. In: *Annals of Statistics* 48.3 (June 2020), pp. 1281–1303. DOI: [10.1214/19-AOS1847](https://doi.org/10.1214/19-AOS1847). URL: <https://hal.archives-ouvertes.fr/hal-01483585>.
- [15] J.-D. Boissonnat, O. Devillers, K. Dutta and M. Glisse. ‘Randomized incremental construction of Delaunay triangulations of nice point sets’. In: *Discrete and Computational Geometry* 64 (2020), p. 33. DOI: [10.1007/s00454-020-00235-7](https://doi.org/10.1007/s00454-020-00235-7). URL: <https://hal.inria.fr/hal-02937624>.
- [16] J.-D. Boissonnat, S. Kachanovich and M. Wintraecken. ‘Triangulating submanifolds: An elementary and quantified version of Whitney’s method’. In: *Discrete and Computational Geometry* (11th Dec. 2020). DOI: [10.1007/s00454-020-00250-8](https://doi.org/10.1007/s00454-020-00250-8). URL: <https://hal.inria.fr/hal-03109814>.
- [17] V. Divol and T. Lacombe. ‘Understanding the Topology and the Geometry of the Space of Persistence Diagrams via Optimal Partial Transport’. In: *Journal of Applied and Computational Topology* (24th Oct. 2020). DOI: [10.1007/s41468-020-00061-z](https://doi.org/10.1007/s41468-020-00061-z). URL: <https://hal.inria.fr/hal-01970466>.
- [18] G. Durand, G. Blanchard, P. Neuvial and E. Roquain. ‘Post hoc false positive control for structured hypotheses’. In: *Scandinavian Journal of Statistics* 47.4 (1st Dec. 2020), pp. 1114–1148. DOI: [10.1111/sjos.12453](https://doi.org/10.1111/sjos.12453). URL: <https://hal.archives-ouvertes.fr/hal-01829037>.
- [19] A. Fischer, C. Levrard and C. BréchetEAU. ‘Robust Bregman Clustering’. In: *Annals of Statistics* (2020). URL: <https://hal.archives-ouvertes.fr/hal-01948051>.
- [20] R. Gribonval, G. Blanchard, N. Keriven and Y. Traonmilin. ‘Compressive Statistical Learning with Random Feature Moments’. In: *Mathematical Statistics and Learning* (2021). URL: <https://hal.inria.fr/hal-01544609>.
- [21] S. Hornus, T. Kuipers, O. Devillers, M. Teillaud, J. Martínez, M. Glisse, S. Lazard and S. Lefebvre. ‘Variable-width contouring for additive manufacturing’. In: *ACM Transactions on Graphics* 39.4 (Proc. SIGGRAPH) (July 2020). DOI: [10.1145/3386569.3392448](https://doi.org/10.1145/3386569.3392448). URL: <https://hal.inria.fr/hal-02568677>.
- [22] C. Jérémy and S. Y. Oudot. ‘Decomposition of exact pfd persistence bimodules’. In: *Discrete and Computational Geometry* (2020). DOI: [10.1007/s00454-019-00165-z](https://doi.org/10.1007/s00454-019-00165-z). URL: <https://hal.inria.fr/hal-01359312>.
- [23] S. Oudot and E. Solomon. ‘Inverse Problems in Topological Persistence: a Survey’. In: *Abel Symposia* (2020). URL: <https://hal.inria.fr/hal-01966676>.
- [24] A. Rastogi, G. Blanchard and P. Mathé. ‘Convergence analysis of Tikhonov regularization for non-linear statistical inverse problems’. In: *Electronic journal of statistics* 14.2 (2020), pp. 2798–2841. DOI: [10.1214/20-EJS1735](https://doi.org/10.1214/20-EJS1735). URL: <https://hal.archives-ouvertes.fr/hal-03082290>.

**International peer-reviewed conferences**

- [25] S. Arya, J.-D. Boissonnat, K. Dutta and M. Lotz. ‘Dimensionality Reduction for k-Distance Applied to Persistent Homology’. In: SoCG 2020 - 36th International Symposium on Computational Geometry. Zurich, Switzerland, 23rd June 2020. DOI: [10.4230/LIPIcs.SoCG.2020.10](https://hal.inria.fr/hal-02873730). URL: <https://hal.inria.fr/hal-02873730>.
- [26] J.-D. Boissonnat and S. Pritam. ‘Edge Collapse and Persistence of Flag Complexes’. In: SoCG 2020 - 36th International Symposium on Computational Geometry. Zurich, Switzerland, 23rd June 2020. DOI: [10.4230/LIPIcs.SoCG.2020.19](https://hal.inria.fr/hal-02873740). URL: <https://hal.inria.fr/hal-02873740>.
- [27] J.-D. Boissonnat and M. Wintraecken. ‘The Topological Correctness of PL-Approximations of Isomanifolds’. In: SoCG 2020 - 36th International Symposium on Computational Geometry. Zurich, Switzerland, 23rd June 2020. DOI: [10.4230/LIPIcs.SoCG.2020.20](https://hal.inria.fr/hal-02873794). URL: <https://hal.inria.fr/hal-02873794>.
- [28] M. B. Botnan, V. Lebovici and S. Oudot. ‘On rectangle-decomposable 2-parameter persistence modules’. In: SoCG 2020 - 36th International Symposium on Computational Geometry. Vol. 164. 36th International Symposium on Computational Geometry (SoCG 2020). Zurich, Switzerland, 8th June 2020, 22:1–22:16. DOI: [10.4230/LIPIcs.SoCG.2020.22](https://hal.archives-ouvertes.fr/hal-02996045). URL: <https://hal.archives-ouvertes.fr/hal-02996045>.
- [29] M. Carriere and A. J. Blumberg. ‘Multiparameter Persistence Images for Topological Machine Learning’. In: NeurIPS 2020 - 34th Conference on Neural Information Processing Systems. Vancouver / Virtuel, Canada, 6th Dec. 2020. URL: <https://hal.inria.fr/hal-03112442>.
- [30] D. Cohen-Steiner, A. Lieutier and J. Vuillamy. ‘Lexicographic optimal homologous chains and applications to point cloud triangulations’. In: SoCG 2020 - 36th International Symposium on Computational Geometry. 36th International Symposium on Computational Geometry (SoCG 2020). Zurich, Switzerland, 23rd June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02391240>.
- [31] M. Dindin, Y. Umeda and F. Chazal. ‘Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks’. In: *33rd Canadian Conference on Artificial Intelligence, May 2020*. CanadianAI 2020 - 33rd Canadian Conference on Artificial Intelligence. Proc. 33rd Canadian Conference on Artificial Intelligence, May 2020. Ottawa, Canada, May 2020. URL: <https://hal.inria.fr/hal-02155849>.
- [32] J. Kim, J. Shin, F. Chazal, A. Rinaldo and L. Wasserman. ‘Homotopy Reconstruction via the Cech Complex and the Vietoris-Rips Complex’. In: *36th International Symposium on Computational Geometry (SoCG 2020)*. SoCG 2020 - 36th International Symposium on Computational Geometry. Vol. 164. LIPIcs, Volume 164, SoCG 2020, Complete Volume. Zurich, Switzerland, 23rd June 2020. URL: <https://hal.inria.fr/hal-02425686>.
- [33] K. Kim, J. Kim, M. Zaheer, J. S. Kim, F. Chazal and L. Wasserman. ‘PLay: Efficient Topological Layer based on Persistence Landscapes’. In: NeurIPS 2020 - 34th Conference on Neural Information Processing Systems. Vancouver / Virtuel, Canada, 6th Dec. 2020. URL: <https://hal.inria.fr/hal-03111277>.
- [34] C. Maria, S. Y. Oudot and E. Solomon. ‘Intrinsic Topological Transforms via the Distance Kernel Embedding’. In: SoCG 2020 - 36th International Symposium on Computational Geometry. Zurich, Switzerland, 2020. URL: <https://hal.inria.fr/hal-02425598>.
- [35] Q. Merigot, A. Delalande and F. Chazal. ‘Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space’. In: AISTATS 2020 - 23rd International Conference on Artificial Intelligence and Statistics. Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics 108. Palermo /Online, Italy, 26th Aug. 2020, pp. 3186–3196. URL: <https://hal.archives-ouvertes.fr/hal-02315998>.

### Conferences without proceedings

- [36] J.-D. Boissonnat, S. Kachanovich and M. Wintraecken. ‘Tracing isomanifolds in  $R^d$  in time polynomial in  $d$  using Coxeter-Freudenthal-Kuhn triangulations’. In: SoCG 2021 - 37th Symposium on Computational Geometry. Buffalo, United States: <https://cse.buffalo.edu/socg21/socg.html>, 7th June 2021. URL: <https://hal.inria.fr/hal-03006663>.
- [37] C. Maria and O. Rouillé. ‘Computation of Large Asymptotics of 3-Manifold Quantum Invariants’. In: ALNEX21. Alexandria, United States, 10th Jan. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03133238>.

### Doctoral dissertations and habilitation theses

- [38] N. Berkouk. ‘Persistence and Sheaves : from Theory to Applications’. Institut Polytechnique de Paris, 24th Sept. 2020. URL: <https://tel.archives-ouvertes.fr/tel-02970509>.
- [39] T. Lacombe. ‘Statistics for Topological Descriptors using optimal transport’. Institut Polytechnique de Paris, 8th Sept. 2020. URL: <https://hal.archives-ouvertes.fr/tel-02979251>.
- [40] S. Pritam. ‘Collapses and persistent homology’. Université Côte d’Azur, 18th June 2020. URL: <https://hal.inria.fr/tel-02962587>.
- [41] R. Tinarrage. ‘Topological inference from measures and vector bundles’. Université Paris-Saclay, 12th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/tel-02970491>.

### Reports & preprints

- [42] G. Blanchard, A. A. Deshmukh, U. Dogan, G. Lee and C. Scott. *Domain Generalization by Marginal Transfer Learning*. 21st Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02974216>.
- [43] G. Blanchard, P. Mathé and N. Mücke. *Lepskii Principle in Supervised Learning*. 21st Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02974206>.
- [44] G. Blanchard, P. Neuvial and E. Roquain. *On agnostic post hoc approaches to false positive control*. 27th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02320543>.
- [45] J.-D. Boissonnat, S. Kachanovich and M. Wintraecken. *A compact data structure for high dimensional Coxeter-Freudenthal-Kuhn triangulations*. 16th Nov. 2020. URL: <https://hal.inria.fr/hal-03006608>.
- [46] J.-D. Boissonnat, S. Kachanovich and M. Wintraecken. *Tracing Isomanifolds of Fixed Dimension in Polynomial Time*. 6th July 2020. URL: <https://hal.inria.fr/hal-02889048>.
- [47] J.-D. Boissonnat and M. Wintraecken. *The topological correctness of PL-approximations of isomanifolds*. 20th Oct. 2020. URL: <https://hal.inria.fr/hal-02386193>.
- [48] M. B. Botnan, V. Lebovici and S. Oudot. *Local characterizations for decomposability of 2-parameter persistence modules*. 9th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02995822>.
- [49] B. Buet and M. Rumpf. *Mean curvature motion of point cloud varifolds*. 23rd Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03019123>.
- [50] M. Carriere, F. Chazal, M. Glisse, Y. Ike and H. Kannan. *Optimizing persistent homology based functions*. 18th Feb. 2021. URL: <https://hal.inria.fr/hal-02969305>.
- [51] F. Chazal, C. Levrard and M. Royer. *Optimal quantization of the mean measure and applications to statistical learning*. 10th Mar. 2021. URL: <https://hal.archives-ouvertes.fr/hal-02465446>.
- [52] O. Cheong, O. Devillers, M. Glisse and J.-w. Park. *Covering families of triangles*. INRIA, 2020, p. 31. URL: <https://hal.inria.fr/hal-03031995>.
- [53] D. Cohen-Steiner and A. C. de Vitis. *Spectral Properties of Radial Kernels and Clustering in High Dimensions*. 6th Jan. 2020. URL: <https://hal.inria.fr/hal-01969956>.
- [54] A. Delalande and Q. Merigot. *Quantitative Stability of Optimal Transport Maps under Variations of the Target Measure*. 9th Mar. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03164147>.

- [55] O. Devillers, P. Duchon, M. Glisse and X. Goaoc. *On Order Types of Random Point Sets*. 28th May 2020. URL: <https://hal.inria.fr/hal-01962093>.
- [56] V. Divol. *A short proof on the rate of convergence of the empirical measure for the Wasserstein distance*. 21st Jan. 2021. URL: <https://hal.inria.fr/hal-03117283>.
- [57] V. Divol. *Minimax adaptive estimation in manifold inference*. 8th June 2020. URL: <https://hal.inria.fr/hal-02440881>.
- [58] V. Divol. *Reconstructing measures on manifolds: an optimal transport approach*. 15th Feb. 2021. URL: <https://hal.inria.fr/hal-03141977>.
- [59] F. Göbel and G. Blanchard. *Volume Doubling Condition and a Local Poincaré Inequality on Unweighted Random Geometric Graphs*. 18th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03082496>.
- [60] C. ICUBAM, L. Bonnasse-Gahot, M. Dénès, G. Dulac-Arnold, S. Girgin, F. Husson, V. Iovene, J. Josse, A. Kimmoun, F. Landes, J.-P. Nadal, R. Primet, F. Quintao, P. G. Raverdy, V. Rouvreau, O. Teboul and R. Yurchak. *ICU Bed Availability Monitoring and analysis in the Grand Est region of France during the COVID-19 epidemic*. 25th May 2020. URL: <https://hal.archives-ouvertes.fr/hal-02620018>.
- [61] C. Maria. *Parameterized complexity of quantum knot invariants*. 6th Jan. 2020. DOI: 10.00477. URL: <https://hal.archives-ouvertes.fr/hal-02429767>.
- [62] H. Marienwald, J.-B. Fermanian and G. Blanchard. *High-Dimensional Multi-Task Averaging and Application to Kernel Mean Embedding*. 12th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03002342>.
- [63] D. Perez. *On  $C^0$ -persistent homology and trees*. 7th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03040819>.
- [64] D. Perez. *On the persistent homology of almost surely  $C^0$  stochastic processes*. 17th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03079171>.
- [65] M. Royer, F. Chazal, C. Levrard, Y. Umeda and Y. Ike. *ATOL: Measure Vectorization for Automatic Topologically-Oriented Learning*. Feb. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02296513>.
- [66] E. M. Saad, G. Blanchard and S. Arlot. *Online Orthogonal Matching Pursuit*. 14th Feb. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03141061>.
- [67] R. Tinarrage. *Computing persistent Stiefel-Whitney classes of line bundles*. 25th May 2020. URL: <https://hal.archives-ouvertes.fr/hal-02619607>.
- [68] R. Tinarrage. *Recovering the homology of immersed manifolds*. 5th June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02396261>.
- [69] O. Zadorozhnyi, G. Blanchard and A. Carpentier. *Restless dependent bandits with fading memory*. 18th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03082512>.