Recent Advances on Topological and Geometric Data Analysis

July 8-10 2009 — workshop program

Program at a glance

Wednesday, July 8th

08:30 - 09:00	Registration
09:00 - 09:20	Opening remarks
09:20 - 10:30	**Key Note:** *Geometric entropy minimization* — Alfred Hero
10:30 - 10:50	Coffee break
10:50 - 12:30	**Session 1: Reconstruction in 3D**
	Scale space meshing — Julie Digne
	Reconstructing 3D compact sets — Frédéric Cazals
12:30 - 14:00	Lunch
14:00 - 15:40	**Session 2: Reconstruction in arbitrary dimensions**
	Manifold Reconstruction from Tangential Complex — Arijit Ghosh
	Model selection for simplicial approximation — Bertrand Michel
15:40 - 16:00	Coffee break
16:00 - 17:40	**Session 3: Geometric inference in the presence of outliers**
	Geometric Inference for Measures based on Distance Functions — Quentin Mérigot
	Efficient Approximation of the Distance to an Empirical Measure — Dmitriy Morozov

Thursday, July 9th

09:00 - 09:20	Registration
09:20 - 10:30	**Key Note:** *An introduction to zigzag persistence* — Vin de Silva
10:30 - 10:50	Coffee break
10:50 - 12:30	**Session 4: Persistence and unsupervised learning**
	Persistence-based clustering — Primoz Skraba
	Persistent cohomology and circular coordinates — Mikael Vejdemo-Johansson
12:30 - 14:00	Lunch
14:00 - 15:40	**Session 5: Signatures for shape classification**
	Topo-geometric Modeling for 3D objects — Hamid Krim
	Gromov-Wasserstein stable signatures for object matching and the role of persistence — Facundo Mémoli
15:40 - 16:00	Coffee break
16:00 - 17:40	**Session 6: Shape matching**
	Heat Kernel Signature: A Concise and Provably Informative Multi-scale Signature Based on Heat Diffusion — Maksims Ovsjanikovs
	Deformable shape matching using linear programming — Qixing Huang

Friday, July 10th

09:00 - 10:40	**Session 7: Reconstruction and mesh generation in 3D**
	Finite Element Analysis of Computer Aided Design Assembly — Kirill Pichon Gostaf
	Reconstruction from Cross-Sections — Pooran Memari
10:40 - 11:00	Coffee break
11:00 - 12:40	**Session 8: Delaunay triangulations**
	Periodic Delaunay triangulations — Manuel Caroli
	A compact data structure to represent the Delaunay Triangulation — Clément Maria
12:40 - 13:00	Closing remarks

Titles and abstracts

Geometric entropy minimization

Alfred Hero, Digiteo Chair and University of Michigan

Geometric entropy minimization (GEM) is a principle that combines combinatorial optimization, information theory, and learning theory. The basis for GEM are the Rényi alpha entropies, which are generalizations of the Shannon entropy. GEM estimates the entropy of a sample by finding graphs that span the data points with minimal properties, e.g. nearest neighbors and minimal spanning trees. We will present theory of GEM and illustrate the method for applications including anomaly detection, intrinsic dimension estimation, and image segmentation.

Scale Space Meshing

Julie Digne, CMLA – École Normale Supérieure de Cachan

Taking the best advantage of high accuracy measurements is a challenge of triangulation laser scanners. This question is particularly relevant for archeological objects where micro textures similar to noise must be preserved. We will show that a scale space approach allows to build a surface texture preserving mesh directly on high precision point data. This contrasts with most mesh reconstruction techniques (level sets, for example) which include a denoising step and therefore entail some smoothing.

Reconstructing 3D compact sets

Frédéric Cazals, ABS (Algorithms Biology Structure) group – INRIA

Reconstructing a 3D shape from sample points is a central problem faced in medical applications, reverse engineering, natural sciences, cultural heritage projects, etc. While these applications motivated intense research on 3D surface reconstruction, the problem of reconstructing more general shapes hardly received any attention. This paper develops a reconstruction algorithm changing the 3D reconstruction paradigm as follows.

First, the algorithm handles general shapes i.e. compact sets as opposed to surfaces. Under mild assumptions on the sampling of the compact set, the reconstruction is proved to be correct in terms of homotopy type. Second, the algorithm does not output a single reconstruction but a nested sequence of *plausible* reconstructions. Third, the algorithm accommodates topological persistence so as to select the most stable features only. Finally, in case of reconstruction failure, it allows the identification of undersampled areas, so as to possibly fix the sampling.

These key features are illustrated by experimental results on challenging datasets, and should prove instrumental in enhancing the processing of such datasets in the aforementioned applications.

Joint work with David Cohen-Steiner.

Manifold Reconstruction from Tangential Complex

Arijit Ghosh, Geometrica group – INRIA

We give a provably correct algorithm to reconstruct a k-dimensional manifold embedded in d-dimensional Euclidean space. Input to our algorithm is a point sample coming from an unknown manifold. Our algorithm and its analysis use ideas from adaptive neighborhood, tangential complexes, and sliver exudation. The running time of our algorithm is quadratic on the size of the input sample of the manifold, linear in d and exponential in k. To the best of our knowledge, this is the first certified algorithm for manifold reconstruction whose complexity depends linearly on the ambient dimension. We also prove that for a dense enough sample the output of our algorithm is homeomorphic to the manifold and a close geometric approximation of the manifold.

Model selection for simplicial approximation

Bertrand Michel, Geometrica group – INRIA

In the computational geometry field, simplicial complexes are currently used to infer a geometric shape knowing a point cloud sampled on it. In this work, an adequate statistical framework is first proposed for the choice of a simplicial complex among a parametrized family. Then, a least squares penalized criterion is introduced to choose a complex in the collection and a model selection theorem is stated to select the *best* simplical complex in a statistical point of view. This result gives the shape of the penalty and next the so called *slope heuristics method* is used to calibrate the penalty from the data. Some experimental studies on simulated and real dataset will illustrate the method for the selection of graphs in two dimensions.

Geometric Inference for Measures based on Distance Functions

Quentin Mérigot, Geometrica group – INRIA

Data often comes in the form of a point cloud sampled from an unknown compact subset of Euclidean space. The general goal of geometric inference is then to recover geometric and topological features (Betti numbers, curvatures,...) of this subset from the approximating point cloud data. In recent years, it appeared that the study of distance functions allows to address many of these questions successfully. However, one of the main limitations of this framework is that it does not cope well with outliers nor with background noise. In this paper, we show how to extend the framework of distance functions to overcome this problem. Replacing compact subsets by measures, we introduce a notion of distance function to a probability distribution in the d-dimensional Euclidean space. These functions share many properties with classical distance functions, which makes them suitable for inference purposes. In particular, by considering appropriate level sets of these distance functions, it is possible to associate in a robust way topological and geometric features to a probability measure. We also discuss connections between our approach and non parametric density estimation as well as mean-shift clustering.

Efficient Approximation of the Distance to an Empirical Measure

Dmitriy Morozov, Geometric Computing group – Stanford University

In order to make geometric and topological inference techniques robust in the presence of significant noise, Chazal, Cohen-Steiner, and Merigot recently introduced the notion of a distance function to a measure. We address the computational questions surrounding this new notion, and in particular analyze two simple algorithms for approximating the distance function to an empirical measure.

Joint work with Leonidas Guibas and Quentin Mérigot.

An introduction to zigzag persistence

Vin de Silva, Digiteo Chair and Pomona College

Zigzag persistence is a new methodology for studying persistence of topological features across a family of spaces or point-cloud data sets. Building on classical results about quiver representations, zigzag persistence generalizes the highly successful theory of persistent homology and addresses several situations which are not covered by that theory. After reviewing standard persistence, I will present theoretical and algorithmic foundations of zigzag persistence with a view towards applications in topological statistics.

Joint work with Gunnar Carlsson and Dmitriy Morozov.

Persistence-based clustering

Primoz Skraba, Geometrica group – INRIA

We present a novel clustering algorithm that combines a mode-seeking phase with a cluster merging phase. While mode detection is performed by a standard graph-based hill-climbing scheme, the novelty of our approach resides in its use of topological persistence theory to guide the merges between clusters. An interesting feature of this algorithm is to provide additional feedback in the form of a finite set of points in the plane, called a *persistence diagram*, which provably reflects the prominence of each of the modes of the density. Such feedback is an invaluable tool in practice, as it enables the user to determine a set of parameter values that will make the algorithm compute a relevant clustering on the next run.

In terms of generality, our approach only requires to know the pairwise distances between the data points, as well as rough estimates of the density at these points. It is therefore virtually applicable in any metric space. In the meantime, its complexity remains reasonable: although the size of the input distance matrix may be up to quadratic in the number of data points, a careful implementation only uses a linear amount of main memory and barely takes more time to run than the one spent reading the input.

Taking advantage of recent advances in topological persistence theory, we give a theoretically sound notion of what the *correct* number k of clusters means, and show that under mild sampling conditions and a relevant choice of parameters (made possible in practice by the persistence diagram) our clustering scheme computes a set of k clusters whose spatial locations are bound to the ones of the basins of attraction of the peaks of the density.

Joint work with Frédéric Chazal, Leonidas Guibas, and Steve Oudot.

Persistent cohomology and circular coordinates

Mikael Vejdemo-Johansson, Applied Topology group – Stanford University

Nonlinear dimensionality reduction (NLDR) algorithms such as Isomap, LLE and Laplacian Eigenmaps address the problem of representing high-dimensional nonlinear data in terms of low-dimensional coordinates which represent the intrinsic structure of the data. This paradigm incorporates the assumption that real-valued coordinates provide a rich enough class of functions to represent the data faithfully and efficiently. On the other hand, there are simple structures which challenge this assumption: the circle, for example, is one-dimensional but its faithful representation requires two real coordinates. In this talk, we present a strategy for constructing circle-valued functions on a statistical data set. We develop a machinery of persistent cohomology to identify candidates for significant circle-structures in the data, and we use harmonic smoothing and integration to obtain the circle-valued coordinate functions themselves. We suggest that this enriched class of coordinate functions permits a precise NLDR analysis of a broader range of realistic data sets.

Joint work with Vin de Silva.

Topo-geometric Modeling for 3D objects

Hamid Krim, Noth Carolina State University

We propose a new method for 3D object representation using weighted skeletal graphs. The geometry of an object is captured by assigning weights to the skeletal graph of the object, which in turn represents its topology. The weights provide necessary information for object reconstruction. The method is rotation, translation, and scaling invariant. Applications include shape representation, compression, and object recognition.

Gromov-Wasserstein stable signatures for object matching and the role of persistence

Facundo Mémoli, Applied Topology group – Stanford University

We review the construction of a variant of the Gromov-Hausdorff distance which is suitable for comparing shapes under invariances. The variant we discuss is itself a metric on the class of measure metric spaces, that is, shapes endowed with a notion of weight at each point. We recall lower bounds for this distance that establish connections with several classical Shape Matching procedures based on assigning certain invariant signatures to a shape. By using concepts from persistence we are able to find a new family of lower bounds for the Gromov-Wasserstein and Gromov-Hausdorff distances. The underlying construction provides a novel class of signatures that can be used for shape matching. In particular, these lower bounds imply the stability of persistence diagrams arising from Vietoris-Rips constructions.

Heat Kernel Signature: A Concise and Provably Informative Multi-scale Signature Based on Heat Diffusion

Maksims Ovsjanikovs, Geometric Computing group – Stanford University

In this talk, I will describe the new robust, multi-scale point signature that is and invariant to isometric deformations. Our point descriptor, called the Heat Kernel Signature, is obtained by restricting the well known heat kernel to the temporal domain. As a result, our signature is both robust to perturbations of the shape, and at the same time captures important information such as curvature in a multi-scale way. This allows us to "compare points" on the same shape, or even across shapes for matching, and finding shared structure. We also show that the set of all Heat Kernel Signatures on the shape defines it up to isometry, which makes it useful for defining robust, isometry invariant shape descriptors.

Joint work with Jian Sun and Leonidas Guibas.

Deformable shape matching using linear programming

Qixing Huang, Geometric Computing group – Stanford University

We present a new technique for deformable shape matching.Our approach uses linear programming to compute correspondences between approximately isometrically deformed surfaces. Our technique is robust to topological noise and can be extended to multi-surface matching. On the technical side, we demonstrate how to formulate robust shape matching as a higher order integer programming problem and how to transform the problem appropriately such that it can be solved efficiently using standard linear programming. We evaluate our technique in a number of applications, such as deformable registration, multi-surface matching and reconstruction, and example-based surface completion.

Joint work with Miachel Wand, Leonidas Guibas and Hans-Peter Seidel.

Finite Element Analysis of Computer Aided Design Assembly

Kirill Pichon Gostaf, Jacques-Louis Lions Laboratory – University Paris VI

In today's product development and engineering process, usage of computer aided design (CAD) platform is obvious. It allows crating of quite realistic models, precisely describing not only the geometry of the developed prototype, but also its physical properties. These models, usually referred as parts, are brought together (assembled) into assemblies, like it occurs in real world. The final digital prototype consists sometimes of thousands of parts and sub-assemblies so, its finite element based optimization is limited by computing architectures that cannot hold and process large models in a timely way. It is therefore desirable to use multiprocessor parallel computers to solve the large mathematical systems. We propose to apply domain decomposition methodology (DDM) to run numerical simulation of CAD large assemblies. Instead of exporting a complex assembly as a single component, and later dividing it in numerous, randomly created sub-domains, we propose to use assembly topology, which was previously created by designer while modeling process period. Thus, each part represents an independent sub-domain, and could be analyzed simultaneously with other parts, by sharing only the contact information. In order to enforce the matching of the local solutions, interface conditions and projection operators have to be written on the boundary between subdomains. These terms and conditions are imposed iteratively. The convergence rate is very sensitive to these interface conditions. The classical Schwarz algorithm without overlap, as well as FETI-based methods are presented. A brief presentation of the finite element open-source FreeFem3D software could be initiated.

Reconstruction from Cross-Sections
Pooran Memari, Geometrica group – INRIA

In this talk, we consider the problem of reconstructing a shape from unorganized cross-sections. The main motivation for this problem comes from medical imaging applications where cross-sections of human organs are obtained by means of a free hand ultrasound apparatus. The position and orientation of the cutting planes may be freely chosen which makes the problem substantially more difficult than in the case of parallel cross-sections, for which a rich literature exists. The input data consist of the cutting planes and (an approximation of) their intersection with the object. Our approach consists of two main steps. First, we compute the arrangement of the cutting planes. Then, in each cell of the arrangement, we reconstruct an approximation of the object from its intersection with the boundary of the cell. Lastly, we glue the various pieces together. The method makes use of the Delaunay triangulation and generalizes the reconstruction method of Boissonnat and Geiger for the case of parallel planes. The analysis provides a neat characterization of the topological properties of the result and, in particular, shows an interesting application of Moebius diagrams to compute the locus of the branching points. We have implemented our algorithm in C++, using the CGAL library. Experimental results show that the algorithm performs well and can handle complicated branching configurations.

Periodic Delaunay triangulations
Manuel Caroli, Geometrica group – INRIA

This work is motivated by the need for software computing 3D periodic triangulations in numerous domains including astronomy, material engineering, biomedical computing, fluid dynamics etc. We design an algorithmic test to check whether a partition of the 3D flat torus into tetrahedra forms a triangulation (which subsumes that it is a simplicial complex). We propose an incremental algorithm that computes the Delaunay triangulation of a set of points in the 3D flat torus without duplicating any point, whenever possible; our algorithmic test detects when such a duplication can be avoided, which is usually possible in practical situations. Even in cases where point duplication is necessary, our algorithm always computes a triangulation that is homeomorpic to the flat torus. To the best of our knowledge, this is the first algorithm of this kind whose output is provably correct. Proved algorithms found in the literature are in fact always computing with 27 copies of the input points in R3, and yield a triangulation that does not have the topology of a torus. Our implementation of the algorithm has been reviewed and accepted by the Cgal Editorial Board.

A compact data structure to represent the Delaunay Triangulation

Clément Maria, Geometrica group – INRIA

One of the difficulties to process the Delaunay triangulation in high dimensions is that the size of the triangulation increases dramatically. In this talk, we present a way to compress the 1-skeleton of the Delaunay triangulation. Our structure must support neighbor listing, addition and deletion of edges in order to work with a previous implementation of the incremental algorithm for constructing Delaunay triangulations.