Data Analysis using Computational Topology and Geometric Statistics
Mar 8 – Mar 13, 2009

MEALS

*Breakfast (Buffet): 7:00–9:30 am, Sally Borden Building, Monday–Friday
*Lunch (Buffet): 11:30 am–1:30 pm, Sally Borden Building, Monday–Friday
*Dinner (Buffet): 5:30–7:30 pm, Sally Borden Building, Sunday–Thursday
Coffee Breaks: As per daily schedule, 2nd floor lounge, Corbett Hall
***Please remember to scan your meal card at the host/hostess station in the dining room for each meal.**

MEETING ROOMS

All lectures will be held in Max Bell 159 (Max Bell Building accessible by walkway on 2nd floor of Corbett Hall). LCD projector, overhead projectors and blackboards are available for presentations. Please note that the meeting space designated for BIRS is the lower level of Max Bell, Rooms 155–159. Please respect that all other space has been contracted to other Banff Centre guests, including any Food and Beverage in those areas.

SCHEDULE

Sunday

16:00	Check-in begins (Front Desk - Professional Development Centre - open 24 hours)
17:30–19:30	Buffet Dinner, Sally Borden Building
20:00	Informal gathering in 2nd floor lounge, Corbett Hall Beverages and small assortment of snacks available on a cash honour-system.

Monday

7:00–8:45	Breakfast
8:45–9:00	Introduction and Welcome to BIRS by BIRS Station Manager, Max Bell 159
9:00–10:00	Stephen Smale, *Hodge Theory*
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Gunnar Carlsson, *Generalized Persistence, Noise, and Statistical Significance*
11:30–13:00	Lunch
13:00–14:00	Guided Tour of The Banff Centre; meet in the 2nd floor lounge, Corbett Hall
14:00	Group Photo; meet on the front steps of Corbett Hall
14:15–14:45	Vin de Silva, *Zigzag persistence*
15:00–15:30	Coffee Break, 2nd floor lounge, Corbett Hall
15:30–16:00	Konstantin Mischaikow, *Topology Guided Sampling of Nonhomogeneous Random Fields*
16:00–16:30	Matt Kahle, *Moduli spaces of hard disks in a box*
16:30–17:00	Sayan Mukherjee, *Conditional Independence Models via Filtrations*
17:30–19:30	Dinner

Tuesday

7:00–9:00	Breakfast
9:00–10:00	Susan Holmes, *How to sample from a manifold: Applications to validation of Computational Topology and its algorithms*
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Peter Bubenik, *Persistent homology and nonparametric regression.*
11:30–13:30	Lunch

13:30–14:00	Stephan Huckeman, *Intrinsic Statistics on Riemannian Manifolds*
14:00–14:30	Yusu Wang, *Approximating Laplace-Beltrami Operator, Integrals and Gradients in Non-statistical Discrete Settings*
14:30–15:00	Moo Chung, *Eigenfunctions of Laplace-Beltrami operator in cortical manifolds*
15:00–15:30	Coffee Break, 2nd floor lounge, Corbett Hall
15:30–16:00	Axel Munk, *Statistical Multiscale Analysis - from signal detection to Image Analysis and Biophotonics*
16:00–16:30	Zhiming Luo, *Asymptotic minimax regression estimate under super-norm loss on Riemannian manifold*
17:30–19:30	Dinner

Wednesday

7:00–9:00	Breakfast
9:00–9:30	Leo Guibas, *Analysis of Scalar Fields over Point Cloud Data*
9:30–10:00	Tamal Dey, *Topology by approximating cut locus from point data*
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:00	Andre Lieutier, *A stable notion of curvature on point clouds*
11:00–11:30	Fred Chazal and David Cohen-Steiner, *Geometric inference for probability distributions*
11:30–13:30	Lunch
13:30–17:30	Free Afternoon
17:30–19:30	Dinner

Thursday

7:00–9:00	Breakfast
9:00–10:00	Vic Patrangenaru, *Asymptotic Statistics and Nonparametric Bootstrap on Manifolds and Applications*
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Louis-Paul Rivest, *Some statistical models for SE(3) data*
11:30–13:30	Lunch
13:30–14:00	Michael Pierrynowski, *Differential geometry reveals differences in the knee motion of elders with osteoarthritis*
14:00–14:30	Dominique Attali, *Persistence-sensitive simplification of functions on surfaces in linear time.*
14:30–15:00	Coffee Break, 2nd floor lounge, Corbett Hall
15:00–15:30	Facundo Memoli, *A Metric Geometry approach to Object Matching*
15:30–16:00	Yuriy Mileyko, *Defining hierarchical order within reticular networks*
17:30–19:30	Dinner

Friday

7:00–9:00	Breakfast
9:00–10:00	Informal Discussions
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Informal Discussions
11:30–13:30	Lunch

Checkout by 12 noon.

** 5-day workshops are welcome to use the BIRS facilities (2nd Floor Lounge, Max Bell Meeting Rooms, Reading Room) until 3 pm on Friday, although participants are still required to checkout of the guest rooms by 12 noon. **

Data Analysis using Computational Topology and Geometric Statistics
Mar 8 – Mar 13, 2009

ABSTRACTS
(in alphabetic order by speaker surname)

Speaker: **Dominique Attali** (CNRS, Grenoble)
Title: *Persistence-sensitive simplification of functions on surfaces in linear time.*
Abstract: Let f be a real-valued function defined on a triangulated surface S. The persistence diagram of f encodes the homological variations in the sequence of sublevel sets $S_t = f^{-1}(-\infty, t]$. A point (x, y) in the persistence diagram of f corresponds to a homological class which appears in S_x and disappears in S_y. The distance $y - x$ of the point (x, y) to the diagonal represents the importance of the associated homological class: the further away a point is from the diagonal, the more important the associated feature. An ε-simplification of f is a map g on S whose persistence diagram consists only of those points in the diagram of f that are more than ε away from the diagonal. In this talk, we give an algorithm for constructing an ε-simplification of f which is also ε-close to f.

This is a joint work with M. Glisse, S. Hornus, F. Lazarus and D. Morozov.

Speaker: **Peter Bubenik** (Cleveland State University)
Title: *Persistent homology and nonparametric regression.*
Abstract: We estimate the persistent homology of sublevel sets of a function on a compact Riemannian manifold, from a finite noisy sample. The Stability Theorem of Cohen-Steiner, Edelsbrunner and Harer bounds the distance between the persistent homologies of the sublevel sets of two functions by the supremum norm of the difference between the two functions. This allows us to convert our topological problem to the statistical nonparametric regression problem on a compact manifold under the sup-norm loss. We calculate the sharp asymptotic minimax bound. Furthermore, the construction of the estimator in the proof is well-suited to calculations of the persistent homology of its sublevel sets. We illustrate these techniques with an application to brain image data. This is joint work with Gunnar Carlsson, Moo Chung, Peter Kim, and Zhiming Luo.

Speaker: **Gunnar Carlsson** (Stanford University)
Title: *Generalized Persistence, Noise, and Statistical Significance*
Abstract: Persistent homology has been shown to be a useful way to detect qualitative structure in various kinds of data sets. Recently, in joint work with V. de Silva and D. Morozov, we have shown that a generalized form of persistence, which we call "zig-zag persistence", can be useful both in removing noise in certain geometric problems as well as in understanding statistical significance of qualitative geometric invariants. I will describe the techniques and the various ways in which it can be applied.

Speaker: **Fred Chazal and David Cohen-Steiner** (INRIA)
Title: *Geometric inference for probability distributions*
Abstract: 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 talk, we will show how to extend the framework of distance functions to overcome this problem. Replacing compact subsets by measures, we will introduce a notion of distance function to a probability distribution in $\mathbb{R}^n$. 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. If time permits, we will also mention a few other potential applications of this framework.

Speaker: **Moo Chung** (University of Wisconsin-Madison)
Title: *Eigenfunctions of Laplace-Beltrami operator in cortical manifolds*
Abstract: In quantifying cortical and subcortical anatomy of the human brain, various differential geometric methods have been proposed. Many such successful methods are inherently implicit and without explicit parametric forms. Although there are few parametric approaches such as spherical harmonic descriptors, the application has been limited to simple subcortical structures. The reason for the lack of more explicit parametric approaches is that it is difficult to construct an orthonormal basis for an arbitrary cortical manifold. We propose to use the eigenfunctions of the Laplace-Beltrami operator, which are computed numerically using the cotan fomula. The eigenfunctions are then used in setting up a regression in the cortical manifold. In the heat kernel smoothing framework, smoothing is done by expanding the heat kernel using the eigenfunctions. The eigenfunction approach offers far more flexibility in setting up a statistical model than implicit approaches.

Speaker: **Vin de Silva** (Pomona)
Title: *Zigzag persistence*
Abstract: 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 generalises the highly successful theory of persistent homology and addresses several situations which are not covered by that theory. I will present theoretical and algorithmic foundations with a view towards applications in topological statistics. As an important example, I discuss a particular zigzag sequence derived from the levelsets of a real-valued function on a topological space. A powerful structure theorem, called the Pyramid Theorem, establishes a connection between this "levelset zigzag persistence" and the extended persistence of Cohen-Steiner, Edelsbrunner and Harer. This theorem resolves an open question concerning the symmetry of extended persistence. Moreover, the interval persistence of Dey and Wenger can be understood in this context; in some sense it carries three-quarters of the information produced by the other two theories. This is joint work with Gunnar Carlsson and Dmitriy Morozov.

Speaker: **Tamal Dey** (Ohio State University)
Title: *Topology by approximating cut locus from point data*
Abstract: A cut locus of a point p in a compact Riemannian manifold M is defined as the set of points where *minimizing* geodesics issued from p stop being minimizing. It is known that a cut locus contains most of the topological information of M. Our goal is to utilize this property of the cut loci to decipher the topology of M from a point sample. Recently it has been shown that Rips complexes can be built from a point sample P of M systematically to compute the Betti numbers, the rank of the homology groups of M. Rips complexes can be computed easily. However, the sizes of the Rips complexes tend to be large. Since the dimension of a cut locus is lower than that of the manifold M, a sub-sample of P approximating the cut locus is usually much smaller in size and hence admits a relatively smaller Rips complex.

We explore the above approach for point data sampled from surfaces embedded in any high dimensional Euclidean space. We present an algorithm that computes a sub-sample P' of a sample P of a 2-manifold where P' approximatesa cut locus. Empirical results show that the first Betti number of M can be computed from the Rips complexes built on these sub-samples.

Speaker: **Leo Guibas** (Stanford)
Title: *Analysis of Scalar Fields over Point Cloud Data*
Abstract: Given a real-valued function f defined over some metric space X, is it possible to recover some structural information about f from the sole information of its values at a finite subset L of sample points, whose pairwise distances in X are given? We provide a positive answer to this question. More precisely,

taking advantage of recent advances on the front of stability for persistence diagrams, we introduce a novel algebraic construction, based on a pair of nested families of simplicial complexes built on top of the point cloud L, from which the persistence diagram of f can be faithfully approximated. We derive from this construction a series of algorithms for the analysis of scalar fields from point cloud data. These algorithms are simple and easy to implement, have reasonable complexities, and come with theoretical guarantees.

Joint work with F. Chazal, S. Y. Oudot, and P. Skraba.

Speaker: **Susan Holmes** (Stanford)
Title: *How to sample from a manifold: Applications to validation of Computational Topology and its algorithms*
Abstract: First I will survey the classical methods of parametric bootstrapping and MCMC for generating samples from non uniform distributions. Then I will present joint work with Persi Diaconis and Mehrdad Shahshahani on how to draw samples from a manifold and show this can be used to compute confidence statements for results from various output from computational topology algorithms such as JPlex.

Speaker: **Stephan Huckeman** (Goettingen)
Title: *Intrinsic Statistics on Riemannian Manifolds*
Abstract: One goal in image analysis consists in describing statistical distributions of characteristic patterns, e.g. shapes of random physical objects. Typically such shapes live on non-Euclidean manifolds, possibly with unbound curvature at singularities (e.g. Kendall's 3D shape space). While over the last decades statisticians have used Euclidean approximations to these manifolds thus making tools of classical multivariate analysis available for "sufficiently concentrated data", this contribution aims at intrinsic generalizations of PCA and MANOVA thus broadening the scope of statistical image analysis.

Speaker: **Matt Kahle** (Stanford University)
Title: *Moduli spaces of hard disks in a box*
Abstract: We discuss a family of moduli spaces which generalize classical configuration spaces for points in the plane. However, the methods used for computing homology of configuration spaces are not easily applicable to these spaces, and even the number of components seems to be a fairly subtle question. So we resort to computational / applied methods in trying to better understand a pure math problem. We apply a combination of techniques, including simulated annealing and the nudged elastic band method, to compute the most basic topological features of these spaces. In this talk we will also briefly discuss the statistical physics setting that motivates the problem, suggested by Persi Diaconis. This is ongoing joint work with Gunnar Carlsson and Jackson Gorham.

Speaker: **Andre Lieutier** (Dassault Systemes)
Title: *A stable notion of curvature on point clouds*
Abstract: Joined work with Frederic Chazal, David Cohen-Steiner and Boris Thibert

We address the problem of curvature estimation from sampled compact sets.

The main contribution is a stability result: we show that the gaussian, mean or anisotropic curvature measures of the offset of a compact set K with positive μ-reach can be estimated by the same curvature measures of the offset of a compact set K' close to K in the Hausdorff sense.

We show how these curvature measures can be computed for finite unions of balls.

The curvature measures of the offset of a compact set with positive μ-reach can thus be approximated by the curvature measures of the offset of a point-cloud sample. These results can also be interpreted as a framework for an effective and robust notion of curvature.

Speaker: **Zhiming Luo** (University of Guelph)
Title: *Asymptotic minimax regression estimate under super-norm loss on Riemannian manifold*
Abstract: Relating to Peter Bubenik's talk "Persistent homology and nonparametric regression", we give more details on the minimax nonparametric regression estimator and the exact constant of the sharp

asymptotic minimax bound on a compact Riemannian manifold. This is joint work with Peter Bubenik, Gunnar Carlsson, Moo Chung, and Peter Kim.

Speaker: **Facundo Memoli** (Stanford)
Title: *A Metric Geometry approach to Object Matching*
Abstract: The problem of object matching under invariances can be studied using certain tools from Metric Geometry. The main idea is to regard objects as metric spaces (or measure metric spaces). The type of invariance one wishes to have in the matching is encoded in the choice of the metrics with which we endow the objects. The standard example is matching objects in Euclidean space under rigid isometries: in this situation one would endow the objects with the Euclidean metric. More general scenarios are possible in which the desired invariance cannot be reflected by the preservation of an ambient space metric. Several ideas due to M. Gromov are useful for approaching this problem. In this talk we discuss different adaptations of these, and in particular we construct an L^p version of the Gromov-Hausdorff distance using mass transportation ideas.

Speaker: **Yuriy Mileyko** (Duke University)
Title: *Defining hierarchical order within reticular networks*
Abstract: While the Strahler Stream Order is a standard method for computing the hierarchal order within non-reticular networks, it cannot handle networks with loops. In this talk I shall present a new algorithm which can perform such a task for planar networks. This algorithm is based on ideas from persistent homology and may be regarded as a generalization of the Strahler Stream Order. From a topological point of view, the latter method defines a filtration of a network (based on tributaries) and updates the order of the edges at critical events, that is, when two connected components merge. Such an event can be regarded as a change in 0-dimensional homology. Therefore, we define critical events for networks with loops as changes in 1-dimensional homology. Taking advantage of the planarity of a network, we can trace a sequence of such critical events and update the order of network edges. This work was motivated by the problem of analyzing the structure of leaf networks, and I shall preset a few preliminary result of such an analysis. I shall also discuss possible generalizations of the new method to arbitrary networks.

Speaker: **Konstantin Mischaikow** (Rutgers University)
Title: *Topology Guided Sampling of Nonhomogeneous Random Fields*
Abstract: Topological measurements are increasingly being accepted as an important tool for quantifying complex structures. In many applications these structures can be expressed as nodal domains of real-valued functions and are obtained only through experimental observation or numerical simulations. In both cases, the data on which the topological measurements are based are derived via some form of finite sampling or discretization. In this paper we present a probabilistic approach to quantifying the number of components of generalized nodal domains of non-homogeneous random fields in one space dimension via fi nite discretizations, i.e., we consider excursion sets of a random eld relative to a non-constant deterministic threshold function. Our results furnish explicit probabilistic a-priori bounds for the suitability of certain discretization sizes and also provide information for the choice of location of the sampling points in order to minimize the error probability. We illustrate our results for a variety of random fields, demonstrate how they can be used to sample the classical nodal domains of de- terministic functions perturbed by additive noise, and discuss their relation to the density of zeros.

Speaker: **Sayan Mukherjee** (Duke University)
Title: *Conditional Independence Models via Filtrations*
Abstract: We formulate a novel approach to infer conditional independence models or Markov structure of a multivariate distribution. Specifically, our objective is to place informative prior distributions over decomposable graphs and sample efficiently from the induced posterior distribution. The key idea we develop in this paper is a parametrization of decomposable hypergraphs using the geometry of points in $\mathbb{R}^m$. This allows for specification of informative priors on decomposable graphs by priors on a finite set

of points. The constructions we use have been well studied in the fields of computational topology and random geometric graphs. We develop the framework underlying this idea and illustrate its efficacy using simulations.

Speaker: **Axel Munk** (Goettingen)
Title: *Statistical Multiscale Analysis - from signal detection to Image Analysis and Biophotonics*
Abstract: In this talk we discuss how to use statistical multiscale analysis (SMA) techniques in order extract jumps from noisy signals in various signal detection problems. This will be applied to reconstruct the open states in ion channel experiments for biomembranes. In the second part SMA will be extended to Image analysis, i.e. to 2D and 3D. The resulting method is locally adaptive, i.e. it automatically adjusts locally any regularisation method to locally varying features, such as edges. This will be illustrated with examples from biophotonic imaging.

Speaker: **Vic Patrangenaru** (Florida State University)
Title: *Asymptotic Statistics and Nonparametric Bootstrap on Manifolds and Applications*
Abstract: Asymptotic statistical analysis and nonparametric bootstrap on smooth geometric ob- jects, or manifolds, is an exciting and challenging field of research, extending multivariate limit theorems to the nonlinear case, where statistical theory and differential geometry are inextricably intertwined, and implementation requires innovative algorithms and high speed computation. The presentation deals with recent developments in this young area of nonparametric statistics, which must also resolve associated geometric issues and problems of implementation. Asymptotic statistics on manifolds have a wide range of applications in many areas of science including geology, meteorology, biology, medical imaging, bioinformatics and machine vision. This is joint work with R. N. Bhattacharya, F. H. Ruymgaart and other collaboratorts.

Speaker: **Michael Pierrynowski** (McMaster University)
Title: *Differential geometry reveals differences in the knee motion of elders with osteoarthritis*
Abstract: Knee motion, force and moment have been used by biomechanists to identify elders with and without knee osteoarthritis (OA). The knee adduction moment has received the most attention since it is associated with the severity and prognosis of OA which then informs clinicians to prescribe effective intervention. However, measuring the knee adduction moment clinically is problematic since it requires synchronized kinematic data acquisition and ground reaction force measures. For potential clinical use we propose a differential geometry analysis of the easier measured knee kinematics [SE(3)] that shows promise to detect the presence or absence of mild to moderate knee OA. This technique sums over repetitive gait cycles the curvatures () and torsions () from the 3 translation component of SE(3) which are then geometrically interpreted using Frechels Theorem. In a similar vein, we examined the length of the paths transcribed on a sphere (S^2) by the three columns (orthonormal vectors) of the SO(3) orientation component. We report that during repetitive normal overground gait, the sum of the curvatures and the path length of the third SO(3) vector are smaller in 52 elders with knee osteoarthritis compared to 47 elders with healthy knees. We will discuss this finding in relation to OA knees having decreased non-linear motion paths and less tibia rotation during gait. We acknowledge the support of Queens University and Sole Supports, Inc. This is joint work with Peter T. Kim.

Speaker: **Louis-Paul Rivest** (Université Laval)
Title: *Some statistical models for SE(3) data*
Abstract: This presentation will begin by reviewing the occurrence, in the biomechanical literature, of data sets whose elements belong to SE(3), the 6-dimensional Lie group of 3D rigid body displacements. The construction of some probability models on SE(3) using distance measures will be presented. These models will be used to describe the dispersion of an observed SE(3) displacement around its true value. They will be used to construct loss functions for the estimation of the parameters of a statistical model for SE(3) data. The SE(3) model used to estimate the directions of the two rotation axes of the ankle will then be

presented. Some statistical challenges associated with the estimation of the parameters of this model will be reviewed, with some of the solutions that have been put forward. Statistical analyses carried out with the R-package Kinematics for the statistical modeling of SE(3) data will be used to illustrate the theory.

Speaker: **Stephen Smale** (Toyota Technological Institute at Chicago)
Title: *Hodge Theory* Abstrac:We will discuss results on extensions of Hodge theory to metric spaces and the relations to the subject "Topology, Geometry and Data"

Joint work with Nat Smale

Speaker: **Yusu Wang** (Ohio State University)
Title: *Approximating Laplace-Beltrami Operator, Integrals and Gradients in Non-statistical Discrete Settings*
Abstract: The Laplace-Beltrami operator of a given manifold (e.g, a surface) is a fundamental object encoding the intrinsic geometry of the underlying manifold. It has many properties useful for practical applications from areas such as graphics and machine learning. For example, its relation to the heat diffusion makes it a primary tool for surface smoothing in graphics. However, many a time, the underlying manifold is only accessible through a discrete approximation, either as a mesh or simply as a set of points. The important question is then how to approximate the Laplace operator and other geometric invariants from such discrete setting. Previously, much work has been done on approximating Laplace operator from points sampled from some probabilistic distribution. In this talk, I will briefly describe our recent results on approximating Laplace operator from either piecewise linear manifolds (e.g, meshes) or simply general point cloud data. I will then focus on several applications of the constructed discrete Laplace operator, including estimating gradient, critical points, and integral of an input function from point cloud data.