Atelier Fields-Ottawa
Fields-Ottawa
Workshop

à Ottawa, février 23 -- 25, 2005
in Ottawa, February 23--25, 2005

uOttawa

Faculté des sciences
Faculty of Science

La Géométrie des Très Grands Échantillons de Données

The Geometry of Very Large Data Sets

Conférenciers invites/ Invited Speakers

Gunnar Carlsson (Stanford University)
Persistent Homology

Alexander Gorban (University of Leicester)
How to discover a geometry and topology in a finite dataset by means of elastic nets.

Peter Kim (Guelph University)
Nonparametrics in High Dimensions

Information
http://www.fields.utoronto.ca/programs/scientific/04-05/data_sets/
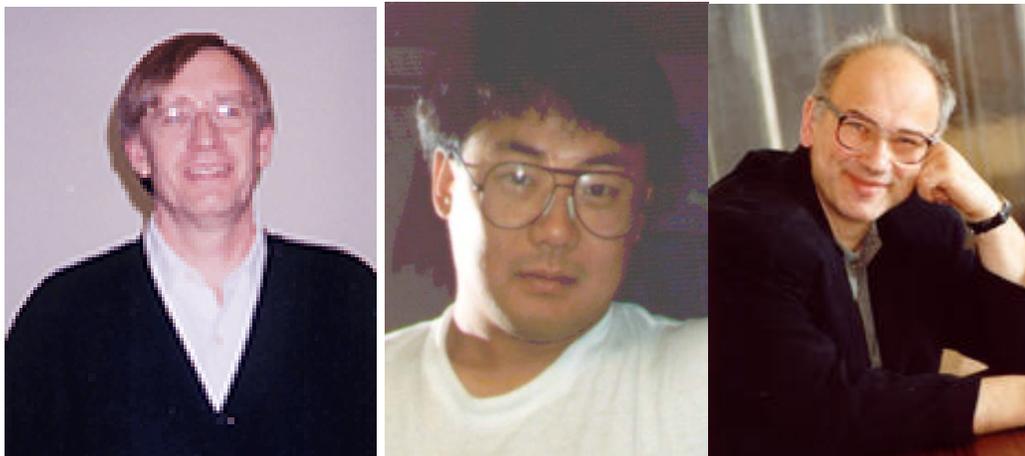http://www.mathstat.uottawa.ca/conf/fields_win_05/main.html

Supported by MITACS

Tentative schedule

Date	Heure/Time	Evenement/Event
February 23	9:00AM	Bienvenue/Welcome coffee
	10:00AM	Elements of topology (P-E Parent and B. Jessup)
	Midi/noon	Lunch
	13:30	Aspects of statistics (A. Dabrowski)
	15:00	Objectives
	17:00	
February 24	9:00AM	Gunnar Carlsson I
	10:30	Coffee
	11:00	Alexander Gorban I
	12:30	Lunch
	14:00	Peter Kim I
	15:30	Gunnar Carlsson II
	17:00	Social event
February 25	8:30AM	Peter Kim II
	10:00	Coffee
	10:30	Maia Lesosky
	11:00	Peter Bubenik
	11:30	Ulrich Fahrenberg
	12:00	Pizza
	13:00	Alexander Gorban II
	14:30	Discussion

All Sessions will be held in STE J0106 (SITE building at King Edward and Mann,
http://www.uottawa.ca/map/index.html)

Abstracts

Paul-Eugène Parent and Barry Jessup(Ottawa)	**Elements of topology.** Assuming only an undergraduate knowledge of mathematics, we will review some basic notions in manifolds and algebraic topology. We will also introduce the concept of an (algebraic) invariant for such objects, one in particular being the homology. We will recall the construction of a combinatorial tool to compute this invariant, namely a simplicial complex.
André Dabrowski (Ottawa)	**Aspects of statistics.** This session will employ the basic elements from introductory undergraduate courses on probability theory and statistics as a springboard to the discussion of more comprehensive results such as Donsker's theorem (Functional Central Limit Theorem) and empirical processes.
Gunnar Carlsson (Stanford)	**Persistent Homology.** Algebraic topology is a mathematical formalism which makes precise mathematics out of certain kinds of intuitive concepts concerning geometrical objects. These concepts come under the heading of "connectivity information", i.e. they include the possibility of decomposing the object into disjoint pieces, the study of holes in the space, the nature of closed loops, and so on. Until recently, these methods have been restricted to situations where the space is given in closed form, and where by hand calculation is feasible. In recent years, methods have been developed which permit the automatic computation of some of this information in situations where we are not given complete information about the space, but only sets sampled from the space. These ideas can be used to study high dimensional data sets qualitatively in situations where actual visualization is not possible. We will present this work, and show examples using some real data sets. We will also show how the ideas can be extended to study qualitative information which is not a priori topological, such as the presence of corners and edges, and apply the results to shape recognition.
Alexander Gorban (Leicester)	**How to discover a geometry and topology in a finite dataset by means of elastic nets.** Principal manifolds were introduced in 1989 as lines or surfaces passing through "the middle" of the data distribution. This intuitive notion, corresponding to the human brain generalization ability, was supported by a mathematical notion of self-consistency: every point of the principal manifold is a conditional mean of all points that are projected into this point. Most scientific and industrial applications of principal manifold methodology were implemented using the SOM (self-organizing maps) approach, coming from the theory of neural networks. In the lecture, algorithms for fast construction of approximate principal manifolds with various topology are presented. These algorithms are based on analogy of principal manifold and elastic membrane and corresponding variational principle. The relation between the classical statistics and the data modelling approaches is discussed. In Introduction, brief review of clustering algorithms and SOM construction is presented. Further steps, principal graphs construction and a graph grammar extraction are outlined
Peter Kim (Guelph)	**Nonparametrics in high dimensions.** This series of talks will investigate the interplay between geometry and statistics. We will begin with a look at a parametric problem on the sphere with the data being the directed unit normals of the elliptic planes of long period cometary orbits. It is the belief by astronomers that the intrinsic distribution of

	the directed normals is the spherical uniform distribution. Nevertheless conventional statistical tests always reject uniformity if applied directly. Part of the difficulty comes from the fact that there is considerable selection bias in the observed directed normals. One can model this selection bias by using properties of spherical geometry and once this selection bias has been accounted for, one can no longer reject spherical uniformity of the directed unit normals of long period cometary orbits. The second topic discusses a deconvolution problem on the space of 3x3 rotation matrices. The technique involves using the irreducible representations of rotation matrices. The main result is to show that one can obtain minimax deconvolution density estimators on the space of 3x3 rotation matrices. This represents a sufficiently rich example so that one can extend the theory to compact Lie groups. Some applications are discussed which include rotational matching in bioinformatics, texture analysis in physical chemistry, encryption in quantum computing as well as an application to persistent homology. The third topic will then be a general approach to what may be termed as a statistical inverse problem on a Riemannian manifold. Here one is interested in recovering a transformation of a density function on a Riemannian manifold. Both rate and sharp minimaxity will be discussed along with additional examples.
Maia Lesosky (Guelph)	**Introduction to Quantum Computing.** Quantum computing has been generating intense interest lately in a large number of fields. I will introduce the concept of quantum error correction, particularly a method known as the noiseless subsystems method. Along the way I will discuss some interesting extensions of classical probability theory. In addition we will see a nice geometrical interpretation of one of the key players in quantum computing, the density matrix.
Peter Bubenik (Lausanne)	**Persistent homology and directional statistics.** We combine statistical and topological approaches to study data sampled from densities on spheres. In particular we define a persistent homology for densities, and calculate barcode and function descriptors for the homology of various densities on spheres. We use the theory of spacings to compare different ways of combining statistics and topology to study very large data sets sampled from the circle.
Ulrich Fahrenberg (Aalborg)	**Parallel composition of automata.** We show how parallel composition of higher-dimensional automata (HDA) can be expressed categorically in the spirit of Winskel & Nielsen. Employing the notion of computation path introduced by van Glabbeek, we define a new notion of bisimulation of HDA using open maps. We derive a connection between computation paths and carrier sequences of dipaths and show that bisimilarity of HDA can be decided by the use of geometric techniques. For a mathematical audience, we concentrate more on the topological aspects, and include material on how equivalence of computation paths is related to dihomotopy of dipaths