

## ÉCOLE D'ÉTÉ FRANCE EXCELLENCE

### TOULOUSE MASTER SCHOOL ON DATA SCIENCE AND GEOMETRY

Arrival at Toulouse airport via Paris on July 1st 2019, first lectures start on Tuesday 2nd July 2019.

**INSTITUTE OF MATHEMATICS PRESENTATION :** The Toulouse Mathematics Institute, CNRS Research Laboratory, federates the mathematics community of the Toulouse area. It is comprised of three main teams, the team of Statistics and Probability, the Mathematics for Industry and Physics Team and the Fundamental Mathematics Team, located at the University of Toulouse III, the Universities of Toulouse I and II, and INSA (Institut National des Sciences Appliquées).

It brings together around 240 tenured researchers and lecturer-researchers whose research interests cover a very broad spectrum of mathematics, ranging from theoretical aspects to applications. It is host to long- and short-term foreign visitors from all over the world, and all mathematical backgrounds. It cultivates international partnerships, as well as industrial contracts, these being particularly at a local level.

Jointly accredited with INSA, SUPAERO and the University of Toulouse I, and associated with the Graduate School of Informatics at the level of the University of Toulouse, its Graduate School has an annual intake of around 120 graduate students.

The Institute Library hosts the region's largest collection of mathematical books and journals, and constitutes a major work tool.

Under the aegis of the University Paul-Sabatier (University of Toulouse III), The Mathematics Institute publishes the "Annales de la Faculté des Sciences de Toulouse (Mathématiques)", one of France's oldest scientific journals, and awards the internationally-renowned "Fermat Prize for mathematical research" every two years

**SCHOOL PRESENTATION :** The aim of this summer school is to provide an insightful overview of two very actual mathematical topics: Geometry and Data Science. These two themes will first be developed separately and then will merge through the study of highly efficient recent algorithm developed to handle complex data.

Toulouse Institute of Mathematics has a high world visibility for its researches in Geometry and Stochastic Analysis. The participant of this summer school will benefit of this high level of competence in the city considered as the jewel of the south west of France.

The school will be organized as follows. the three first weeks, the students will follow both 10 hours in Geometry (2h daily) and 10 hours of Statistics (2h daily). While the last week will be devoted to implementation in Python of Data Science algorithms built using geometric tools.

The summer school is a satellite event associate to the thematic CIMI semester Statistics with Geometry and Topology (August-November 2019).

## 1. FACILITIES

## ARRIVAL IN FRANCE

Participants will be welcomed at Toulouse Blagnac international airport on Monday 1st July 2019.

Courses will start on Tuesday 2nd July 2019.

## ACCOMMODATION

The school will provide accommodation for all the participants in one of the CROUS student's residence on the Université Paul Sabatier Campus.

## MEALS

Noon and afternoon meals are included in the program and will take place in the university restaurants.

## TRANSPORTATION

The residence is located some meters away from the mathematical institute where the lectures are going to be held. However the school will provide a one month public transport pass that gives unlimited travel in buses, metro and tramway in Toulouse metropolitan area.

## RETURN TO CHINA

The school will end on Friday 26th July 2019. The return flight to China from Toulouse Blagnac international airport will take place on Saturday 27th July 2019.

## 2. SCIENTIFIC SYLLABUS AND REQUIRED LEVEL

### EXPECTED LEVEL OF PARTICIPANTS

The expected level of the participants may vary from the end of the third undergraduate year of mathematics to the end of the first master year. That is, we expect the students have followed and completed 3 to 5 years of mathematics.

### LECTURES, SCHOOL DESCRIPTION

The school will count with 6 ten hours courses and 1 twenty hours course on Geometry and Statistics. The first three weeks, the students will follow one course on Geometry and one on Statistics. The last week a 20 hours course, including student projects, is scheduled. This last course will explore some interesting interactions between geometry and data sciences. For more details we now list the lectures titles and contents by week :

#### FIRST WEEK 2-5 july 2019

Geometry Lecture (10h): «An Introduction to Riemannian Geometry»

Speaker: Paulo CARRILLO ROUSE

Description:

1. Lecture 1 (2 hours) Differentiable manifolds, tangent spaces, vector fields.
2. Lecture 2 (2 hours) Riemannian metrics and Riemannian connections.
3. Lecture 3 (2 hours) Geodesics; convex neighborhoods.
4. Lecture 4 (2 hours) Curvature. Ricci and scalar curvature. Tensors on Riemannian manifolds.
5. Lecture 5 (2 hours) Jacobi fields. Morse Index theorem.

Statistics Lecture (10h): «Probability background for statistical learning»

Speakers: Clément Pellegrini and Thierry Klein

Description:

1. Lecture 1 and 2 (4 hours) Convergence of random variables, SLLN, CLT Delta method and Slutsky lemma, Gaussian vectors, Classical concentration inequalities.
2. Lecture 3 (2 hours) Conditional expectation.
3. Lecture 4 (2 hours) Parameter estimation in statistics. Moments methods and maximum likelihood estimation. Confidence sets.
4. Lecture 5 (2 or 4 hours) Basic methods in statistical learning. PCA, Regression, k-nearest neighbours algorithm, theoretical study of the rate of convergence of the k-nearest neighbours algorithm.

SECOND WEEK 8-12 july 2019

Geometry Lecture (10h): «Analysis on manifolds »

Speaker: Jerome BERTRAND

Description:

1. Lecture 1 (2 hours) Manifolds with boundary, Bishop inequality, Laplacian of the distance function.
2. Lecture 2 (2 hours) Differential operators and their formal adjoints, the Hodge-de Rham theorem. Basic spectral geometry.
3. Lecture 3 (2 hours) Some examples of Spectra, The minimax principle.
4. Lecture 4 (2 hours) Eigenvalues estimates, Bishop's theorem, Lower bounds for the first eigenvalue.
5. Lecture 5 (2 hours) Paul Levy's isoperimetric inequality

Statistics Lecture (10h): «Statistical Learning»

Speakers: Jean-Michel Loubes and Laurent Risser

Description:

1. Lecture 1 (2 hours) From linear regression to non linear regression. GLM and other extensions.
2. Lecture 2 and 3 (4 hours) Regression trees. Bagging. Boosting.
3. Lecture 4 (2 hours) Introduction to deep learning.
4. Lecture 5 (2 or 4 hours) Stochastic gradient for online learning

THIRD WEEK 15-19 july 2019

## Geometry Lecture (10h) «Wasserstein geometry and Optimal transport »

Speaker : Max FATHI

Description :

1. Lecture 1 (2 hours) Introduction to the optimal transport problem on Euclidean space. Formulations of Monge and Kantorovitch, history, applications. Explicit solution in dimension one. Existence of solutions to the Kantorovitch problem.
2. Lecture 2(2 hours) Kantorovitch duality, existence of a transport map solving the Monge problem. Connection with the Monge-Ampere PDE. Extension to Riemannian manifolds.
3. Lecture 3 (2 hours) Transport cost as a distance on the space of probability measures, and applications in statistics.
4. Lecture 4 (2 hours) The geometry of optimal transport: Benamou-Brenier formula and Riemannian structure of the space of probability measures. Application: gradient flow structure of the heat equation.
5. Lecture 5 (2 hours) Long-time behavior of stochastic processes, and applications to numerical schemes.

## Statistics Lecture (10h): «Using the geometry of the Wasserstein space in Statistical learning»

Speaker : Thierry Klein and Alice Le Brigant

Description:

1. Lecture 1 (2 hours) PCA for a closed convex subset of an Hilbert Spaces. The Wasserstein space in dimension 1, the exponential and logarithm maps, PCA for a closed convex subset of an Hilbert Spaces, functional PCA in the Wasserstein Space.
2. Lecture 2 (2 hours) Wasserstein Barycentre, existence and uniqueness of empirical and population Wasserstein barycentre, Minimax convergence rate for estimating the Wasserstein barycentre of random measures on the real line.
3. Lecture 3 (2 hours) The Fisher information in statistical inference Fisher information; Kullback-Leibler divergence;; Search for the best estimator : Cramér-Rao bound and efficient estimators.
4. Lecture 4 (2 hours) Information geometry manifold. Dual affine connections; Information alpha-manifolds; Divergences; Expected alpha-manifolds of a parametric family of probability distributions.
5. Lecture 5 (2 hours) The Fisher-Rao geometry. Uniqueness of the Fisher-Rao metric; The exponential model; Hyperbolic geometry of Gaussian distributions.

## FOURTH WEEK 22-26 july 2019

Geometry and Statistics (20h): «Implementation and projects. Both in Python using the Anaconda distribution (<https://www.anaconda.com>)»

Speakers: Reda CHHAIBI and Laurent RISSER

Description:

1. Practice 0 (2 hours) Introduction to Python and Standard libraries (NUMPY, PANDAS,...)
2. Practice 1 (4 hours) Statistical inference in image processing using Wasserstein metric. Optimal transport; Segmentation of images.
3. Practice 2 (4 hours) Riemannian techniques for brain images (2D or 3D). Deformation of images; Groups of diffeomorphisms; Euler-Lagrange equations; Geodesic shooting.

4. Practice 3 and 4 (6 hours) Geometry of Toeplitz covariance matrices in signal processing. Orthogonal polynomials on the circle ; Reflection coefficients ; Riemannian p-means of probability distributions.
5. Defense of projects by the students (4 hours)

### 3. CULTURAL SYLLABUS

Three visits will be organized during the month:

Visit 1 : Cité de l'espace <https://www.cite-espace.com/>

Visit 2 : Airbus/Aéroscopia <http://www.musee-aeroscopia.fr/>

Visit 3 : Halle aux Machines <http://www.halledelamachine.fr/>

### 4. RESUME

Toulouse Master School on Data Sciences and Geometry.

Welcome date on 1st July 2019.

The total number of places is limited to 20 students.

The price does not include the trip from China to Toulouse.

Accommodation, main meals (not breakfast) and local transportation are included.

The summer school will deliver a participation diploma.

The summer school is open to BACHELOR/MASTER students.