Mathieu Carrière

Topological Data Analysis and Machine Learning

✆ +1 917-941-5182
✉ mathieu.carriere3@gmail.com
Adress: 390 Ft Washington Ave,
10033 NYC, USA
https://mathieucarriere.github.io/website
Skype: mathieu.carriere
French and American citizenship

Education

Now **Postdoctoral Research Scientist**, *Rabadan Lab, Columbia University, New York, USA*.

08/2018 **Ph.D. in Applied Mathematics and Informatics**, *DataShape, Inria Saclay, Palaiseau, France*.
Title: On metric and statistical properties of topological descriptors for geometric data.

02/2015 **Engineering Degree**, *Ecole Centrale Paris, Châtenay-Malabry, France*.

12/2014 **M.Sc. in Mathematics, Vision and Learning**, *ENS Cachan, Cachan, France*.

Research contributions and impact

My research focuses on Topological Data Analysis and statistical Machine Learning, with an application to biology and genomics. I contributed to the definition and analysis of topological descriptors, as well as on the use of kernels and deep learning methods for them. This work resulted in several scientific articles in top conference proceedings and scientific journals, and has been used in different fields, like bioinformatics and computer graphics. I implemented my work in the C++ GUDHI library and in various Python packages. I have also contributed to the community as a reviewer for several conferences and journals (ICML, JMLR, SoCG, SODA, JACT, DCG, JoCG, TKDE, GD), and through the organization of the New-York Applied Topology Meeting Group at Columbia University.

Research Articles

Machine Learning and Statistics with Persistence Diagrams

- Stable Topological Signatures for Points on 3D Shapes. M. Carrière, S. Oudot, M. Ovsjanikov, *SGP,* 2015.
  Used topological descriptors to improve accuracy in point classification on 3D shape datasets.

- Sliced Wasserstein Kernel for Persistence Diagrams. M. Carrière, M. Cuturi, S. Oudot, *ICML,* 2017.
  Derived a new kernel for Persistence Diagrams improving accuracy on various datasets, such as texture images.

- On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert spaces. M. Carrière, U. Bauer, *SoCG,* 2019.
  Established negative results about the equivalence of Persistence Diagram distances and Hilbert space metrics.

- PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures. M. Carrière, F. Chazal, Y. Ike, T. Lacombe, M. Royer, Y. Umeda, *AISTATS,* 2020.
  Defined general Neural Net architecture for Persistence Diagrams with application in graph classification.

Machine Learning and Statistics with Mapper

- Structure and Stability of the 1-Dimensional Mapper. M. Carrière, S. Oudot, *FoCM,* 2017.
  Defined appropriate metrics to prove the stability of the Mapper clustering algorithm.

- Local Equivalence and Intrinsic Metrics between Reeb graphs. M. Carrière, S. Oudot, *SoCG,* 2017.
  Established a local equivalence between the bottleneck and the Gromov Hausdorff distances for Reeb graphs.

- Statistical Analysis and Parameter Selection for Mapper. M. Carrière, B. Michel, S. Oudot, *JMLR, 2018*.
  Defined confidence intervals and convergence rates for the Mapper clustering algorithm.

- Two-Tier Mapper: a user-friendly clustering method for global gene expression based on topology. R. Jeitziner, M. Carrière, J. Rougemont, S. Oudot, K. Hess, C. Brisken, *Bioinformatics,* 2019.
  Used a modified Mapper clustering algorithm for genomics.

- Topological Data Analysis of single-cell Hi-C contact maps. M. Carrière, R. Rabadán, *Abel Symposium,* 2019.
  Used the Mapper clustering algorithm to establish confidence regions for single-cell Hi-C contact maps.

- Approximation of Reeb spaces with Mappers and applications to stochastic filters. M. Carrière, B. Michel, *Preprint*, 2019.
  Studied applications of Mapper in machine learning.

  ### Optimal transport
- MREC: a fast and versatile framework for aligning and matching point clouds with applications to single cell molecular data. A. Blumberg, M. Carrière, M. Mandell, R. Rabadán, S. Villar, *Preprint*, 2020.
  Established recursive approximation scheme for fast computation of optimal transport.

Skills

Languages	French (native), English (professional TOEFL 627/677), Spanish (B1 level).
Code	C++, Python (proficient), R, Matlab (prior experience).

Coding projects

- Cover complex module of the C++ GUDHI library, *http://gudhi.gforge.inria.fr/doc/latest/group__cover__complex.html*.
- sklearn-tda: a scikit-learn compatible Python library implementing various vectorization and kernel methods used in TDA, *https://github.com/MathieuCarriere/sklearn_tda*.
- PersLay: a neural network layer for optimizing vectorizations of persistence diagrams, *https://github.com/MathieuCarriere/perslay*.
- MREC: a fast computational tool for optimal transport and applications to genomics, *https://github.com/MathieuCarriere/mrec*.
- My other projects can be found on my GitHub account, *https://github.com/MathieuCarriere*.

Grants

- Mobility Grant (1000 euros) from the DAAD exchange program.
- Mobility Grant (1000 euros) from the STIC doctoral school.
- Best Scientific Contribution 2017 (2nd Prize – 600 euros) from the STIC doctoral school.
- Funding Support (1800 dollars) from ICML 2017.
- Thiessé de Rosemont / Schneider Prize (10,000 euros) from Chancellerie des Universités de Paris.

Teaching Activities

I was a teaching assistant for the following courses.

2015–2017	*Topological Data Analysis,* Ecole Polytechnique, Palaiseau, France.
2016–2017	*Basics of Algorithmic and Progamming,* Ecole Polytechnique, Palaiseau, France.

Presentations for Workshops and Conferences

I gave presentations at the following international conferences.

11/2015	*Journées de Géométrie Algorithmique,* IESC, Cargèse, Corsica.
12/2015	*Computational and Methodological Statistics,* London University, London, UK.
04/2016	*Stochastic Geometry and its Applications,* Université de Nantes, Nantes, France.
07/2016	*Applied Topology: Methods, Computation and Science,* Politecnica di Torino, Torino, Italy.
09/2017	*France Japan Machine Learning Workshop,* ENS Ulm, Paris, France.
10/2017	*Amazon Graduate Research Symposium,* Amazon Meeting Center, Seattle, USA.
12/2017	*Journées de Géométrie Algorithmique,* Centre Paul Langevin, Aussois, France.
08/2018	*TRIPODS Bootcamp on Topology and Machine Learning,* ICERM, Providence, USA.
01/2019	*AMS Special Session on Topological Data Analysis,* Convention Center, Baltimore, USA.
02/2019	*Applied Algebraic Topology Research Network,* via Bluejeans.
04/2019	*Topology, Geometry and Data Seminar,* OSU, Columbus, USA.
06/2019	*Symposium on Computational Geometry,* PSU, Portland, USA.
07/2019	*International Congress on Industrial and Applied Mathematics,* UPV, Valencia, Spain.

01/2020   *TDA Meeting,* UF, Gainesville, USA.

03/2020   *Probability and Society Symposium,* Columbia University, New-York, USA.

05/2020   *SIAM Conference on Mathematics of Data Science,* Cincinnati, USA.

References

Steve Oudot
DataShape team
Inria Saclay
91120 Palaiseau, France
steve.oudot@inria.fr
+33 174 854 216

Marco Cuturi
CREST - ENSAE
Université Paris-Saclay
91120 Palaiseau, France
marco.cuturi@ensae.fr
+33 170 266 857

Raúl Rabadán
Systems Biology Department
Columbia University
New-York, USA
rr2579@columbia.edu