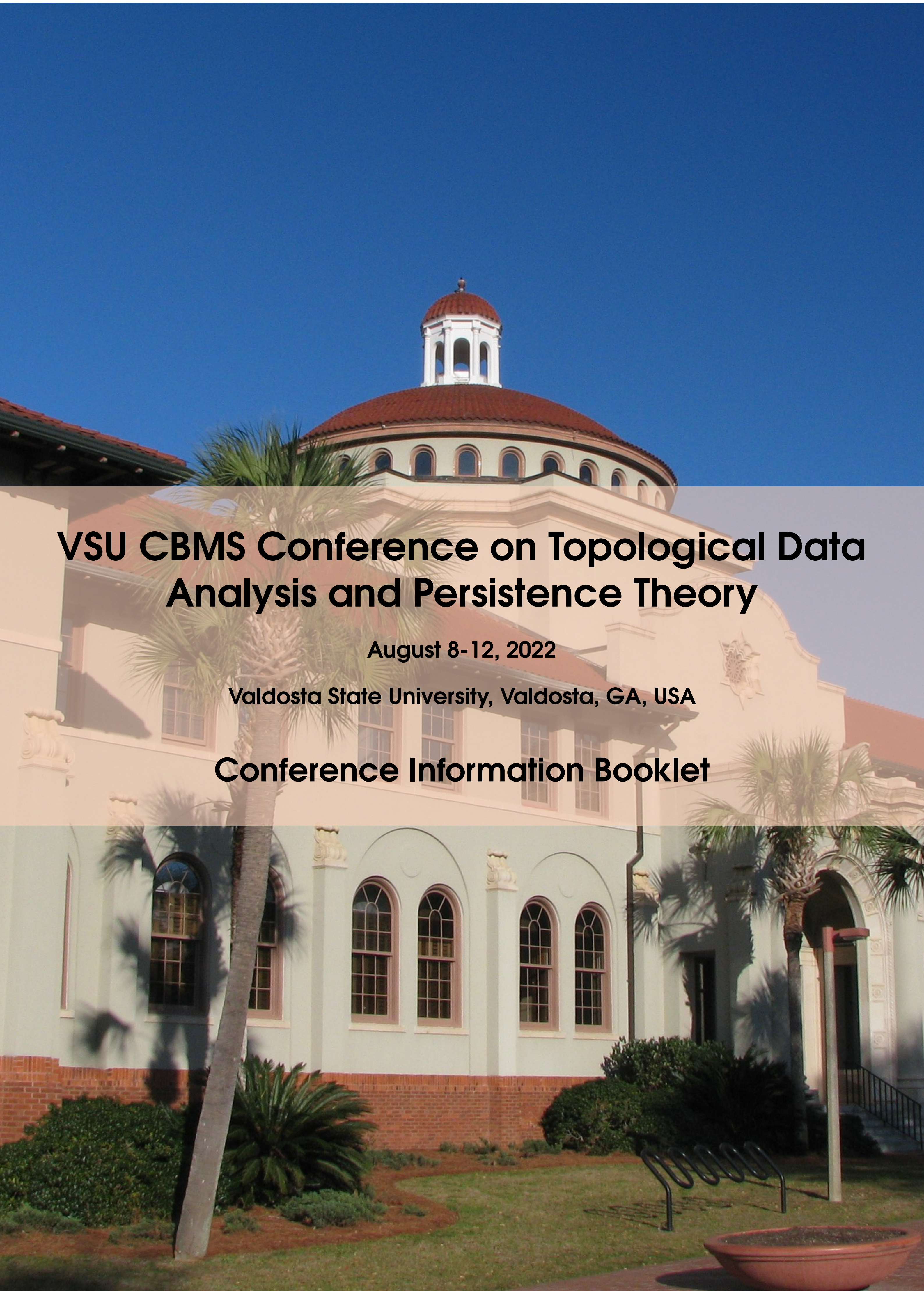


VSU CBMS Conference on Topological Data Analysis and Persistence Theory

August 8-12, 2022

Valdosta State University, Valdosta, GA, USA

Conference Information Booklet

Organizing Committees and Sponsoring Institutions

Sponsoring Institutions

This conference is supported by the National Science Foundation (NSF 19-539 NSF-CBMS – Award Number 2132497).

Lecturer

- PETER BUBENIK, Professor, *Department of Mathematics, University of Florida*, `Gainesville, FL, USA`

Organizing Committee

- SHAUN V. AULT, Professor and Department Head, *Department of Mathematics, Valdosta State University*, `Valdosta, GA, USA`
- JOSÉ A. VÉLEZ-MARULANDA, Professor, *Department of Mathematics, Valdosta State University*, `Valdosta, GA, USA`

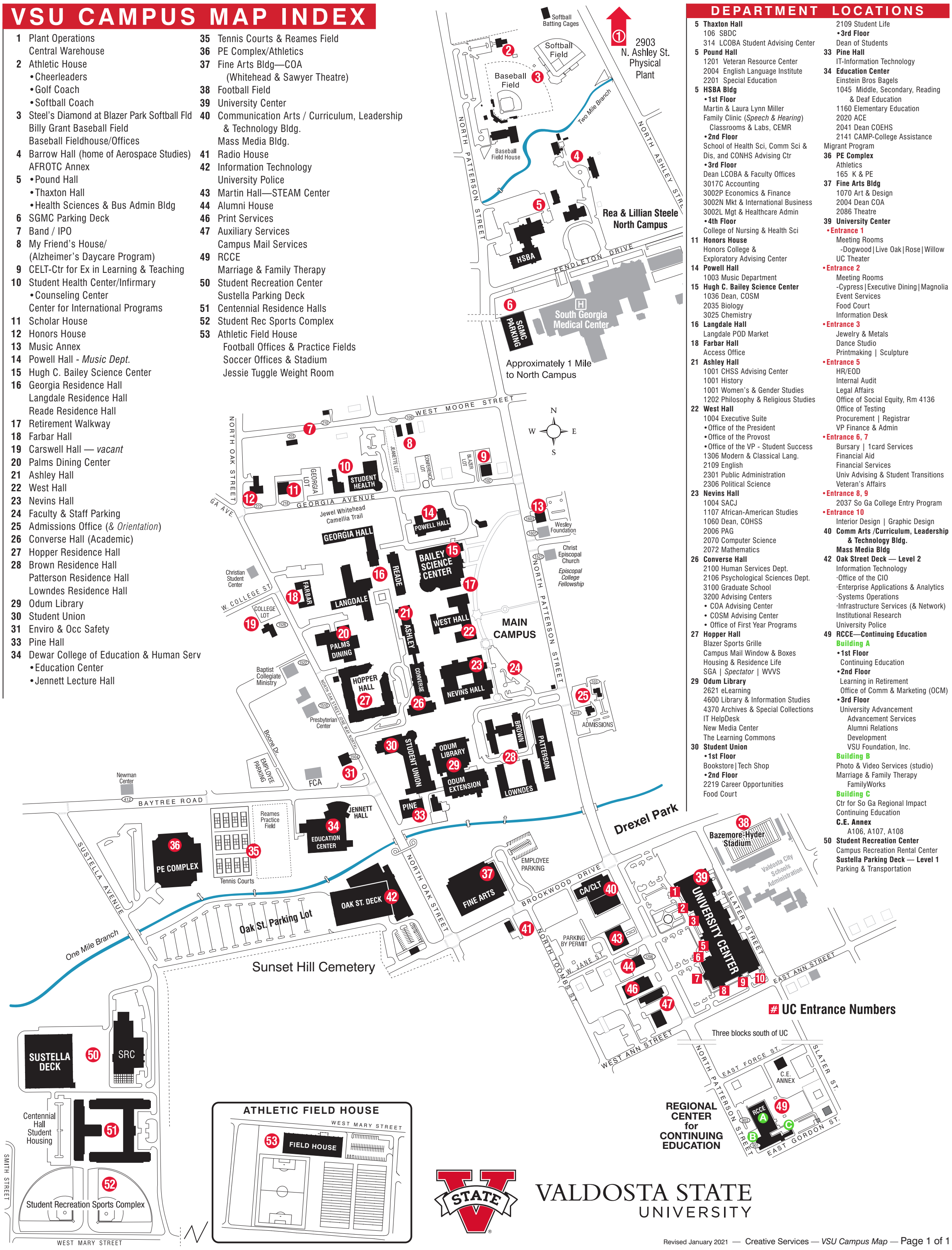
Contents

1. General Information

1.1 About the conference

The Department of Mathematics at the Valdosta State University is pleased to announce the VSU CBMS Conference on Topological Data Analysis and Persistence Theory which will be held from August 8 to August 12, 2022. The main goal of this conference is to provide an introduction to topological data analysis (TDA) and persistence theory (PT) to a broader audience. TDA and PT are relatively recent methods useful for finding important features in large data sets using ideas from traditionally theoretical branches of mathematics such as algebra and topology. This conference consists in a series of daily lectures given by **Dr. Peter Bubenik** at the University of Florida in Gainsville, FL. The topics of these lectures include a review of the basic mathematical concepts related to TDA and PT, interactions with statistical methods and machine learning as well as current applications and software implementation. Finally, this lecture series will end with a discussion of advanced topics and current research related to TDA and PT. There will also be three structured Lab Sessions where the participants will be introduced to software that can be used to compute various TDA functions on data sets.

1.2 Campus map



1.3 Description of the Lectures

- **All lectures take place in the STEAM Center, First Floor, Main Room** (Building 43 in the Campus Map, Address: 1302 N Patterson St, Valdosta, GA 31601)
- **Lab sessions take place in Odum Library, Room 3270** (Building 29 in the Campus Map)

Day 1. Introduction to Homology and Persistent Homology

- Lecture 1. Motivation and Basic Constructions – The shape of data, simplicial and cubical complexes, homology, persistent homology, persistence diagrams, Čech complexes, Vietoris-Rips complexes, digital images
- Lecture 2. Foundational Results – The persistence algorithm, Wasserstein distance, stability, flavors of persistence (e.g. zigzag, multiparameter)

Day 2. Mathematics of Persistent Homology

- Lecture 3. Algebra of Persistence Modules – Commutative algebra, representations of quivers, graded modules, algebraic stability
- Lecture 4. Geometry and Combinatorics – Geometric stability, Möbius inversion, coarse geometry

Day 3. Statistics and Machine Learning

- Lecture 5. Statistics – Hilbert spaces, kernels, persistence landscapes, averages, variance, hypothesis testing, permutation tests, principal component analysis, subsampling
- Lecture 6. Machine Learning – Classification, regression, support vector machines, deep learning, multilayer perceptrons, convolutional neural networks, topological loss, topological layers

Day 4. Applications and Software

- Lecture 7. Applications – Preprocessing, mathematical encoding of data, time series, case studies
- Lecture 8. Software and Algorithms – Computational advances, guide to current software

Day 5. Advanced Topics and Current Research Problems

- Lecture 9. Multiparameter Persistent Homology – Theory, algorithms, software, open problems
- Lecture 10. Mathematics of Persistent Homology – Graded persistence diagrams, virtual persistence diagrams, categorical stability, universal constructions, Cerf theory, open problems”

1.4 Schedule

VSU CBMS CONFERENCE ON TOPOLOGICAL DATA ANALYSIS AND PERSISTENCE THEORY					
	Monday 08/22	**Tuesday 09/22**	**Wednesday 10/22**	**Thursday 11/22**	**Friday 12/22**
9:00–9:30	WELCOME	COFFEE	COFFEE	COFFEE	COFFEE
9:30–10:30	LECTURE 1	LECTURE 3	LECTURE 5	LECTURE 7	LECTURE 9
10:30-11:00	COFFEE	COFFEE	COFFEE	COFFEE	COFFEE
11:00-12:00	WORK & GROUP DISCUSSION	WORK & GROUP DISCUSSION	LECTURE 6	WORK & GROUP DISCUSSION	LECTURE 10
12:00-13:30	LUNCH	LUNCH	LUNCH	LUNCH	CONFERENCE ENDS
13:30-14:30	LECTURE 2	LECTURE 4	CONFERECE EXCURSION	LECTURE 8	
14:30-15:00	COFFEE	COFFEE		COFFEE	
15:00-16:00	LAB SESSION	LAB SESSION		LAB SESSION	
16:00-17:00	DISCUSSION	DISCUSSION		DISCUSSION	
17:00-19:00	TOUR TO VALDOSTA'S DOWNTOWN	TOUR TO THE PLANETARIUM	FREE	CONFERENCE BANQUETTE	

- **All lectures take place in the STEAM Center, First Floor, Main Room** (Building 43 in the Campus Map, Address: 1302 N Patterson St, Valdosta, GA 31601)
- **Lab sessions take place in Odum Library, Room 3270** (Building 29 in the Campus Map)

1.5 Places of interests

- **GŪD Craft Company Co.**, `133 North Patterson Street, Valdosta, GA`. Open from 7AM to 5 PM.
- **Georgia Beer Company**, `109 S. Briggs St, Valdosta, Georgia 31601`. Open from 4 PM to 9 PM.
- **The Southern Cellar**, `120 N. Patterson Street, Valdosta, GA 31601`. Open from 3:30 PM to 9:30 PM (Closed on Monday).
- **Wild Adventures**, `3766 Old Clyattville Rd, Valdosta, GA`. Open from 10 AM to 7 PM.
- **Mom & Dad's Italian Restaurant**, `4143 North Valdosta Road, Valdosta, GA`. Open 5PM to 9PM
- **Smok'n Pig BBQ**, `4228 N Valdosta Rd, Valdosta, GA` . Open 11AM - 9PM
- **Parking** is available at the following areas (not parking permit is required at the moment):
  - In the lot of the University Center `(Building 39 in the Campus Map)`
  - In the Oak St. Parking Lot next to the Oak St. Deck `(Building 42 in the Campus Map)`
  - In the Conference Lot near the Student Health `(Building 10 in the Campus Map)`

2. Further documentation (by P. Bubenik)

Topological Data Analysis and Persistence Theory
Peter Bubenik

Introduction

The goal of this conference is to introduce many aspects of Topological Data Analysis (TDA) and Persistence Theory. We will start with the basic constructions and fundamental results and work our way up to recent research progress and open problems. There will be ten lectures, three group work and discussion sessions, three lab sessions, and three open discussion sessions.

First a disclaimer. The subject at hand is already much larger than what can be covered in the time allotted. For example, I will focus on persistent homology and not discuss other important tools in TDA such as the Mapper algorithm and Reeb graphs. Many interesting theoretical results and exciting applications will be omitted. Hopefully, this introduction will encourage participants to discover other parts of this subject and make it easier for them to do so.

The main part of the conference is a sequence of ten lectures, which I am looking forward to be giving. Each of these will tackle one topic and together they are roughly equally divided between TDA and persistence theory. The lectures will be supplemented by three short sets of exercises which may be done with pencil and paper. Time is allotted for participants to work on these in small groups. An important aim of these exercises in addition to learning the material is to get to know some of the other participants and to have fun! Perhaps the second main part of the conference is a sequence of three lab sessions in which participants will learn how to perform TDA computations using the programming language R. These sessions will be led by Iryna Hartsock and John Bush.

Historical background

The roots of persistent homology may already be found in the work of Marston Morse [Mor25]. As observed by Raoul Bott [Bot88], the following is a central result of Morse's theory. Let f be a Morse function on a manifold M. Morse proved that not only is the ith Betti number a lower bound on the number of critical points of f of degree i, but the excess critical points may be partitioned into pairs that differ in degree by 1. What Morse did not show is that this pairing is canonical and that if we consider the corresponding pairs of critical values we obtain what is now called the persistence diagram of f, for which there is an efficient algorithm and which is useful in applications.

Vanessa Robins [Rob99] defined persistent homology groups and studied their ranks, which she called persistent Betti numbers. Herbert Edelsbrunner, David Letscher, and Afra Zomorodian gave an efficient algorithm for computing the persistence diagram and showed that it could be used to obtain the persistent Betti numbers. Zomorodian and Gunnar Carlsson [ZC05] used a structure theorem to justify the pairing of critical values in the persistence diagram. An early application of persistent homology in topological data analysis was the study of the space of natural images by Carlsson, Tigran Ishkhanov, Vin de Silva, and Zomorodian [CIdSZ08].

References

[Bot88] Raoul Bott. Morse theory indomitable. *Inst. Hautes Études Sci. Publ. Math.*, (68):99–114 (1989), 1988.

[CB15] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *J. Algebra Appl.*, 14(5):1550066, 8, 2015.

[CCR13] Joseph Minhow Chan, Gunnar Carlsson, and Raul Rabadan. Topology of viral evolution. *Proc. Natl. Acad. Sci. USA*, 110(46):18566–18571, 2013.

[CCSG$^+$09] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the 25th annual symposium on Computational geometry*, SCG '09, pages 237–246, New York, NY, USA, 2009. ACM.

[CdSGO16] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. SpringerBriefs in Mathematics. Springer, [Cham], 2016.

[CdSO14] Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geom. Dedicata*, 173:193–214, 2014.

[CIdSZ08] Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *Int. J. Comput. Vision*, 76(1):1–12, 2008.

[CSEH07] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.

[Les15] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Found. Comput. Math.*, 15(3):613–650, 2015.

[Mor25] Marston Morse. Relations between the critical points of a real function of n independent variables. *Trans. Amer. Math. Soc.*, 27(3):345–396, 1925.

[Rob99] V. Robins. Towards computing homology from finite approximations. In *Proceedings of the 14th Summer Conference on General Topology and its Applications (Brookville, NY, 1999)*, volume 24, pages 503–532 (2001), 1999.

[ZC05] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.

NSF/CBMS Conference Valdosta State University August 8–12, 2022

Topological Data Analysis and Persistence Theory

Peter Bubenik

Outline

Lectures

Lecture 1: Motivation and Basic Constructions. We will start with motivations for topological data analysis (TDA) and an overview of TDA. I will introduce the some of the basic mathematical objects and constructions of TDA and discuss the corresponding categories and functors. I will end with an introduction to persistent homology.

Lecture 2: Foundational Results. In the second lecture, we will study the persistence algorithm, Wasserstein distances, interleaving, the isometry theorem, sublevelset stability, Wasserstein stability, and various flavors of persistence.

Lecture 3: Combinatorics and Geometry. In the third lecture, we will see how a combinatorial construction, Möbius inversion, produces persistence diagrams and graded persistence diagrams. I will present the Gromov-Hausdorff stability of persistent homology and introduce generalized persistence diagrams.

Lecture 4: TDA and Statistics. In the fourth lecture, we will discuss feature maps and kernels for TDA. I will use the erosion of persistence modules to define the persistence landscape and discuss some of its properties. I will show how TDA may be used for statistics and exploratory data analysis and discuss dimensionality reduction and convergence results.

Lecture 5: The Algebra of Persistence Modules. In the fifth lecture, we will consider persistence modules as representations of quivers and as graded modules and discuss the classification theorem of persistence modules. We will study maps of persistence modules and the algebraic stability theorem. I will introduce q-tame persistence modules and also consider the algebra of generalized persistence modules.

Lecture 6: TDA and Machine Learning. In the sixth lecture, I will discuss learning problems, clustering, classification, regression, and deep learning, and their connections to TDA.

Lecture 7: Noise and multiparameter persistence. In the seventh lecture, I will discuss noise in data and the resulting instability of persistent homology and how it may be addressed. This will lead us to multiparameter persistent homology (MPH). I will discuss the mathematics of MPH and computational approaches to MPH.

Lecture 8: Applications of TDA. I will present case studies that illustrate a number of important topics in TDA: how long and short bars reveal topology and geometry; the importance of preprocessing and how to represent data as a filtered simplicial complex; how to apply TDA to time series data; and the use of representative cycles in TDA.

Lecture 9: Mathematics and Persistence. In the second-last lecture, I will discuss some interesting mathematical connections to persistence that have arisen in recent research: formal sums on metric spaces; optimal transportation; generalized Morse theory; and abelian categories.

Lecture 10: Software and Algorithms. In the final lecture, I will discuss algorithmic improvements to persistence computations and various persistent homology and TDA software.

Group Work and Discussion

Group Work 1. These exercises with give you practice with the basic constructions of TDA as well as computing homology and the ranks of persistent homology vector spaces.

Group Work 2. These exercises will lead you through the computation of barcodes and persistence diagrams via matrix reduction and via Möbius inversion. You will also compute graded persistence diagrams.

Group Work 3. These exercises will lead you through the computation of the persistence landscape.

Lab Sessions

Lab 1. You will learn how to sample points, construct Voronoi cells, Delaunay complexes and Vietoris-Rips complexes. You will compute barcodes and persistence diagrams, and visualize representative cycles.

Lab 2. You will compute death vectors and persistence landscapes, and learn how to average them, plot their differences, and compute p values. You will perform principal components analysis (PCA) and use it to plot low-dimensional projections, loading vectors, and the explained variance.

Lab 3. You will combine TDA with support vector machine (SVM) to classify data and to perform regression. You will clean noisy data using k-nearest neighbors (kNN).

NSF/CBMS Conference Valdosta State University August 8–12, 2022

Topological Data Analysis and Persistence Theory

Peter Bubenik

Preparation

Assumed background

If you are not familiar with any of these topics, or have forgotten about them, then please look them up on Wikipedia, for example, before the conference.

Linear algebra: vector space; linear map; matrix; kernel, image, cokernel, and rank of a linear map; rank-nullity theorem; row reduction of a matrix; (real) inner product space; (real) Hilbert space; the sequence space, ℓ^2; the space of square-integrable functions, L^2

Statistics: statistical hypothesis testing; p value

Topology: simplicial complex; abstract simplicial complex; simplicial map; simplicial homology (with coefficients taken to be in a field instead of being integers so that one obtains vector spaces instead abelian groups)

Category theory: category; functor; natural transformation

Preparation for the lab sessions

Install R and RStudio. Download and install the following free software on your laptop. First R from `https://cran.rstudio.com/` and then RStudio Desktop from `https://www.rstudio.com/products/rstudio/download/`.

Introduction to R. Work through the following R introduction by typing (or copying and pasting) commands from `https://www.r-tutor.com/r-introduction` into the console in RStudio.

Suggested reading

To obtain a thorough background, it is recommended that you read (or skim) the beginning of one of the following two excellent recent books.

For the mathematically inclined reader: Chapters 1 and 2 in Topological Data Analysis for Genomics and Evolution, by Raúl Rabadán and Andrew Blumberg, Cambridge University Press, 2020.

For the computationally inclined reader: Chapters 1–3 in Computational Topology for Data Analysis, by Tamal Krishna Dey and Yusu Wang, Cambridge University Press, 2022. `https://www.cs.purdue.edu/homes/tamaldey/book/CTDAbook/CTDAbook.pdf`