

## COURSE OUTLINE: MH4517

Course Title	<b>Data Applications in Natural Sciences</b>		
Course Code	<b>MH4517</b>		
Offered	Study Year 3, Sem 2   Study Year 4, Sem 2		
Course Coordinator	Xia Kelin (Asst Prof)	XIAKELIN@ntu.edu.sg	6513 7464
Pre-requisites	MH1402 OR MH1403 OR CZ2001		
Mutually exclusive	MH4600		
AU	4		
Contact hours	Lectures: 39, Tutorials: 12		
Approved for delivery from	AY 2020/21 semester 2		
Last revised	19 Oct 2020, 09:19		

### Course Aims

This course aims to provide the recent progress on geometric data analysis, topological data analysis, the combination with machine learning models, and their applications in nature data analysis. In this course you will develop skills in geometric and topological modeling, and the analysis of complicated physical/chemical/biological data.

### Intended Learning Outcomes

Upon successfully completing this course, you should be able to:

1. Analyze physical/chemical/biological data with tools from topological data analysis
2. Analyze physical/chemical/biological data with tools from geometric data analysis
3. Design geometric and topological models for problems in nature sciences
4. Solve nature science problems with geometry and topology based learning models

### Course Content

Topological data analysis models, including simplicial complex, nerve theorem, homology, cohomology, filtration, persistent homology, Morse theory, Hodge-Laplacian, Reeb graph.

Geometric data analysis models, including multidimensional scaling, isomap, diffusion map, spectral graph, manifold learning, differential forms.

Geometry and topology based learning, including data representation, feature engineering, molecular/chemical descriptors, graph neural network.

## Assessment

Component	Course ILOs tested	SPMS-MAS Graduate Attributes tested	Weighting	Team / Individual	Assessment Rubrics
<b>Continuous Assessment</b>					
<b>Tutorials</b>					
Project	1, 2, 3, 4	1. a, b, c, d 2. a, b, c, d 3. a, b 4. a 5. a	20	both	See Appendix for rubric
<b>Mid-semester Quiz</b>					
Short Answer Questions	1, 2, 3	1. a, b, c, d 2. a, b, c, d	30	individual	See Appendix for rubric
<b>Examination</b>					
Short Answer Questions	1, 2, 3, 4	1. a, b, c, d 2. a, b, c, d	50	individual	See Appendix for rubric
<b>Total</b>			<b>100%</b>		

These are the relevant SPMS-MAS Graduate Attributes.

### 1. Competence

- a. Independently process and interpret mathematical theories and methodologies, and apply them to solve problems
- b. Formulate mathematical statements precisely using rigorous mathematical language
- c. Discover patterns by abstraction from examples
- d. Use computer technology to solve problems, and to communicate mathematical ideas

### 2. Creativity

- a. Critically assess the applicability of mathematical tools in the workplace
- b. Build on the connection between subfields of mathematics to tackle new problems
- c. Develop new applications of existing techniques
- d. Critically analyse data from a multitude of sources

### 3. Communication

- a. Present mathematics ideas logically and coherently at the appropriate level for the intended audience
- b. Work in teams on complicated projects that require applications of mathematics, and communicate the results verbally and in written form

### 4. Civic-mindedness

- a. Develop and communicate mathematical ideas and concepts relevant in everyday life for the benefits of society

### 5. Character

- a. Act in socially responsible and ethical ways in line with the societal expectations of a mathematics professional, particularly in relation to analysis of data, computer security, numerical computations and algorithms

## Formative Feedback

In the beginning of each class, there will be a 5 to 10 minutes review with several questions to help you understand the critical concepts in geometric and topological data analysis (1,2,3,4).

In the project, you will be asked to apply the geometric and topological tools and models to physical/chemical/biological data analysis. There will be a midterm report of the project. Individual meeting will be scheduled for the discussion of the project during the semester (1,2,3,4).

Several in-class quiz will be scheduled to help review the critical concepts, models and tools (1,2,3,4).

Midterm exam will cover half of the materials (1,2). A detailed explanation of the midterm problems after the exam will contribute to a better understanding.

Examiner's Report for the final exam will be uploaded to NTULearn for reviewing the class materials.

## Learning and Teaching Approach

<b>Lectures</b> (39 hours)	In each lecture, a short review about 5 to 10 minutes will be given to test the understanding of 1,2,3,4. Group discussion will be conduct for 3,4.
<b>Tutorials</b> (12 hours)	Tutorials will focus on the analysis of problem with various models, tools, and algorithms (1,2,3,4)

## Reading and References

Edelsbrunner, Herbert. A short course in computational geometry and topology. (No. Mathematical methods). Berlin, Germany:: Springer, 2014. (eBook ISBN 978-3-319-05957-0)

Edelsbrunner, Herbert, and John Harer. Computational topology: an introduction. American Mathematical Soc., 2010. (ISBN-13: 978-0821849255)

A Mathematical Introduction to Data Analysis, Yuan Yao, Hong Kong University of Science and Technology. (Online lecture notes [https://github.com/yao-lab/yao-lab.github.io/blob/master/book\\_datasci.pdf](https://github.com/yao-lab/yao-lab.github.io/blob/master/book_datasci.pdf))

Bronstein, Michael M., Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric deep learning: going beyond euclidean data." IEEE Signal Processing Magazine 34, no. 4 (2017): 18-42. (Research paper)

Duc Nguyen, Zixuan Cang, and Guo-Wei Wei, A review of mathematical representations of biomolecular data, Physical Chemistry Chemical Physics, 22, 4343-4367 (2020). (Research paper)

## Course Policies and Student Responsibilities

### (1) General

You are expected to complete all lectures and tutorials, take the quizzes, and complete the project. You are expected to take responsibility to follow up with course notes, assignments and course related announcements if you are absent.

### (2) Absenteeism

Absence from examination without a valid reason will affect your overall course grade. Valid reasons include falling sick supported by a medical certificate and participation in NTU's approved activities supported by an excuse letter from the relevant bodies.

## Academic Integrity

Good academic work depends on honesty and ethical behaviour. The quality of your work as a student relies on adhering to the principles of academic integrity and to the NTU Honour Code, a set of values shared by the whole university community. Truth, Trust and Justice are at the core of NTU's shared values.

As a student, it is important that you recognize your responsibilities in understanding and applying the principles of academic integrity in all the work you do at NTU. Not knowing what is involved in maintaining academic integrity does not excuse academic dishonesty. You need to

actively equip yourself with strategies to avoid all forms of academic dishonesty, including plagiarism, academic fraud, collusion and cheating. If you are uncertain of the definitions of any of these terms, you should go to the [Academic Integrity website](#) for more information. Consult your instructor(s) if you need any clarification about the requirements of academic integrity in the course.

## Course Instructors

Instructor	Office Location	Phone	Email
Xia Kelin (Asst Prof)	SPMS-MAS-05-18	6513 7464	XIAKELIN@ntu.edu.sg

## Planned Weekly Schedule

Week	Topic	Course ILO	Readings/ Activities
1	Basic introduction of physical/chemical/biological data in nature sciences.	1, 2	Bronstein, Michael M., Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric deep learning: going beyond euclidean data." IEEE Signal Processing Magazine 34, no. 4 (2017): 18-42.  Duc Nguyen, Zixuan Cang, and Guo-Wei Wei, A review of mathematical representations of biomolecular data, Physical Chemistry Chemical Physics, 22, 4343-4367 (2020).
2	Topological data analysis models, including simplicial complex, nerve theorem, homology, cohomology, filtration, persistent homology, Morse theory, Hodge-Laplacian, Reeb graph.	1, 3, 4	A short course in computational geometry and topology-Chapter 8 to 11; Computational topology: an introduction-chapter 3 to 7
3	Topological data analysis models, including simplicial complex, nerve theorem, homology, cohomology, filtration, persistent homology, Morse theory, Hodge-Laplacian, Reeb graph.	1, 3, 4	A short course in computational geometry and topology-Chapter 8 to 11; Computational topology: an introduction-chapter 3 to 7
4	Topological data analysis models, including simplicial complex, nerve theorem, homology, cohomology, filtration, persistent homology, Morse theory, Hodge-Laplacian, Reeb graph.	1, 3, 4	A short course in computational geometry and topology-Chapter 8 to 11; Computational topology: an introduction-chapter 3 to 7
5	Topological data analysis models, including simplicial complex, nerve theorem, homology, cohomology, filtration, persistent homology, Morse theory, Hodge-Laplacian, Reeb graph.	1, 3, 4	A short course in computational geometry and topology-Chapter 8 to 11; Computational topology: an introduction-chapter 3 to 7
6	Topological data analysis models, including simplicial complex, nerve theorem, homology, cohomology, filtration, persistent homology, Morse theory, Hodge-Laplacian, Reeb graph.	1, 3, 4	A short course in computational geometry and topology-Chapter 8 to 11; Computational topology: an introduction-chapter 3 to 7
7	Geometric data analysis models, including multidimensional scaling, isomap, diffusion map, spectral graph, manifold learning, differential forms.	2, 3, 4	A Mathematical Introduction to Data Analysis- chapter 5 to 7
8	Geometric data analysis models, including multidimensional scaling, isomap, diffusion map, spectral graph, manifold learning, differential forms.	2, 3, 4	A Mathematical Introduction to Data Analysis- chapter 5 to 7
9	Geometric data analysis models, including multidimensional scaling, isomap, diffusion map, spectral graph, manifold learning, differential forms.	2, 3, 4	A Mathematical Introduction to Data Analysis- chapter 5 to 7
10	Geometric data analysis models, including multidimensional scaling, isomap, diffusion map, spectral graph, manifold learning, differential forms.	2, 3, 4	A Mathematical Introduction to Data Analysis- chapter 5 to 7
11	Geometry and topology based learning, including data representation, feature engineering, molecular/chemical descriptors, graph neural network.	1, 2, 3, 4	Bronstein, Michael M., Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric deep learning: going beyond euclidean data." IEEE Signal Processing Magazine 34, no. 4 (2017): 18-42.  Duc Nguyen, Zixuan Cang, and Guo-Wei Wei, A review of mathematical representations of biomolecular data,

			Physical Chemistry Chemical Physics, 22, 4343-4367 (2020).
12	Geometry and topology based learning, including data representation, feature engineering, molecular/chemical descriptors, graph neural network.	1, 2, 3, 4	<p>Bronstein, Michael M., Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric deep learning: going beyond euclidean data." IEEE Signal Processing Magazine 34, no. 4 (2017): 18-42.</p> <p>Duc Nguyen, Zixuan Cang, and Guo-Wei Wei, A review of mathematical representations of biomolecular data, Physical Chemistry Chemical Physics, 22, 4343-4367 (2020).</p>
13	Geometry and topology based learning, including data representation, feature engineering, molecular/chemical descriptors, graph neural network.	1, 2, 3, 4	<p>Bronstein, Michael M., Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric deep learning: going beyond euclidean data." IEEE Signal Processing Magazine 34, no. 4 (2017): 18-42.</p> <p>Duc Nguyen, Zixuan Cang, and Guo-Wei Wei, A review of mathematical representations of biomolecular data, Physical Chemistry Chemical Physics, 22, 4343-4367 (2020).</p>

## Appendix 1: Assessment Rubrics

### Rubric for Tutorials: Project (20%)

Grading Criteria	Exceptional (17-20)	Effective (14-16)	Acceptable (10-13)	Developing (0-9)
Accuracy	The interpretation is highly accurate, concise and precise.	The interpretation is mostly accurate. Some parts can be better explained or more succinct.	The interpretation is somewhat accurate. However, it contains some inaccuracies, missing points or ideas that are not related to the interpretation.	The interpretation are mostly inaccurate.
Thoroughness	The literature review was comprehensive and rigorous. It includes several different perspectives, including a good spread of the first and latest ideas on the topic.	The literature review was mostly comprehensive and rigorous. It can improve in terms of the selection of the works relating to the topic.	The literature review was adequate. It covers some of the major works relating to the topic. References to primary source is largely missing.	The literature review was not thorough. It is based on a single source of information and/or inaccurate or unreliable secondary sources.
Presentation	Very clear and organized. It is easy to follow your train of thought	Mostly clear and organized. Some parts can have better transitions.	Somewhat clear. It requires some careful reading to understand what you are writing.	Mostly unclear and messy. It is difficult to understand what you are writing as there is no clear flow of ideas.
Originality	Evidence of extensive synthesis of ideas from different perspectives such that there is a very convincing original interpretation and that goes beyond what is already discussed in literature.	Evidence of some synthesis of ideas which lead to an original interpretation. The interpretation is good original summary of what is discussed in literature.	Evidence of an attempt to synthesise ideas. However, the attempt contains some misunderstandings.	No synthesis of ideas or originality. It is a repetition of what people have said or a laundry list of ideas with little interpretation.

Please Note: In principle, students in the same group share the same group marks. However, there can be some individual variation within a group, depending on the evaluation of the tutor and the feedback from the peers. Students may be awarded more marks for showing exemplary contribution to other team members' learning that goes beyond what is required, whereas students who have not contributed sufficiently may receive lower marks than the rest of the team members.

### Rubric for Mid-semester Quiz: Short Answer Questions (30%)

Point-based marking

### Rubric for Examination: Short Answer Questions (50%)

Point-based marking