Mathematical Methods in Linguistics

Fall 2021

Course	Info
Course#	lin539
Time	MWF 10:30-11:25am
Location	Frey 216
Website	lin539.thomasgraf.net
Instructor	Thomas Graf
Email	[coursenumber]@thomasgraf.net
Office hours	MF 11:30am – 1:00pm
Office	SBS N-249/SBS frontyard

A friendly plug: If you are interested in this class, also consider attending the department's Mathematical Linguistics Reading Group (https://mlrg.thomasgraf.net). Due to COVID, we don't quite know yet how we'll run it this year, I'll keep you posted.

1 Course Outline

1.1 Bulletin Description

An overview of the mathematical foundations of theoretical and computational linguistics. Topics covered include set theory, morphisms, logic and model theory, algebra, lattices, lambda calculus, probability theory, information theory, and basics of formal language theory. A strong emphasis is put on the linguistic application of the mathematical concepts in the study and analysis of natural language data.

1.2 Full Description

This course is an introduction to mathematics in linguistics. It aims to help students familiarize themselves with mathematical concepts and applications that are widely relevant to theoretical and/or computational linguistics. This covers a wide range of topics, mostly from *discrete mathematics*. The course is also very different from what you did in high school, there's precious

few numbers here and we don't care much about trigonometry or calculating compound interest. In contrast to a proper mathematics course, we also focus more on techniques and tools rather than theorems and proofs. This means that you will learn how to work with things like matrices, semirings, and lattices, but you won't have to prove things about them. So this is more like a CS methods course than a proper math class.

For more information about the content, see the Selected Topics section. You will see that the schedule for this class is very ambitious. It has to be: this class serves an integral function of our Computational Linguistics MA and must get students to a level where they can take courses and read textbooks on mathematically demanding topics such as machine learning. This is not your typical graduate-level course, it is a **boot camp**, so be prepared to invest a fair amount of blood, sweat, and tears.

1.3 Prerequisites

No prior mathematical or computational experience is required. I have run this course as an independent study with high school students who had no prior experience in linguistics, and they could follow along for the most part and only needed some help in specific areas.

2 Mode of Instruction: Flipped Classroom Lite

The most important mathematics skill is the ability to learn mathematics on your own, without the help of an instructor. Whether you're doing computational linguistics in the industry or as an academic, sooner or later you will come across some tool or technique that builds on an area of math you have never encountered before. It is vital for your future career that you can pick up a textbook or survey paper and teach yourself how this unfamiliar kind of math works.

Unfortunately, math is already a very challenging topic for most people, and learning it without somebody's help seems impossible to many. In order to teach you how to make the transition from listener to autonomous learner, this course is run as a **hybrid class** with a small lecture component. This means that a lot of the learning takes place outside the class room, but each week still starts out with a lecture component to get you going. The rest of the week is dedicated to helping you in your journey of autonomous learning. Specifically, this will work as follows:

Day	**Activities**
Monday	intro to this week's topic assign chapters to read assign exercises collect previous week's homework
Wednesday	Q&A for assigned chapters
Friday	homework discussion

Expect the workload outside the class room to be much higher than usual, but on the flipside you do not even have to come to the Monday and Wednesday lectures if you have no questions

and feel confident that you can handle the chapters on your own. Only Friday lectures have mandatory attendance.

All lecture notes will be made available as PDFs in the course repository (check the `pdf` folder, feel free to ignore the rest).

3 Selected Topics

A brief selection of the topics to be covered (we will probably deviate from this order):

1. Basics of mathematics
    - Topics: sets, multisets, tuples, functions
    - Applications: bag of words model of text, n-gram models of grammaticality
2. Types of infinity
    - Topics: bijections, function inverse
    - Applications: is language infinite?
3. Relations and orders
    - Topics: properties of orders, posets, lattices, antimatroids
    - Applications: mereology, string extension learners, OT, feature systems, adjunct algebras, syntactic relations, linguistic universals
4. Graph theory
    - Topics: (un)directed graphs, connectedness, components
    - Application: parse forest representation, autosegmental phonology, AVMs, unification grammars
5. Automata theory
    - Topics: finite-state automata and transducers, regular expressions, push-down automata
    - Application: complexity of phonology & morphology VS syntax
6. Logic
    - Topics: propositional logic and first-order logic, types, lambda calculus
    - Application: semantics, model-theoretic syntax, subregular linguistics, CCG
7. Linear algebra
    - Topics: vectors and vector spaces, matrices, tensor product
    - Application: vector space semantics, spatial semantics, inflectional morphology
8. Abstract algebra
    - Topics: monoids, groups, semirings
    - Application: violation semirings in OT, semiring parsing
9. Probability theory
    - Topics: calculating probabilities with addition and multiplication
    - Application: weighted context-free grammars, corpus-based techniques
10. Information theory
    - Topics: entropy, cross-entropy
    - Application: probabilistic machine learning, surprisal for processing

4 Teaching Goals

- master essential concepts and techniques in mathematics and theoretical computer science
- apply mathematical techniques to the study of language
- formalize linguistic ideas in mathematical terms
- develop learning autonomy and the ability to expand your mathematical knowledge through self-study

5 Grading

This course can only be taken for 0 or 3 credits. Student grades are determined by the following components:

1. **Taking the initial assessment (10%)**
   At the beginning of the semester, students are asked to take a survey to assess their mathematical aptitude. Participation is worth 10 percentage points. It is perfectly alright to fail every single question, the goal is not to weed out "bad" students but to figure out the best way to pair up students in learning teams. Bring the completed survey to the last session of the first week. Performance is P/F depending on whether a filled-out survey was submitted (answering "Don't know" on each question is perfectly fine).

2. **Assignments (60%)**
   Assignments are the main component of this course and follow a specific procedure oriented around group work:

   1. *Monday: Team assignment*
      Students are matched into teams of 2 or 3, and each team is assigned a number of exercises from the lecture notes.
   2. *Friday: Homework discussion*
      Each team brings a copy of their solutions to class, and team members present their solutions on the board. We discuss key aspects of the exercises.
   3. *Next Monday: Team solutions collected* Each team hands in its solutions for all assigned exercises (not just the subset it had to prepare for the Friday session). This version is graded on a P/F basis.

3. **Final oral exam (30%)**
   At the end of the semester students take an oral exam in groups of 2 or 3. Each student gets a different exercise and must present their solution on the board while the other students ask clarification questions. The format loosely resembles that of our weekly homework discussions. Performance is evaluated based on the correctness of the solution, the clarity of presentation, and the questions asked during the other student's presentation.

6 Policies

6.1 Contacting me

- Emails should be sent to [coursenumber]@thomasgraf.net. Disregarding this policy means late replies and is a sure-fire way to get on my bad side.
- Reply time < 24h in simple cases, possibly more if meddling with bureaucracy is involved.
- If you want to come to my office hours and anticipate a longer meeting, please email me so that we can set apart enough time and avoid collisions with other students.

6.2 Student Accessibility Support Center Statement

If you have a physical, psychological, medical, or learning disability that may impact your course work, please contact the Student Accessibility Support Center, Stony Brook Union Suite 107, (631) 632-6748, or at sasc@stonybrook.edu. They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and the Student Accessibility Support Center. For procedures and information go to the following website: https://ehs.stonybrook.edu//programs/fire-safety/emergency-evacuation/evacuation-guide-disabilities and search Fire Safety and Evacuation and Disabilities.

6.3 Academic Integrity Statement

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at http://www.stonybrook.edu/commcms/academic_integrity/index.html

6.4 Critical Incident Management

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Student Conduct and Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Until/unless the latest COVID guidance is explicitly amended by SBU, during Fall 2021 "disruptive behavior" will include refusal to wear a mask during classes.

For the latest COVID guidance, please refer to: https://www.stonybrook.edu/commcms/strongertogether/latest.php