APMA4903: Seminar in Applied Mathematics, Fall 2025

Mondays and Wednesdays 11:40am-12:55am, IN PERSON, NOT ON ZOOM
428 Pupin
prof: Chris Wiggins chris.wiggins@columbia.edu
this document available online via http://bit.ly/4903info
schedule available online via http://bit.ly/f2025seminar

The seminar in applied mathematics is a required course for Columbia juniors and seniors majoring in applied mathematics, and is open only to undergraduates majoring in applied mathematics.

The senior seminar meets Monday and Wednesday and is a chance for seniors to crystallize their research interests, to begin writing about these interests, and to transition from student to creative scientists, engineers, and researchers. Seniors will themselves choose the topics to be discussed, external speakers of interest, and in the later part of the course will present, in groups of 2-4, topics of their choosing to the juniors and to their fellow seniors.

Both Juniors and Seniors can use these lectures to formulate possible research topics and postgraduate plans.

1 Grading policy

Your grade will be a function of 3 things:

1. attendance
2. an oral presentation to the juniors or to your fellow seniors on a research topic of your own choosing. (see "timeline" below for how this is submitted). Note:

- The presentation file (Google Slides or PDF preferred) should be uploaded, world-readable, with URL shared on the schedule page. Code should be linked to (Colab or GitHub repo preferred).
- Data used should also be world-readable (e.g., uploaded to Google drive so that the Colab or GitHub can be run reproducibly)

3. **There are two ways to take the seminar for 4 instead of 3 points,**

a) a *research proposal* to be turned in before **5 pm** 1 week after the last lecture. This document should be at least 5 substantive pages (e.g., not 1 page of "title page") and should summarize the field in which you are most interested, introduce a specific research question, and describe what you will do to answer it. If possible, you should include your preliminary results, data, and/or figures. You should take advantage of the oral presentation as a chance to concretize your individual research interests and you should definitely take advantage of the feedback you get from the juniors and from

your fellow seniors. All assignments must be sent *from your Columbia email to apma4903@gmail.com.*

b) Write up your results narratively either

- in a less formal style, posted on medium.com, or
- in a more formal style, posted on arxiv.org

and, in either case, include in that write-up links

- to your reproducible code as well as
- to your world-readable presentation file .

All assignments must be sent *from your Columbia email to apma4903@gmail.com.*

2 What are we learning in this class?

2.1 collaboration

Part of being a scientist, engineer, or researcher is collaboration. If/when you go to the real world, no one will mind if you got help. They will care if you understand what was requested of you and if you got the job done well. For that reason, I want you to collaborate, and your final project can be a collaborative one.

I also want you to communicate your results. That is why there will be final presentations.

However, I expect the same person-hours from everyone. If your group is twice as big, your project should be twice as impressive. Each presenter should have one `nugget` (see below)

2.2 communication

Part of being a good collaborator and a good communicator is to acknowledge the sources you used and the people who helped you. Be sure and include these acknowledgments to people who helped, and cite sources for text, images, or ideas you use.

Lastly, I want you to be able to think critically, so these projects will be in front of your peers, who will of course engage you with the spirit of participatory inquiry one expects from Columbia undergraduates.

2.3 the finiteness of time

Getting this work done, the research and the communication, on deadline, will also involve:

- sizing: you will need to be able to tell the difference between a “small”, “medium”, and “large” project

- time management: whenever your talk is, you will need to plan, break into parts, and execute on time

3 the timeline for and content of your talks

Please read this section. If you don't understand it, let me know now. After your grade has been submitted is *too late* to discuss things (see below). If you don't read this section, you will be confused when I bring up these written rules during your talk, asking you please to read this document, which will be too late to be constructive.

3.1 content

1. The first slide should contain, in addition to anything else you'd like, the names of the 1 or 2 presenters. Please start the talk by describing why you got interested in the subject.
2. The second slide should contain a list of your 1-3 primary references, including author names, years, titles, and, if from a journal, the name of the journal. If you have other references, put that on a second slide (choose the most important 1-3 or so)
3. If you used websites for research (e.g., Wikipedia), please include them among your references. Note that a reference is not a URL. No one will know what "http://doi.org/alksjdq9wue80193" means , so please use standard citations not just URLs.
4. All talks should include at least one a tech nugget (see below) for each presenter.
5. Do not show graphics taken from websites without attribution on the same slide. Do not quote extensively from websites or other sources without quotation marks and listing the source of the quote. If an unattributed quote is long enough for me to identify it by googling, it's plagiarism (see below).
6. Use complete sentences rarely on your slides, if at all. You will be speaking at the same time as you present, and an audience will be overwhelmed trying to follow you and the text at the same time.

3.2 the nugget, and the role of math

The most important part of your talk as a senior is the technical nugget. This should be a derivation or a demo of code presented along with the pseudocode.

3.2.1 demo

Code that's presented in this class should be scientific computation (not fizzbuzz or other logical but non-mathematical algorithms), using a computer to explore predictions or understand a model. Pseudocode, like derivations, present the *essential* parts, and knowing what's essential requires understanding.

Pseudocode should represent code that you will demo, i.e., run during your presentation.

3.2.2 derivation (or proof)

3.2.2.1 what is it If you're not demoing code and pseudocode, a nugget should be a proof or a derivation.

If your previous math classes only presented final results without showing how those results were obtained, the idea of a derivation may be unfamiliar. A derivation is like a map of a walk: it traces the journey from point A to point Z, highlighting only the most important turning points along the way. The goal is to show which steps are essential—not every detail—so that the key transitions are clearly understood.

In a derivation, it's important to specify

- where you are starting,
- where you want to end up, and
- what mathematical tools you'll use.

The main purpose is to emphasize the critical steps needed to move from the start to the end using algebra and manipulation of mathematical symbols. Some people misinterpret a derivation as simply quoting a series of equations from a paper, but that's not sufficient. Understanding the path—from code to pseudocode, or from a full derivation to its summary—demonstrates true understanding. If this is unclear, be sure to ask for clarification before your presentation.

3.2.2.2 infrastructure In olden times this would have been done on blackboards, in chalk. If you wish to do so you can try, but it's far easier to typeset the equations (esp. in TeX/LaTeX) and then present them digitally as part of a Google Colab, Jupyter notebook or PDF or Google Slides presentation. This also will prevent you from lecturing to a chalkboard, which mathematics professors consider an example of "a bad time."

Let me know — *before your presentation* — if this does not make sense.

3.3 timing

1. 2 weeks in advance:

(a) meet with me regarding potential topic,
(b) present an outline of what you plan to discuss
(c) identify what your primary references will likely be

2. 1 week in advance:

(a) email me *from your columbia address* and to apma4903@gmail.com a copy (preferably a google doc or google presentation file) of your presentation files or notes on what chalk you plan to present, and cc your co-presenters,

using their columbia addresses. we will then go over your talk content together (This instruction used to be 'hand me a hard copy', meaning print it out, but really who does that anymore?)

(b) edit online schedule (http://bit.ly/f2025seminar) to include your title and, if topic is not clear from your title, your topic. 2

3. before your talk: email me the soft copies of your presentation files or notes on what chalk you plan to present to apma4903@gmail.com
4. before your talk: edit the schedule page to include a link to the presentation file so that students can follow along, as well as a link to any original code (e..g, at GitHub.com).

4 Appendix: prior topics

4.1 2013

- Modeling Labor Market Dynamics
- Adaptive anonymity via b-matching
- Traffic Modeling and Analysis of Traffic Jams
- Autism & Schizophrenia
- Optimizing Peptides for Epitope-Based Vaccines
- Finite Volume Methods for the Simulation of Skeletal Muscle
- q-learning for poker
- Modern Portfolio Theory
- spring break
- guest lecture
- Electricity Pricing with Kalman Filters
- Mathematical proof of riffle card shuffling
- Evolutionary Game Theory/Replicator Dynamics
- Computer Vision and Object Recognition
- Statistical Estimation Methods: MM, ML & GMM
- Options Pricing and Black-Scholes Model
- Copulas in Finance
- Black-Litterman Model
- Using Google's PageRank Model to Rank Sports Teams
- Measuring Dispersion of Innovation for Firms
- Optimal Harvesting in Fishery Model
- Interpreting Quantum Mechanics

4.2 2014

- Modelling Cortical Synchronization on Complex Networks
- TSUNAMIs
- Ocean Tides
- Markowitz portfolio optimization and the efficient frontier
- SIR models

- Quantum Computing: Shor's Algorithm
- Partial Differential Equations
- Simulated Annealing
- Lotka-Volterra & May-Leonard Model
- General Relativity
- naive bayes
- city data
- CT
- Network Theory in Financial Systems
- Stochastic Matching Models: Equilibrium Solutions from Simulation Data
- sports statistics
- stokes flow
- Modelling Scores in the Premier League

4.3 2015

- Markowitz (portfolio optimization)
- agent-based modeling
- PageRank
- Mathematical Genetics
- fluid simulation
- rowing project/database
- benford's law
- Perceptron Algo
- Recommendation system
- Boosting and Classifier cascade (feat. Face Detection)
- k-means clustering with applications to weather data
- black scholes + the smile
- Number Theory, Codes, and Cryptography
- Neural Networks
- Kernel Methods and Their Applications
- Queueing Theory in the Labor Market
- Sabermetrics: Lineup Construction and Markov Chains
- stochastic calculus
- AI
- SIR modeling
- Value at Risk

4.4 2016

- Predicting Transiters with ANNs
- Human Detection using HOG + SVM
- Mathematical Modelling of a Hormonal Cycle
- Natural Language Processing
- Methods in coevolution analysis
- Branching Process Models of Cancer

- quantum computation
- Consensus in Collective Dynamics
- Probabilistic Topic Models and User Behavior
- Riemann problem and application with Shallow Water Equations
- Modelling HIV using the SIR model
- Text Compression, Huffman Coding and PPM Models
- The Traveling Salesman Problem
- Road Traffic Modeling with PDEs and Cellular Automata
- How to Lose Less $$$ in Blackjack
- Mathematical Modeling of How We Hear
- Markowitz Portfolio Optimization
- Mathematics of Fantasy Football
- Statistics and Machine Learning for Precision Medicine
- Markov Chain Monte Carlo w Application to Cryptography

4.5 2017

- American Options Pricing: Black Scholes vs. models that incorporate early exercise
- BiTcOiN
- machine learning, computer vision, and open source
- Control Theory
- Dynamic Stochastic General Equilibrium: Modeling Real Business Cycles
- Applying a Cellular Automata Model to the Study of Gentrification and Social Migration Within Neighborhoods
- Hidden Markov Models in Genomics
- Poker and Machine Learning
- Matrix factorization models in recommender systems
- Assessment of GWO predictability for tornado frequency
- NLP & valuing companies
- Evolutionary Computation and Symbolic Regression
- Application of game theroy model on environmental accidents and subsequent apologies
- Modern Portfolio Theory
- Rationality assessment and decision making
- Predictive analytics in soccer
- Modeling Terrorism
- Pricing and the Newsvendor Model
- Efficient query into database using k-regret

4.6 2018

- What basic statistical modeling reveals about the microstructure of the US stock market
- Tropical cyclone risk assessment in a changing climate
- Holdout Randomization Tests

- META-LEARNING
- ML for dota2
- Tropical Cyclone Prediction Model
- Equity Portfolio Optimization
- Derivation of one-dimensional model of blood flow
- neutrino oscillations and unknown dark matter
- real estate machine learning
- Quantum Computing, specifically focusing on Shor's algorithm
- h1b dataset: predicting fails
- stroke prediction ml
- developing an algorithmic trading strategy in the cryptocurrency market
- predicting revenue of movies
- black scholes and the smile
- multi-arm bandits problem
- pairs trading - Kalman filtering and its application
- game theory
- Predicting Yelp Stars from Yelp Reviews
- D'Alembert's Paradox and Flight (Application to Wing Configurations)
- Hyperbolic Geometry in Machine Learning

4.7 2019

- census 2020 and GIS
- computer vision
- math + biz/ $
- prediction markets
- pokerbots
- deep nets and PDE
- Who Responds to the Census?
- bio and image segmentation
- hyperbolic space embeddings
- Differential Privacy + Census
- sports analytics
- cyclones
- Curve Shortening Flow
- Libraries and The Census
- Census and Gentrification
- style in NYT articles; number of followers of authors
- pairs trading on quantopian
- "iterative proportional fitting and census data"
- opioid
- TrueSkill Rating System
- NFL Elo Rankings
- Effect of Market disagreement on CAPM
- "predictive policing"

4.8 2020

- Using finite elements to model ice flow
- Application of Kalman Filter in Finance
- Understanding Pregnancy Predition from a Self-Reporting Women's Health Application
- Reinforcement Learning for No Limit Texas Hold 'em
- "Fokker-Planck Equation and its application to Ornstein-Uhlenbeck"
- Incorporating Reddit Sentiment Data into Portfolio Optimization
- Analyzing initial transmission of COVID-19 through a Bayesian model
- SEIR Model to predict COVID 19
- Extreme weather forecasting/modeling
- "Merton & KMV Modeling"
- Applying Elo Ratings to Squash
- Defenses Against Physical Adversarial Attack
- Shot selection in the NBA
- Modeling Language Competition with Dynamical Systems
- Quantum Supremacy: Breaking Cryptography with Quantum Computers
- Urban Data Science
- (Covid) testing regime simulations
- machine learning+music
- Volatility Smile & Jump-Diffusion Processes
- Bayesian neural net

4.9 2021

- Carbon Capture by Adsorption
- Revising REVISE: Analysing Bias in Computer Vision Datasets
- Foundations of Probability Theory
- Valued & policy based reinforcement learning on multiple stock trading
- Optimizing portfolios incorporating investor's ESG ethical considerations
- Analyzing the Runner on Second Rule with Markov Chains and a Logistic Classifier
- The Mathematics of Generating Music
- The Theory and Practice of Modeling Extreme Precipitation Events
- Natural Gradient
- Forecasting Elections: Polls & Betting Market
- Valuation of Amazon Stock Price by DCF Model, MCS Simulation and BS Model
- Stock Prediction Using LSTM and GRU Models
- Discrete Hazard Modelling of Eviction Risk
- Heston Model
- Pseudorandom Keystream Generation Using Cellular Automata
- The Mathematics of Bitcoin Mining
- Modeling Gender Bias
- Applying Mean Field Game Theory to Human Migration

- predicting MLB game outcomes
- federated machine learning (slides)

4.10 2022

- Fintech, blockchain, and crypto
- Structured State Spaces (S4) and Applications
- Markov Chains to Evaluate Strategies in Baseball and Tennis
- Speculative bubbles in cryptocurrencies
- Exoplanet, Goldilocks Zone, Machine Learning Atm.
- Building a Recommender System for Podcasts
- Options Pricing: Beyond the Gaussian
- Computer Vision in Medical Imaging
- ML in Private Equity Fund Selection
- Heisenberg Uncertainty Principle
- Atlantic Meridional Overturning Circulation (AMOC) Tipping Point
- Adomian Decomposition methods for first order nonlinear PDE
- Blackjack Analysis
- 2D Shallow Water Equations (Link to GitHub at the last slide)
- Gerrymandering Detection with Markov Chain Monte Carlo
- Applications of the Coordinate Ascent Variational Inference Algorithm
- ML Meets Systems Biology: Modeling Glucose-Insulin Response
- Coding Theory: Decoding Methods and Applications
- "Introduction to Spectral Clustering Algorithms"
- Introduction to Autoencoder and SELU
- Modeling Migration
- Variants of gradient descent optimization in neural networks (with breast cancer diagnostic data)

4.11 2023

- data integrity, hamming codes, sodoku
- Image Processing: using PDE based models for image restoration, image denoising
- Climate Migration in the U.S. - Modeling "Climate Destinations"
- Cloaking via Change of Variables in Electrostatics and Electrodynamics
- Climate Dynamics' Effects on Financial Systems
- The Invasion of the Spotted Lantern Fly
- Modeling Food Waste
- MLB win percentage prediction for starting pitchers
- Auction Design
- Gradient Descent and other training methods for neural network
- Predicting the Soccer Transfer Market
- Energy Management of a Migrogrid with Flexible Demand
- AES-RSA Hybrid Encryption System
- spectral methods for PDEs of geophysical fluid motion

- Math of Finance
- Math of Music Recommendation Systems
- Using Machine Learning for Medical Diagnosis: Neurodegenerative Disease in the Gut Microbiome
- Markov Chains, Machine Learning, and Tennis
- Feasibility and spatial heterogeneity of lineage conversion processes

4.12 2024

- Modeling Chromatin Regulatory Effects with Variational Inference
- Simulation of Discharge Initiation with Runaway Electrons in DIII-D
- Expensive Coffee Incentive: Non-cooperative games and Optimal Pricing Strategy
- Schramm-Loewner Evolution
- Singular Structures in Bar Frameworks
- ML Techniques in Silent Speech Interfaces
- Why does the equatorial ocean flow upwards?
- Event Driven Stock Prediction
- Aviation & Fluid Dynamics
- Transit Dynamics
- Optimizing Air, Land, and Sea
- Online Poker and Portfolio Optimization
- LLMs for Literature Analysis
- Drug Sensitivity Modeling
- Markov Chain Traffic Assignment and Optimization
- Predicting Premier League Outcomes
- Kolmogorov Arnold Networks for Options Pricing
- Using Alpha-Beta Pruning and Deep Reinforcement Learning to solve 4x4 Minichess
- Queueing Theory Applications for Improving Mental Health Call Centers
- Recommendation Systems for Letterboxd
- Combinatorial Game Theory for Go Endgame
- Neural Network for nanomaterial flake searching

5 legal disclaimers:

5.1 grading

Please note that the Faculty Handbook makes explicit constraints on what I can and can not do when it comes to grading. Although it is precedented to make unusual arrangements for a student before a grade is assigned, e.g., for health reasons or other extenuating circumstances with approval of a dean, "Once qualitative grades are given, they may not be improved through the submission of additional work by the student." "Changes to final grades normally may be made only to correct an error.... The grounds for changing a disputed grade

are very limited."

5.2 plagiarism

https://bulletin.columbia.edu/columbia-engineering/university-school-policies-procedures-regulations/academic-integrity-discipline/ makes clear Columbia's (and therefore my) policies regarding plagiarism, singling out the use of online material:

> One of the most prevalent forms of plagiarism involves students using information from the Internet without proper citation. While the Internet can provide a wealth of information, sources obtained from the Web must be properly cited just like any other source. If you are uncertain how to properly cite a source of information that is not your own, whether from the Internet or elsewhere, it is critical that you do not hand in your work until you have learned the proper way to use in-text references, footnotes, and bibliographies. Faculty members are available to help as questions arise about proper citations, references, and the appropriateness of group work on assignments. You can also check with the Undergraduate Writing Program. Ignorance of proper citation methods does not exonerate one from responsibility.

The site also makes clear what sanctions will result.

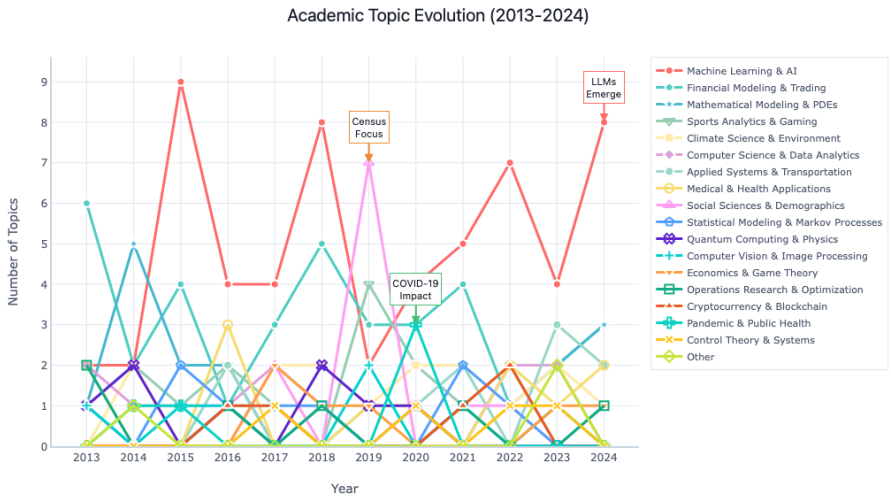
5.3 Generative AI

The provost https://provost.columbia.edu/content/office-senior-vice-provost/ai-policy has set a policy:

> Disclose the use of Generative AI tools: Columbia community members who leverage Generative AI to produce any written materials or other work product must disclose that those materials and that work product is based on or derives from the use of Generative AI. Always be transparent if you are relying on the output of a Generative AI tool.

For example: this chart, along with this online navigator of prior topics from senior presentations:

- chart:



- site: https://apma4903.github.io/topic-evolution/

6 logistics and laptops

- I used to sit in the back of the room, emailing students I see on Facebook during class. But now I try to sit near the front, off to the side. If you see me using my laptop rest assured it'll be to research the article being presented, or a Wikipedia page on the technique being discussed. I encourage you to do the same, but please sit in the back if you're using a laptop so as not to distract students who are paying attention.

7 links

If you want to look back at the links from day one of the class, you can find them here:

- instructions, 4901: http://bit.ly/4901info
- instructions, 4903: http://bit.ly/4903info
- day 1 presentation file: http://bit.ly/4901day1
- 'good talk': http://bit.ly/exampletalk

also:

- 4901/3 anonymous survey: http://bit.ly/4901f2025s (complete asap!)

8 Canvas

There is no canvas for this class. This instruction document is canonical

9 Coursworks

There is no courseworks for this class. This instruction document is canonical

10 Slack

I actually made a Slack for this class! You are required to join it but you are not required to look at it since the canonical information is all in this document. Here it is! https://apma2025.slack.com/