CPSC 7430: Topics in flow-based generative modeling and optimal transport

Aram-Alexandre Pooladian

Fall 2026

E-mail: aram-alexandre.pooladian@yale.edu
Web: arampooladian.com
Office hours: Thu. 11h00 to 12h00, KT 1321
Class time: Mon. 9h25 to 11h20
Classroom: 17 Hillhouse, room 115

Course description

Generative modeling has had a major impact across many scientific disciplines. A driving force of this change has been the development of scalable algorithms that transport, or "flow," simple reference distributions into complex data distributions. This course will discuss the mathematical foundations of flow-based generative models and optimal transport, with the goal of understanding benefits and limitations of various procedures. Core topics include optimal transport maps, entropic optimal transport, flow matching and stochastic interpolants, diffusion models, Schrödinger bridges, temporal data, and data on Riemannian manifolds. In addition to discussing architectural aspects of mondern generative modeling (such as UNets, Vision transformers, VAEs), we will also connect these tools to fine-tuning, stochastic optimal control, approximate inference, and inverse problems, and conditional generation.

Prerequisites

Mathematical maturity is necessary to enroll in the course. A working understanding of the following courses is also recommended:

- **Advanced probability:** S&DS 400, S&DS 600, or MATH 330;
- **Optimization:** S&DS 431, S&DS 631, S&DS 432, or S&DS 632;
- **Stochastic processes:** S&DS 351 or S&DS 551.

Prior exposure to neural networks, optimal transport, or generative modeling is not required. Enrollment is limited and requires the instructor's permission.

Learning objectives

By the end of the course, students will be able to:

1. understand basic definitions pertaining to optimal transport maps, entropic optimal transport, and their computational and mathematical differences;

2. understand the difference between deterministic and stochastic measure transport, and describe the similarities and differences of various generative modeling algorithms (such as stochastic interpolants, flow matching, Schrödinger bridges, etc.);

3. implement algorithms discussed in the course at small or larger scales;

4. connect generative modeling to broader areas of research such as stochastic control, approximate inference, and in specialized instances such as when the data is temporal in nature, manifold data, and for inverse problems;

5. read, present, and critically assess recent research in the area.

Course materials and format

The primary reference is the evolving course notes. Additional readings will draw from recent research papers, especially during the second half of the term.

YCRC resources will be available for course-related experiments.

Course outline

I. Foundations of optimal transport

1. **Defining the optimal transport problem**
    - Definition of transport maps and plans
    - The Monge problem, Kantorovich relaxation, and duality
    - The 2-Wasserstein distance and its semi-dual formulation
2. **Brenier maps, univariate and Gaussian OT**
    - Brenier's theorem and optimal transport paths.
    - Univariate and Gaussian optimal transport maps
3. **Stability, statistics, and basic OT algorithms**
    - Stability results for optimal transport maps
    - Statistical estimation of Wasserstein distances
    - Map estimation

4. **Entropic optimal transport**
   - Primal and dual entropic OT
   - Sinkhorn's algorithm and convergence guarantees
5. **Entropic maps and regularization limits**
   - Entropic Brenier map
   - Approximations to unregularized OT
   - Implementation of Sinkhorn's algorithm

II. Introduction to dynamic transport

6. **Flow-based transport**
   - Deterministic and stochastic flows; continuity and Fokker–Planck equations
   - Brownian motion, Itô processes, forward and reverse SDEs, Girsanov's theorem, and Doob transforms.
7. **Dynamic optimal transport and Wasserstein geometry**
   - Benamou–Brenier and dynamic dual formulation of optimal transport
   - Primer on Otto calculus.

III. Flow matching and diffusion models

8. Interlude: Modern architectures for generative modeling
   - Variational autoencoders
   - UNets and ViTs
   - Implicit biases of architectures
   - Basics of conditional generation (see later)
9. **Neural ODEs and flow matching**
   - Stochastic interpolants (and flow matching) derivation and pseudocode
   - Mini-batch sampling, Riemannian flow matching, and Wasserstein flow matching.
10. **Diffusion models**
    - Derivation of score-matching via forward and reverse Ornstein–Uhlenbeck processes
    - Theoretical properties of the reverse heat-flow map
    - DDPM/DDIM sampling
    - Blind denoising diffusion models
11. **Schrödinger bridges**
    - Path-space formulation and connections with entropic optimal transport
    - Iterative Markovian Fitting and convergence guarantees

IV. Advanced topics

12. **Fine-tuning**
    - Connections to stochastic optimal control
13. **Temporal snapshots and acceleration matching**
    - Acceleration-matching derivation and implementation.
14. **Few-step and one-step generative algorithms**
    - Flow-map matching and distillation
    - Generative drifting models
15. **Conditional generation and inverse problems**
    - Conditional Brenier maps and estimators for conditional simulation
    - Classifier-free guidance
    - Large-scale inverse problems (e.g., diffusion posterior sampling)

Assessments and grading

Bi-weekly quizzes (15-30%)

Short quizzes (5–10 minutes long) will be given roughly every two weeks. They will assess core definitions, mathematical formulations, derivations, and conceptual connections from the lectures and assigned readings.

Final Course Project (70-85%)

This can either be a solo project or in groups of 2. The "difficulty" of the project is at the students' discretion (to be decided by **October 15**). In all instances, I expect the students to communicate progress with me as the semester unfolds.[1]

Option 1: Carefully summarize an emerging research area. This can encompass anywhere from 1-3 papers or 5-10 papers. Students will be largely evaluated on clarity of exposition and correctness with respect to the larger literature. Students are expected to explain the main ideas of the chosen papers in their own words and also critically assess how they relate to the broader literature.

Option 2: Re-implement existing methods in generative modeling. For example, a public repository for training a diffusion model from scratch, accompanied by applications in fine-tuning or solving inverse problems. Exact reproduction of published results is not required. Instead, the goal is to produce a working repository that others can learn from.

This option would be graded based on readability of code (i.e., a clear README, easy-to-use notebooks) and ability to easily execute the code and traverse through the files.

[1] Note that there is no mandatory in-class presentation, though students are welcome to take class time to present their projects at the end of the semester if they are interested.

Option 3: Write an original research paper. The goal of this project is to work towards something that might be suitable for publication at a top conference (in either NeurIPS, ICML, ICLR, AISTATS, ALT, COLT, etc.). Papers can be theoretical or empirical (or a mix of the two options) in nature.

This option is of course the most ambitious and also the hardest to grade. Students pursuing this type of project are encouraged to start as early as possible and should expect to update me every two to three weeks. This is to ensure that the final project is not rushed in the last week of class.

Finall, I reserve the right to curve the scale based on overall class scores at the end of the semester; any curve will only make it easier to obtain a given letter grade.

Course Policies

During class

To encourage class discussion and participation, laptop use will only be permitted when we are reading a paper together (with the exception of accommodations requests).

Policy on AI use

Students are allowed (and encouraged) to use AI tools for their final projects. However, the goal of the final project is for students to gain a deep understanding of their project (either summarization, implementation, or research project), and assume *ownership* of whatever task they choose. In particular, if you use AI, then you are to be held to a higher standard. A standard AI disclosure statement is expected (what and how AI was used).

Academic integrity and collaboration

Students must comply with the Yale Graduate School Personal Conduct and Academic Integrity Standards. Project discussions are encouraged, but submitted work must accurately represent the group's contributions and cite all sources.

Accessibility and accommodations

Students needing disability-related accommodations should contact Yale Student Accessibility Services as early as possible. Approved accommodations will be implemented with the appropriate university office.