Mathematics of Reinforcement Learning: Towards a PDE-Based Approach Fall 2026

Maria Han Veiga, Vlad Kobzar

March 26, 2026

Reinforcement learning is one of the main machine learning paradigms with extensive industrial applications, ranging from state-of-the-art large language models, robotics and self-driving cars to personalized medicine, digital health, recommender systems and online marketing and advertising[1, 2]. This paradigm blends diverse strands of ideas about reward maximization by trial and error in behavioral psychology, design of clinical trials, Monte-Carlo and other sampling methods going back to the first half of the 20th century[3].

The first part of this course provides a self-contained introduction of reinforcement learning and bandit problems, focusing on open problems and existing classical mathematical results. Then, this graduate course explores applications of PDE methods to study reinforcement learning and related sequential decision problems. Although many of these problems can be characterized by discrete-time dynamic programs, such programs are typically computationally and analytically intractable. Conceptually our approach is "numerical analysis in reverse": we view a discrete-time dynamic program as a numerical PDE scheme and find the PDE that is discretized by it.

The concrete objective of this course is to cover two recent results and focus on related open problems. The first result *A PDE-based analysis of bandit learning with policy gradients* (working title) by the instructors characterizes smooth bandit policies by a degenerate hypoelliptic Fokker Planck equation. The second result *Discrepancy minimization and mean curvature flow* (working title) by Vlad with Nestor Guillen characterizes the *combinatorial discrepancy minimization* by evolution of level sets under a mean curvature flow (this problem is closely related to quasi-Monte Carlo). Important tools in the above will be (a) the linear parabolic regularity theory, particularly hypoellipticity, and (b) viscosity theory of the surface evolution equations. Much of this theory has been developed in the context of mathematical physics and geometry in the second part of the 20th century. The adaptation of this theory to the present setting is interdisciplinary, blending in classical statistics and probability with recent advances in learning theory. All the relevant tools we will be studying have been crucial to progress throughout mathematical sciences beyond our specific motivating applications. As such this course should be beneficial to students with a broad range of interests.

The instructors will provide self-contained notes with a list of open problems. A provisional list of topics/references includes:

- *Background on reinforcement and bandit learning*
  - From prediction to actions: sequential decision making
  - Interaction with environment: incomplete information and exploration-exploitation trade-offs
  - The bandit problem, Markov decision processes and the policy gradient theorem
- *Classical results in reinforcement and bandit learning*
  - Introduction to optimization techniques
  - Policy Iteration methods
  - Convergence results for Policy Gradient
  - Decision trees and Monte Carlo
- *Degenerate Fokker-Planck equation, hypoellipticity, the Hormander bracket condition, and connections to smooth bandit policies*
  - A PDE-based analysis of bandit learning with policy gradients (working title), M. Han Vega, V.K.
  - Hypoelliptic Fokker-Planck characterization of Thompson Sampling (working title), T. Akhunov, Q. Du, D. Lin, V.K.
- *Mean curvature flow and its connections to combinatorial discrepancy and Blackwell approachability in statistics*

- Blackwell, An analogue of minmax theorem for vector payoffs, Pacific J. Math, 1956,
- Kohn, Serfaty, A deterministic-control-based approach to motion by curvature, Comm. Pure Appl. Math. (2006),
- Abernethy, Bartlett and Hazan, Blackwell Approachability and No-Regret Learning are Equivalent, Proceedings of the 24th Annual Conference on Learning Theory (COLT), 2011,
- Discrepancy minimization and mean curvature flow (working title), N. Guillen, V.A.K.

- *Optimal stopping / parabolic free boundary problem, exploration-exploitation trade-off*
  - A. Bather and G. Simons. The minimax risk for clinical trials, J. R. Statist. Soc. B, 1985.
  - Lai and Lim, Optimal stopping for brownian motion with applications to sequential analysis and option pricing. Journal of Statistical Planning and Inference, Herman Chernoff: Eightieth Birthday Felicitation Volume (2005).
  - Harvey, Liaw, Perkins, and Randhawa, Optimal anytime regret with two experts, IEEE 61st Annual Symposium on Foundations of Computer Science, IEEE, 2020.

- The Bellman function method of Burkholder and the Monge-Ampere equations
  - Foster, Rakhlin and Sridharan, Online Learning: Sufficient Statistics and the Burkholder Method, Proceedings of the 31st Conference On Learning Theory (COLT), 2018
  - Vasyunin and Volberg, Burkholder's Function via Monge-Ampere Equation, Illinois Journal of Mathematics (2010)

Tentative schedule

Week 1: Introduction to reinforcement learning

Week 2: Prediction, sampling, and sequential decision making / exploration-exploitation trade-offs

Week 3: Markov decision processes and bandit problem

Week 4: Policy iteration methods, actor-critic methods

Week 5: Policy gradient methods and policy gradient theorem

Week 6: Classical convergence results, regret minimization framework

Week 7: Modern convergence results for policy gradient

Week 8: Hypoellipticity and continuous limit of smooth bandit problems

Week 9: Quasi-Monte Carlo, and continuum limit of combinatorial discrepancy

Week 10: Blackwell approachability theorem and connections to regret minimization

Week 11: Dynamic programming and connections to optimal stopping and exploration/exploitation trade-offs

Week 12: Bellman function method of Burkholder and connections to Monge-Ampere

Week 13: Student presentations

Week 14: Student presentations

Prerequisites: No machine learning or computing experience is required. Some familiarity with the basics of real analysis and PDEs at the level taught in an introductory graduate course is expected, but this course will be intended to be as self-contained as possible.

Participation: The class will be run more like a seminar than a lecture; regular attendance, as well as collegial and dynamic participation in discussions is expected. There will be opportunities for students to conduct independent research and present it to the class.

Grading: The grade in the course will be based on attendance and participation, which may include regular class discussions and/or presentation of papers or research projects on relevant topics of interest.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[2] T. Lattimore and C. Szepesvari, *Bandit Algorithms.* Cambridge University Press, 2020.

[3] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25(3–4), p. 285–294, 1933.

Statements required by College of Arts and Sciences Ad Hoc Committee on State and Federal Policy, and the College of Arts and Sciences Curriculum Committee

Statement on intellectual diversity: Ohio State is committed to fostering a culture of open inquiry and intellectual diversity within the classroom. This course will cover a range of information and may include discussions or debates about controversial issues, beliefs, or policies. Any such discussions and debates are intended to support understanding of the approved curriculum and relevant course objectives rather than promote any specific point of view. Students will be assessed on principles applicable to the field of study and the content covered in the course. Preparing students for citizenship includes helping them develop critical thinking skills that will allow them to reach their own conclusions regarding complex or controversial matters.

Statement on academic misconduct: It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-48.7 (B)). For additional information, see the Code of Student Conduct.

Statement on religious accommodations: Ohio State has had a longstanding practice of making reasonable academic accommodations for students' religious beliefs and practices in accordance with applicable law. In 2023, Ohio State updated its practice to align with new state legislation. Under this new provision, students must be in early communication with their instructors regarding any known accommodation requests for religious beliefs and practices, providing notice of specific dates for which they request alternative accommodations within 14 days after the first instructional day of the course. Instructors in turn shall not question the sincerity of a student's religious or spiritual belief system in reviewing such requests and shall keep requests for accommodations confidential.

With sufficient notice, instructors will provide students with reasonable alternative accommodations with regard to examinations and other academic requirements with respect to students' sincerely held religious beliefs and practices by allowing up to three absences each semester for the student to attend or participate in religious activities. Examples of religious accommodations can include, but are not limited to, rescheduling an exam, altering the time of a student's presentation, allowing make-up assignments to substitute for missed class work, or flexibility in due dates or research responsibilities. If concerns arise about a requested accommodation, instructors are to consult their tenure initiating unit head for assistance.

A student's request for time off shall be provided if the student's sincerely held religious belief or practice severely affects the student's ability to take an exam or meet an academic requirement and the student has notified their instructor, in writing during the first 14 days after the course begins, of the date of each absence. Although students are required to provide notice within the first 14 days after a course begins, instructors are strongly encouraged to work with the student to provide a reasonable accommodation if a request is made outside the notice period. A student may not be penalized for an absence approved under this policy.

If students have questions or disputes related to academic accommodations, they should contact their course instructor, and then their department or college office. For questions or to report discrimination or harassment based on religion, individuals should contact the Civil Rights Compliance Office. (Policy: Religious Holidays, Holy Days and Observances)

Statement on disability services: The university strives to maintain a healthy and accessible environment to support student learning in and out of the classroom. If you anticipate or experience academic barriers based on your disability (including mental health, chronic, or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. If you are ill and need to miss class, including if you are staying home and away from others while experiencing symptoms of a viral infection or fever, please let me know immediately. In cases where illness interacts with an underlying medical condition, please consult with Student Life Disability Services to request reasonable accommodations. You can connect with them at slds@osu.edu; 614-292-3307; or slds.osu.edu.