

OPTIMIZATION FOR AI

INTRODUCTION

Dongyoung Lim
UNIST

AI51101, IE55101

COURSE INFORMATION

- ▶ Course Title: **Optimization for AI (AI51101, IE55301)**
- ▶ Instructor:
 - Dong-Young Lim (Industrial Engineering & AI Graduate School, UNIST)
- ▶ Teaching Assistants:
 - Jonghun Lee (jh.lee@unist.ac.kr)
 - Jeongsik Yun (jeongsik@unist.ac.kr)

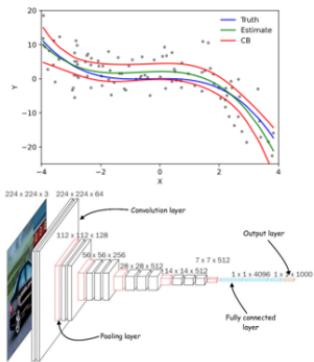
OFFICE HOURS

- ▶ Time: **Wednesday, 1:00 – 2:30 PM**
- ▶ Office: **Room 301-12, Building 112**
- ▶ You are welcome to visit during office hours without an appointment.
- ▶ Outside of these hours, please contact me via email to arrange a meeting.

MOTIVATION

- ▶ Why Optimization for AI?
 - Optimization is the mathematical foundation of AI and ML.
 - Every learning algorithm involves solving an optimization problem.
 - Advances in optimization have enabled deep learning, diffusion models, reinforcement learning, and large-scale AI systems.
 - This course bridges mathematical theory, algorithm design, and applications in AI.

MOTIVATION



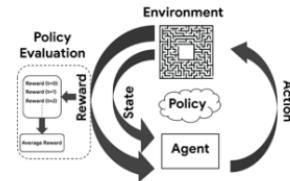
Supervised Learning

$$\min_{\theta} \frac{1}{n} \mathcal{L}_{\theta}(x_i, y_i)$$



Unsupervised Learning

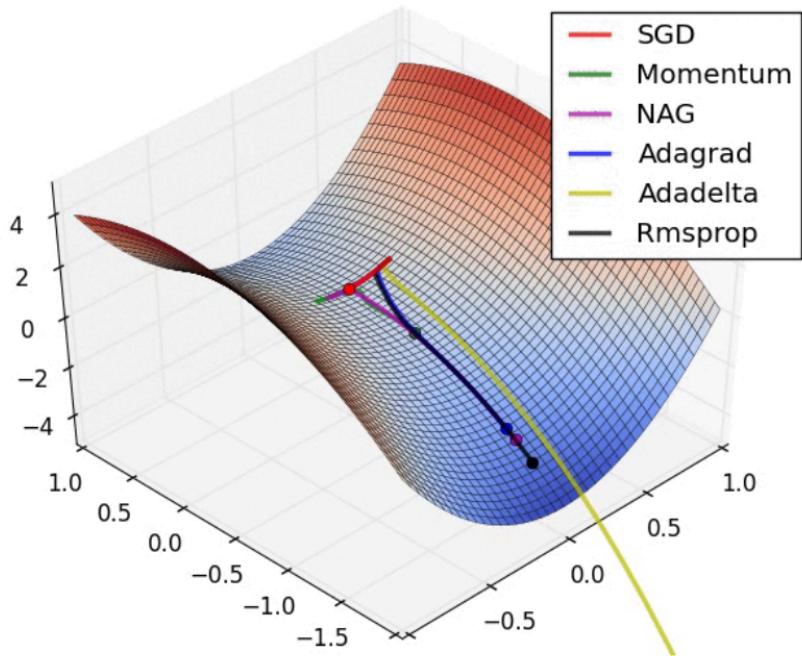
$$\min_{\theta} \frac{1}{n} \mathcal{L}_{\theta}(x_i)$$



Reinforcement Learning

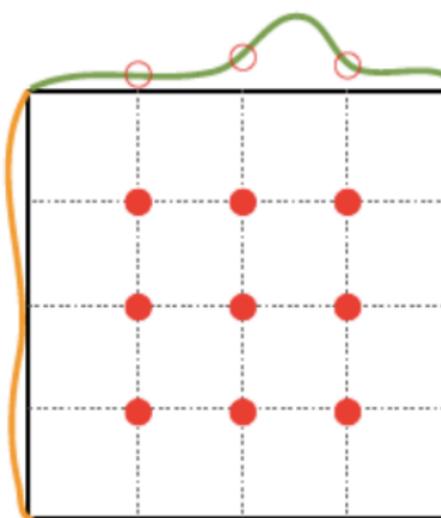
$$\max_{\theta} \mathbb{E}_{\tau \sim p_{\theta}} [\mathcal{R}_{total}(\tau)]$$

MOTIVATION



MOTIVATION

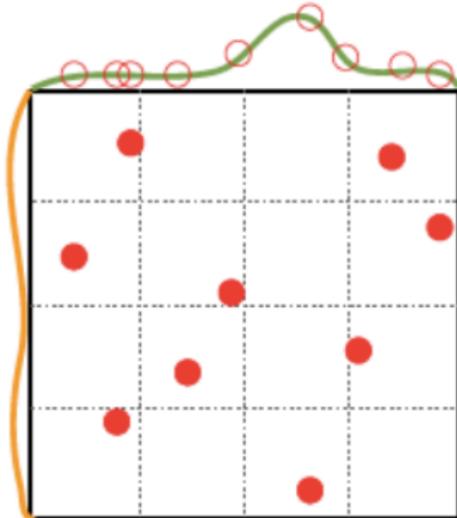
Grid Layout



Important Parameter

Unimportant Parameter

Random Layout



Important Parameter

Unimportant Parameter

PREREQUISITES

- ▶ **Required background knowledge:**
 - Calculus (derivatives, gradients, Jacobians, Hessians, Taylor expansion, ...)
 - Linear Algebra (vector spaces, span, inner products, orthogonality, bases, ...)
 - Real Analysis (metric and normed spaces, convergence, Cauchy sequences, ...)
 - Probability (random variables, expectations, variances, inequalities, ...)
 - Convex Optimization (fundamental concepts and basic algorithms)
- ▶ This course will provide only a *brief review* of real analysis, calculus, linear algebra, and convex optimization.
- ▶ Students are expected to already have working knowledge of these areas.

TEXTBOOK AND REFERENCES

- ▶ **Main material:**
 - Lecture slides are self-contained and will serve as the main textbook.
 - All materials will be available at the course homepage.
- ▶ **References:**
 1. *Convex Optimization* by Stephen Boyd and Lieven Vandenberghe
 2. *Convex Optimization: Algorithms and Complexity* by Sébastien Bubeck
 3. *Introduction to Online Convex Optimization* by Elad Hazan
 4. *Nonconvex Optimization for Machine Learning* by Prateek Jain
- ▶ Students can download lecture slides and communicate with the instructor and TAs through blackboard.unist.ac.kr.

COURSE STRUCTURE

► Phase I: Lectures

- Core topics in optimization:
 - ▶ Convex Optimization
 - ▶ Gradient Descent
 - ▶ Stochastic Gradient Descent
 - ▶ Advanced Optimization Algorithms
 - ▶ Fundamental Analysis Tools for Optimizers
- Fundamental theoretical analysis of these methods

► Phase II: Journal Club and Projects

- Weekly reading and in-depth review of recent optimization papers
- Team projects with presentations

PHASE I: LECTURES

Chapter	Topic
0	Preliminaries
1	Convex Optimization
2	Stochastic Gradient Descent
3	Regret Analysis
4	Nonconvex Optimization
5	PAC-Bayes Bounds

- ▶ This outline is based on last year's lecture notes.
- ▶ For this year, several lectures will be removed, and more time will be dedicated to Phase II (Journal Club and Projects).

PHASE II: PAPER REVIEW (JOURNAL CLUB STYLE)

- ▶ Each team consists of up to **two members**.
- ▶ Teams select one **recent paper** (2022–2025) on optimization algorithms published in top AI venues such as *JMLR*, *SIMODS*, *ICML*, *ICLR*, *NeurIPS*. You may also choose from the top venues in your own field (e.g., *CVPR*, *EMNLP*, etc.).
- ▶ Each team gives a **18-minute presentation** including Q&A.
- ▶ The presentation should cover:
 1. **Motivation** – Why is this problem important? Why should we care about this paper?
 2. **Approach** – What methods or algorithms are proposed in the paper?
 3. **Main Results** – What are the key theoretical findings or experimental results?
 4. **Critical Review** – Strengths, weaknesses, and possible future directions.
- ▶ Focus on delivering a **clear and critical review**, not just a summary.

PHASE II: TEAM PROJECT — OVERVIEW

- ▶ Goal: Conduct an in-depth literature review, implement key methods, run carefully controlled experiments, and deliver an in-depth analysis.
- ▶ Scope: Choose one topic in optimization for AI and study it end-to-end (theory → algorithms → empirical evaluation).
- ▶ Constraints: Ensure **fair comparisons** under the same compute budget and report all experimental settings for **reproducibility**.

PHASE II: TEAM PROJECT — DELIVERABLES

- ▶ Deliverables:
 - **Report** (7 pages): problem setup, methods, experimental design, results, analysis, limitations, takeaways.
 - **Code repository**: clean, documented, reproducible (scripts + config + README) via **GitHub**.
 - **Presentation**: 25 minutes including Q&A.
- ▶ After selecting a topic, your project should include:
 - Survey of **recent methods, applications, theoretical insights, and limitations**
 - Careful **implementation** of key variants or algorithms
 - **Comparison and analysis** of these methods under the same experimental protocol
 - Experiments conducted in **fair and reproducible settings**
 - In-depth discussion of strengths, weaknesses, and potential future directions

PHASE II: TEAM PROJECT — EXAMPLE TOPICS

► Optimization for Hyperparameter Tuning

- How do grid/random search and optimization-based methods (Bayesian optimization, Hyperband/ASHA, gradient-based HPO, etc.) compare under the same compute budget?
- What are the strengths and weaknesses of each method in terms of sample efficiency, stability, and reproducibility?

► Variants of Adam

- How do Adam variants (AdamW, AMSGrad, AdaBelief, Adafactor, LAMB, etc.) perform on modern benchmarks?
- Do these variants improve convergence speed, stability, or generalization under equal compute settings?
- In which scenarios does one variant outperform the others?

PHASE II: TEAM PROJECT — EXAMPLE TOPICS

► Variants of SAM

- Compare the performance of SAM variants on various datasets
- Does the flat minima hypothesis hold in empirical evaluations?
- How do these methods affect sharpness, robustness, and calibration on benchmark datasets?
- What are the limitations or open challenges in applying SAM variants?

► Zero-Order Optimization

- Can zero-order methods (random search, SPSA, NES, CMA-ES, etc.) be competitive with gradient-based optimization?
- How efficient are they in terms of query complexity and scalability?
- Review recent applications of zeroth-order optimizations. In what settings (e.g., black-box LLM tuning) do zero-order methods provide unique advantages?
- What limitations and bottlenecks remain unsolved?

PHASE II: TEAM PROJECT — EXAMPLE TOPICS

► **Second-Order Optimization**

- Are second-order methods (Newton, quasi-Newton, K-FAC, Shampoo, etc.) still impractical due to computation and memory cost?
- With modern hardware, when do they become feasible and effective?
- Do second-order methods truly suffer from overfitting compared to first-order methods?
- What do recent studies reveal about their potential applications?

► **LoRA for Foundation Models**

- How does LoRA enable parameter-efficient fine-tuning of foundation models?
- What are the advantages and limitations of LoRA compared to full fine-tuning or other PEFT methods?
- How do LoRA variants perform across tasks with different data sizes and domains?
- What practical guidelines (rank selection, module choice, deployment) emerge from recent studies?

GRADING POLICY

Exam	Homework	Paper Review Presentation	Team project	Attendance
30%	15%	15%	40%	0%

- ▶ (Important) As per the UNIST Academic Regulations, Article 29, a student shall attend 75% or more of the total class hours. That is, student absenting for more than 25% of the total class hours will fail the class.
- ▶ Use the MOBILE ATTENDANCE SYSTEM to check your attendance by yourself at every class.

CLASS SCHEDULE AND CANCELLATIONS

- ▶ Planned Cancellation Dates: September 3, September 29, October 1, October 29, November 10, November 12
- ▶ For some cancelled sessions, **pre-recorded online lecture materials** will be uploaded.
- ▶ Please check the course **Blackboard** regularly for updates and announcements.