

Minhak Song

Undergraduate Student, KAIST

Personal website: songminhak.github.io Contact: minhaksong@kaist.ac.kr

Research Interests

Theoretical Foundations of Modern Machine Learning, Optimization, Statistics

Education

Korea Advanced Institute of Science and Technology (KAIST) , Daejeon, South Korea B.S. in Industrial and Systems Engineering & Mathematical Sciences (double major) > Tuition and stipend fully covered by National Presidential Science Scholarship. > Completed 2 years of mandatory alternative military service (02/2023 – 11/2024).	03/2020 – Present GPA: 4.19/4.3
University of Washington (UW) , Seattle, WA Exchange Student > Tuition and stipend fully covered by Korea-U.S. Student Exchange Program Scholarship.	01/2025 – 06/2025
University of California, Berkeley , Berkeley, CA Exchange Student > Tuition and stipend fully covered by KAIST Presidential Fellowship.	06/2022 – 08/2022
Korea Science Academy of KAIST , Busan, South Korea Science High School for Gifted Students	03/2017 – 02/2020

Research Experience

Paul G. Allen School of Computer Science & Engineering @ UW , Seattle, WA Undergraduate Researcher with Prof. Simon Du > Focus: Reinforcement Learning Theory.	01/2025 – 06/2025
Optimization & Machine Learning Laboratory @ KAIST , Seoul, South Korea Undergraduate Researcher with Prof. Chulhee Yun > Focus: Deep Learning Theory and Optimization.	03/2022 – 12/2024

Publications

(* denotes equal contribution)

- [3] **Does SGD really happen in tiny subspaces?**
[Minhak Song](#), [Kwangjun Ahn](#), [Chulhee Yun](#) [\[arXiv:2405.16002\]](#)
ICML 2024 Workshop on High-dimensional Learning Dynamics: The Emergence of Structure and Reasoning [ICMLW 2024]
- [2] **Linear attention is (maybe) all you need (to understand Transformer optimization)**
[Kwangjun Ahn*](#), [Xiang Cheng*](#), [Minhak Song*](#), [Chulhee Yun](#), [Ali Jadbabaie](#), [Suvrit Sra](#) [\[Paper\]](#) [\[arXiv:2310.01082\]](#)
International Conference on Learning Representations [ICLR 2024]
- [1] **Trajectory Alignment: Understanding the Edge of Stability Phenomenon via Bifurcation Theory**
[Minhak Song](#), [Chulhee Yun](#) [\[Paper\]](#) [\[arXiv:2307.04204\]](#)
Conference on Neural Information Processing Systems [NeurIPS 2023]

Research Projects

Effectiveness of Schedule-Free AdamW in LLM pre-training [Ongoing] with Dr. Kwangjun Ahn , Beomhan Baek , Prof. Chulhee Yun > Co-leading research on the advantages of Schedule-Free AdamW for language model pre-training. > Theoretically analyzing performance compared to modern learning rate schedulers (e.g., Warmup-Stable-Decay) based on loss landscape properties.	09/2024 – Present Microsoft Research, KAIST
Implicit bias of stochastic Adam on separable data [Ongoing] with Beomhan Baek , Prof. Chulhee Yun > Co-leading project to study the implicit bias of stochastic Adam in linear logistic regression. > Discovered that the implicit bias of Adam heavily depends on the choice of batch size and training data.	09/2024 – Present KAIST

SGD dynamics along Hessian eigenspaces in deep learning [3]
with [Dr. Kwangjun Ahn](#), [Prof. Chulhee Yun](#)

10/2023 – 05/2024
Microsoft Research, KAIST

- > Led research investigating whether SGD can be trained in the tiny subspace spanned by top eigenvectors of the Hessian.
- > Proved that SGD fails to decrease training loss when constrained to this tiny subspace.
- > Extended our findings to the Edge of Stability regime, Sharpness-Aware Minimization (SAM), and Adam.
- > Proposed the hypothesis of *ill-conditioned valley* loss landscapes to explain this phenomenon.

Optimization characteristics of linear Transformers [2]
with [Dr. Kwangjun Ahn](#), [Prof. Xiang Cheng](#), [Prof. Ali Jadbabaie](#), [Prof. Suvrit Sra](#), [Prof. Chulhee Yun](#)

08/2023 – 03/2024
MIT, KAIST

- > Co-led a project to develop a simplified abstraction of the Transformer from an optimization perspective.
- > Demonstrated that a simplified shallow linear Transformer replicates key aspects of training dynamics of Transformer, including Adam’s superior performance over SGD.

Understanding progressive sharpening and Edge of Stability phenomena [1]
with [Prof. Chulhee Yun](#)

01/2023 – 10/2023
KAIST

- > Led a project to analyze progressive sharpening and Edge of Stability phenomena in deep learning optimization.
- > Discovered and rigorously proved the novel phenomenon of *trajectory alignment*, which enables precise characterization of gradient descent trajectories in the Edge of Stability regime.

Industry Experience

Upstage, Seoul, South Korea

09/2022 – 12/2022

AI Research Engineer Intern

- > AI startup led by [Prof. Sung Kim](#) @ HKUST.
- > Designed personalized recommendation models using contextual bandit algorithms for e-commerce service.

Selected Honors and Awards

National Presidential Science Scholarship (45,000 USD), Korea Student Aid Foundation.	2020 – 2026
KAIST Presidential Fellowship (30,000 USD), KAIST.	2020 – 2026
KAIST Alumni Academic Scholarship (15,000 USD), KAIST Alumni Scholarship Foundation.	2021 – 2026
Korea-U.S. Student Exchange Program Scholarship (9,000 USD), Minister of Trade, Industry and Energy.	2025
Travel Award, ICLR 2024. Vienna, Austria	2024
Travel Award, NeurIPS 2023. New Orleans, LA	2023
Top Student Award (rank #1 at department), KAIST ISE.	Spring 2021, Fall 2021, Spring 2022
Dean’s List (top 2%), KAIST.	Spring 2021, Fall 2021, Spring 2022
Talent Award of Korea (50 high school students in Korea), Deputy Prime Minister and Minister of Education.	2019
Hanseong Scholarship for Gifted Students (10,000 USD), Hanseong Sonjaehan Scholarship Foundation.	2018 – 2019
Grand Prize, Korean Young Physicists’ Tournament.	2018

Teaching and Academic Activities

Participant, Deep Learning Theory Workshop and Summer School, Simons Institute. Berkeley, CA

Summer 2022

- > Part of “Summer Cluster: Deep Learning Theory” program at Simons Institute for the Theory of Computing.

Academic Tutor, KAIST. Daejeon, South Korea

2021

- > Courses: Calculus I (Spring 2021), Calculus II (Fall 2021).

Conference Reviewer: NeurIPS 2024, ICML 2025, ICLR 2025, AISTATS 2025

Skills

Languages: Korean (native), English (fluent) — TOEFL iBT: 108/120

Computer Languages & Software: Python, \LaTeX , MATLAB