Hanseul Cho (조한슬)

Room 9410, Building No.9., 85 Heogi-ro, Dongdaemun-gu, Seoul, Republic of Korea

✉ jhs4015@kaist.ac.kr | 🏠 hanseuljo.github.io | 🐙 github.com/HanseulJo

in linkedin.com/in/hanseul-cho | 🐦 @hanseuljo | 🎓 Google Scholar (Hanseul Cho)

Personal Profile

I am a Ph.D. student in the Optimization & Machine Learning (OptiML) Laboratory, advised by Prof. Chulhee Yun at Kim Jaechul Graduate School of AI (GSAI) in Korea Advanced Institute of Science and Technology (KAIST). Before this, I received my Bachelor's degree in Mathematical Sciences (major) and Computing Sciences (minor) at KAIST in 2022.

My primary research interests lie in optimization, machine learning, and deep learning, mainly focusing on mathematical/theoretical analysis and empirical improvements (especially based on theoretical understanding). Recently, I have been looking into understanding and mitigating the fundamental limitations of modern language models (e.g., length generalization and compositional generalization problems). Moreover, I am always interested in topics on various constrained and multi-level optimization problems, including minimax optimization, fair machine learning, reinforcement learning, continual/incremental learning, and more, with particular interest.

Publications

PREPRINTS

- Cho, Hanseul*, Jaeyoung Cha*, Srinadh Bhojanapalli, Chulhee Yun. "Arithmetic Transformers Can Length-Generalize in Both Operand Length and Count." *Under Review*. 2024. [arXiv]

- Jung, Hyunji*, Hanseul Cho*, Chulhee Yun. "Convergence and Implicit Bias of Gradient Descent on Continual Linear Classification." *Under Review*. 2024.

INTERNATIONAL CONFERENCES/JOURNALS

- Cho, Hanseul*, Jaeyoung Cha*, Pranjal Awasthi, Srinadh Bhojanapalli, Anupam Gupta, Chulhee Yun. "Position Coupling: Improving Length Generalization of Arithmetic Transformers Using Task Structure." **NeurIPS 2024** & Short version in ICML 2024 Workshop on Long-Context Foundation Models (LCFM). [arXiv] [OpenReview]

- Shin, Baekrok*, Junsoo Oh*, Hanseul Cho, Chulhee Yun. "DASH: Warm-Starting Neural Network Training in Stationary Settings without Loss of Plasticity." **NeurIPS 2024** & Short version in ICML 2024 Workshop on Advancing Neural Network Training (WANT): Computational Efficiency, Scalability, and Resource Optimization. [arXiv] [OpenReview]

- Lee, Jaewook*, Hanseul Cho*, Chulhee Yun. "Fundamental Benefit of Alternating Updates in Minimax Optimization." **ICML 2024** & Short version in ICLR 2024 Workshop on Bridging the Gap Between Practice and Theory in Deep Learning (BGPT). [arXiv] [OpenReview]
  - **Spotlight at ICML 2024. (Top 3.5%: (144+191) of 9,473 valid submissions)**

- Lee, Junghyun*, Hanseul Cho*, Se-Young Yun, Chulhee Yun. "Fair Streaming Principal Component Analysis: Statistical and Algorithmic Viewpoint." **NeurIPS 2023**. [arXiv] [OpenReview]

- Lee, Hojoon*, Hanseul Cho*, Hyunseung Kim*, Daehoon Gwak, Joonkee Kim, Jaegul Choo, Se-Young Yun, Chulhee Yun. "PLASTIC: Improving Input and Label Plasticity for Sample Efficient Reinforcement Learning." **NeurIPS 2023**. [arXiv] [OpenReview]

- Cho, Hanseul and Chulhee Yun. "SGDA with Shuffling: Faster Convergence for Nonconvex-PŁ Minimax Optimization." **ICLR 2023**. [arXiv] [OpenReview]

DOMESTIC CONFERENCES/JOURNALS

- Jung, Hyunji*, Hanseul Cho*, Chulhee Yun. "Convergence and Implicit Bias of Gradient Descent on Continual Linear Classification." *Short version in the 11th Joint Conference of Korean Artificial Intelligence Association* (JKAIA 2024).
  - **Best Paper Award (Top 3) & Oral presentation.**

- Cho, Hanseul and Chulhee Yun. "SGDA with Shuffling: Faster Convergence for Nonconvex-PŁ Minimax Optimization." *Short version in the 7th Joint Conference of Korea Artificial Intelligence Association* (JKAIA 2022).
  - **NAVER Outstanding Theory Paper Award (Top 3) & Oral presentation.**

Education

Korea Advanced Institute of Science and Technology (KAIST) *Seoul, Republic of Korea*
Ph.D. in Artificial Intelligence *Sept. 2023 – Current*
- Advisor: Prof. Chulhee Yun (Optimization & Machine Learning (OptiML) Laboratory, Kim Jaechul Graduate School of AI (GSAI), KAIST)

Korea Advanced Institute of Science and Technology (KAIST) *Seoul, Republic of Korea*
M.Sc. in Artificial Intelligence *Mar. 2022 – Aug. 2023*
- Advisor: Prof. Chulhee Yun (Optimization & Machine Learning (OptiML) Laboratory, Kim Jaechul Graduate School of AI (GSAI), KAIST)
- Thesis: "Improved Convergence Rate of SGDA by Shuffling: Focusing on the Nonconvex-PŁ Minimax Problems" (Approved by Chulhee Yun, Se-Young Yun, & Donghwan Kim)
- GPA: 4.22/4.3

Korea Advanced Institute of Science and Technology (KAIST) *Daejeon, Republic of Korea*
B.Sc. in Mathematical Sciences *Mar. 2017 – Feb. 2022*
- Minor in Computing Sciences
- Summa Cum Laude (GPA: 4.05/4.3)

University of Twente *Enschede, Netherlands*
Exchange Student Program *Feb. 2020 – Jul. 2020*
- Major in Applied Mathematics

Incheon Science High School *Incheon, Republic of Korea*
High School *Mar. 2015 – Feb. 2017*
- Early graduation by one year (i.e., two-year course)

Awards

2024	**Best Paper Award (Top 3)**, JKAIA 2024	*Republic of Korea*
2024	**Top Reviewer (Top 8.6%: 1,304 of 15,160 reviewers)**, NeurIPS 2024	*Vancouver, Canada*
2022	**NAVER Outstanding Theory Paper Award (Top 3)**, JKAIA 2022	*Republic of Korea*
2022	**Summa Cum Laude**, Bachelor's, KAIST	*Republic of Korea*
2017 – 2020	**The National Scholarship for Science and Engineering**, Korea Student Aid Foundation	*Republic of Korea*
2017 Fall	**Dean's List**, The School of Freshman, KAIST	*Republic of Korea*

Skills

Programming	Familiar: Python (PyTorch, JAX, NumPy, Scikit-learn, Jupyter, Pandas, etc.), MATLAB. Novice: C, C++, R, HTML/CSS, Scala
Computer Misc.	Familiar: LaTeX (Overleaf/VSCode), Git, Microsoft Office. Novice: Adobe (Lightroom, Premiere Pro, After Effects, Photoshop)
Music	Playing the drums and percussions

Ongoing Projects

Length Generalization in Language Models (e.g., Transformers) *Seoul, Republic of Korea*
KAIST AI + Google *Nov. 2023 – Current*
- Transformer (Vaswani et al., 2017) models have exhibited tremendous success in various areas of artificial intelligence such as natural language processing, computer vision, and reinforcement learning.
- However, Transformers often suffer from the impossibility of 'length generalization': they usually fail to extrapolate their performance onto longer sequences than those at the training time.
- We aim to come up with a unifying framework to solve the length generalization problem on several synthetic tasks and arithmetic tasks, such as addition, Nx1 multiplication, copying/reversing under the presence of duplicate tokens, and the parity task. The exact structures for those tasks are known, so we might be able to inject the structure into our Transformer model directly.
- Furthermore, we would like to find a method that automatically discovers the hidden structure of the tasks to apply our methodology for length generalization to general NLP tasks.

Fair (Streaming) Principal Component Analysis - Complexity Lower Bound *Seoul, Republic of Korea*
KAIST AI *Mar. 2023 – Current*
- In Lee et al. (2023), we have studied the memory-efficient & group-wise fair principal component analysis (PCA) methodology called 'Fair Streaming PCA'.
- We proved the upper bound of sample complexity of our main algorithm 'Fair Noisy Power Method (FNPM)' mainly exploiting matrix Bernstein inequality. However, the tightness of the bound has not been checked yet.

Implicit Bias of Continual Learning in Logistic Regression *Seoul, Republic of Korea*
KAIST AI *Jul. 2023 – Current*
- Theoretical understanding of continual learning is mostly done in a simple regression setting, where the objective function is mostly based on a quadratic loss. In such a setting, gradient-based algorithms typically found a minimum for each task at a finite point.
- In contrast, however, the trajectory of continual learning has not been well understood in the classification setup, including even a logistic regression. In logistic regression, the iterates of the gradient-based algorithm diverge to infinity, which makes the theoretical analysis difficult.

Overcoming Loss of Plasticity of Neural Network *Seoul, Republic of Korea*
KAIST AI *Nov. 2023 – Current*
- Neural networks often struggle with the problem called 'loss of plasticity,' which refers to the phenomenon that a model loses the ability to adapt to the shifts in data distribution, especially when new data points appear.
- Interestingly, re-initializing the model partially or entirely, so-called "re-learning," greatly helps the model to recover its adaptability in practice.
- Why is it beneficial? Furthermore, can we develop a "smart" way of re-learning instead of re-initializing the whole model?

Contextual Model-based Black-box Optimization   *Seoul, Republic of Korea*

KAIST AI   *Nov. 2023 – Current*

- Related Keywords: offline policy optimization, contextual optimization/recommendation, black-box optimization with side information

Experiences

Machine/Deep Learning Theory + Physics (MDLTP) Seminar   *Seoul, Republic of Korea*

(Co-)Organizer   *Jul. 2022 – Feb. 2023*

- Homepage: sites.google.com/view/mdlt-p
- Jointly organized by OSI Lab, OptiML, and CSSPL
- Topics: Learning theory, loss landscape, trajectory analysis, (stochastic) optimization, high-dimensional statistics, statistical/mathematical physics, scientific machine learning, and more.

KAIST 2021 Post-AI Research Project   *Daejeon, Republic of Korea*

Undergraduate Researcher   *May 2021 – Dec. 2021*

- Jointly advised by Prof. Sangyoon Yi (DS Lab, GSFS, KAIST) & Prof. Jinkyoo Park (Sys. Int. Lab, ISysE, KAIST)
- Project: Research on 'AI-augmented Organizations' of Collaborative Decision Making and Learning
- Contribution: (1) Devised a model-based randomized algorithm for a single-player finite-horizon NK landscape optimization game; (2) Conducted some experiments on human-AI cooperation based on the algorithm that I devised

Korea Advanced Institute of Science and Technology (KAIST)   *Daejeon, Republic of Korea*

Individual Study   *Mar. 2021 – Jun. 2021*

- Advised by Prof. Jinwoo Shin (ALIN Lab, GSAI, KAIST)
- (1) gradient-based optimizers for large-batch setting (*e.g.*, LARS & LAMB); (2) theoretical analysis on gradient clipping (paper reading)

Korea Advanced Institute of Science and Technology (KAIST)   *Daejeon, Republic of Korea*

Individual Study   *Sep. 2020 – Feb. 2021*

- Advised by Prof. Jong-chul Ye (BISPL, BBE, KAIST)
- Assignment: Semantic Segmentation of Kidney Tumor with U-Net (with KiTS19 Challenge Dataset)

Korea Advanced Institute of Science and Technology (KAIST)   *Daejeon, Republic of Korea*

Individual Study   *Jun. 2020 – Aug. 2021*

- Advised by Prof. Yeonseung Chung (MAS, KAIST)
- Statistical learning theory

Services

Top-tier ML Conference Reviewer

- ICLR 2025. (3 papers)
- NeurIPS 2024. (6 papers) — awarded Top Reviewer (Top 8.6%: 1,304 of 15,160 reviewers) + Free Registration
- ICLR 2024. (2 papers)
- NeurIPS 2023. (2 papers)

GPU server manager of OptiML lab

- June 2022 – Feb 2024
- Being involved in installing OptiML lab's very first 5 GPU servers and a storage server
- Allocating GPU nodes to lab members
- Managing errors occurred in the servers

Languages

English   Professional Proficiency (i.e., sufficient for academic activities)

Korean   Native proficiency

Others   Had some introductory courses on French, German, Classical Latin, & Chinese.