



# Changyu Gao

Email: changyu.gao@wisc.edu  
 cyugao  in/changyu-gao

## EDUCATION

---

### University of Wisconsin–Madison

*Ph.D. Student in Industrial Engineering, Optimization Track*

Madison, WI

*Sep 2019 – Present*

### University of Wisconsin–Madison

*M.S. Student in Computer Science*

Madison, WI

*Feb 2021 – Present*

### University of Science and Technology of China

*B.S., Mathematics and Applied Mathematics (Outstanding Graduate)*

Hefei, China

*Aug 2015 – June 2019*

## RESEARCH EXPERIENCE

---

*PhD advisor:* **Prof. Stephen Wright**

### Differentially Private Optimization

- Designed a differentially private optimization algorithm for finding an approximate second-order stationary point with convergence guarantees.
- Applied the algorithm to the empirical risk minimization problem with adaptive mini-batching.
- Implemented the algorithm using Pytorch and demonstrated its performance through experiments.

### Neural Collapse and Loss Function Design

*Collaborator:* **Dr. Yiqiao Zhong**

- Carried out experiments to the influence of various loss functions on the neural collapse phenomenon.
- Characterized and analyzed the convergence behavior for the cross entropy loss function with label smoothing, for linearly separable data.

### Optimization Methods for Probabilistic Soft Logic (PSL)

- Implemented and tested HOGWILD! and Frank-Wolfe methods for PSL framework using Java.
- Carried out inference experiments on real datasets.
- Proved theoretical guarantees for the proposed formulation of PSL.

### Parameter Learning with Derivative-Free Optimization (DFO) Methods

- Implemented the parameter learning procedure for Lorenz96 model using DFO methods in Python.
- Performed optimization with uncertainty function values using soft interpolation and Gaussian Process methods.

## WORK EXPERIENCE

---

### Research Scientist Intern, Meta

Menlo Park, CA

*Team: Meta AI (FAIR) – Reality Labs*, supervised by **Dr. Min Xu**

*Sep 2022 – Jan 2023*

- **Adaptive Training for Transformer-based Models:**

- \* Contributed to the open source project fairscale. Fixed gradient accumulation bugs.
- \* Implemented methods to collect and analyze the gradient statistics during training of the Translation task using the Transformer model. Based on the gradient statistics, designed and implemented an adaptive training method to improve the training efficiency.
- \* Our algorithm uses the gradient statistics to dynamically adjust the gradient accumulation steps and the learning rate, achieving the baseline with better efficiency.

### Applied Scientist Intern, Amazon

Seattle, WA

*Team: Delivery Experience (DEX) – AI*

*May 2021 – Aug 2021*

- **Mining Inconsistency Issues using Semantic Search Model:**

- \* Applied the semantic search model to the customer contact data, facilitating inconsistency detection.
- \* Collected and refined the queries for semantic search; oversaw the data annotation process.
- \* Implemented two fine-tuning schemes of the encoder used in the semantic search model in Tensorflow and thus improved the search model accuracy.
- \* Important inconsistency issues discovered were escalated to the corresponding issue owners.

## PROGRAMMING SKILLS

---

**Languages:** Python, SQL, MATLAB, R, C, C++, Java

**Frameworks:** Pytorch, Tensorflow, JAX, Pandas, Numpy, Scipy

## PUBLICATIONS

---

**Changyu Gao** and Stephen J Wright. Differentially Private Optimization for Smooth Nonconvex ERM arXiv preprint arXiv:2302.04972, 2023. TPDP 2023 Poster.

Charles Andrew Dickens, **Changyu Gao**, Connor Pryor, Stephen Wright, Lise Getoor. Convex and Bilevel Optimization for Neuro-Symbolic Inference and Learning. OpenReview. Submitted to ICLR.

**Hobbies and interests:** music, guitar, hiking, tech, society, language, reading.