Anton Sugolov

Toronto, ON sugolov.github.io
asugolov@gmail.com

Profile

- Highly motivated **MSc. Mathematics** student with 2 years of **research experience in deep learning** and statistics
- Completed **advanced coursework in probability, statistics**, numerics, optimization, and differential equations
- **Accepted to ICLR 2025** (link) on structure in linearized LLM layers (38+ open LLMs) and the impact on accuracy
- Experienced in **PyTorch** (3+ years), **LLM inference via HuggingFace**, distributed training with DeepSpeed
- Succeeded in balancing advanced coursework, delivering on research deadlines, and teaching assistantship

Education

MSc. Mathematics 9/24 – 12/25 (exp.)
University of Toronto

- **Key courses:** Probability I & II, Optimization, PDE I, Topics in Machine Learning
- **Topics:** Markov chains, stochastic calculus, numerical linear algebra, elliptic PDE theory, generative modeling

HBSc. Applied Mathematics and Statistics 9/20 – 6/24
University of Toronto *3.84*

- **Key courses:** Statistical Machine Learning I & II, Software Design I, Data Analysis I & II
- **Topics:** probabilistic learning, time series analysis, MCMC sampling, regression, software design

Publications

1. Aubry, M.[1], Meng, H.[1], **Sugolov, A.**[1], Papyan, V. Transformer Block Coupling and its Correlation with Generalization in LLMs. **Accepted to ICLR 2025**. *Equal contribution.*[1]
2. **Sugolov, A.**, Emmenegger, E., Paterson, A.D., Sun L. Statistical Learning of Large-Scale Genetic Data: How to Run a Genome-Wide Association Study of Gene-Expression Data Using the 1000 Genomes Project Data. *Statistics in Biosciences* (2023).

Experience

Vector Institute, Faculty Affiliate Researcher 11/23 - Present
Prof. Vardan Papyan *University of Toronto*

- Discovered coupled structure in linearized LLM layers, to appear in ICLR 2025 (repository)
- Implemented vectorization for Jacobians of transformer blocks and efficient SVD algorithms
- Created inference pipelines on SLURM cluster to collect performance metrics while balancing compute resources
- Developing mathematical results to describe the coupling phenomenon and low-rank implicit biases in ResNets

Research Assistant 6/20 - 6/21
Prof. Lei Sun and Dr. Andrew Paterson *University of Toronto*

- Implemented large-scale ($> 10^6$ dim.) PCA, linear, and logistic regression for ERAP2 gene expression
- Created and led a workshop for 15 first-year level students to successfully replicate statistical tests (repository)

Projects

SkipNorm | *PyTorch, DeepSpeed* (repository)
- Trained classification ViT with data parallelization and DeepSpeed acceleration on Slurm cluster with 4xRTX6000
- Building transformer training improvements through encouraging low-rank structure in the residual stream

Software Design Course Project | *Java* (repository)
- Created Java academic timetable builder adhering to SOLID design principles and patterns for OOP

Score Based Sampling | *JAX* (repository)
- Implementation of score-based generative models (SSM, DSM), and Langevin MC sampling in JAX

Skills

Programming: Python, Java, R
Technical: PyTorch, JAX (optax + eqx), DeepSpeed, Slurm, Linux, HuggingFace
Languages: Ukrainian, French

Honours

Vector Scholarship in AI - Masters' 2024
Vector Institute
ICSA Best Paper Award 2025
International Chinese Statistical Association, Invited Speaker
NSERC Undergraduate Summer Research Award 2023
University of Toronto