Anton Sugolov

Toronto, ON anton.sugolov@mail.utoronto.ca
sugolov.github.io github.com/sugolov

Education

MSc. Mathematics 9/24 – 8/25 (exp.)
University of Toronto

· Probability I & II, PDE I, Optimization, Variational Methods in Generative Neural Networks

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

· Statistical Machine Learning I & II, Data Analysis I & II, Software Design I

Experience

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

· Investigating the presence of consistent SVDs of Jacobians in residual blocks after training
· Implemented inference for 38+ LLMs (Llama, Gemma, Phi) and evaluation metrics for SVDs of layers
· Extracted Stein unbiased risk estimate for denoising generative models (consistency, rectified flow)

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

· Performed statistical tests to preprocess genetic data, displayed population structure with PCA, performed linear and logistic regressions on human genome with Plink v2.0 and R
· Created and led a workshop for 15 first-year level students to successfully replicate statistical tests

Projects

SkipNorm (repository)

· Trained classification ViT with data parallelization and DeepSpeed acceleration on 4xNvidia RTX6000
· Implemented data augmentation (MixUp, CutMix) and hyperparameter tuning scripts on Slurm cluster
· Working towards stabilizing transformer training through normalizing the residual stream

Score Based Sampling in JAX (repository)

· Implementation of score-based generative models (SSM, DSM), and Langevin MC sampling in JAX

Timetable Scheduler, Software Design Course Project (repository), (design doc)

· Created Java academic timetable builder adhering to SOLID design principles and patterns for OOP

Topics Course Projects

· **MAT1855.** An introduction to Otto calculus and some applications. (write-up)
· **MAT1510.** Results and experiments based on '*The Emergence of Clusters in Self-Attention Dynamics*' by Geshkovski et al. Presented at Applied[2] Graduate Seminar. (write-up, slides)

Skills

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

Publications and Preprints

1. (Preprint, arXiv) Aubry, M.[1], Meng, H.[1], **Sugolov, A.**[1], Papyan, V. Transformer Block Coupling and its Correlation with Generalization in LLMs. *Submitted 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).

Teaching

MAT187H: Calculus II 1/25 – 4/25
Teaching Assistant

MAT133Y: Calculus and Linear Algebra for Commerce 9/25 – 4/25
Teaching Assistant

Genome Wide Association Workshop 6/21
Instructor and Organizer (repository)

Presentations

1. Results and experiments based on '*The Emergence of Clusters in Self-Attention Dynamics*' by Geshkovski et al. Aubry, M. and **Sugolov, A.** Applied ^{2} Graduate Seminar. Podium. November 2023.

2. Short lectures on the Inverse Kasteleyn Matrix, and Introduction to the Ising, Potts, Percolation, and Random Cluster Models. **Sugolov, A.** Seminar in the Dimer Model and Discrete Riemann Surfaces. November 2023.

Honours

Vector Scholarship in AI - Masters' 2024
Vector Institute *Unable to accept*

NSERC Undergraduate Summer Research Award 2023
University of Toronto

Best Poster (1/3) 2023
Data Sciences Institute, Summer Undergraduate Research Day

Regent's Scholarship I, II, III, and Dean's List 2021 – 2024
University of Toronto

Volunteering

Unissued Diplomas Project 2/24 – 3/24