Aleksei G Sorokin

+1 (630) 297 6261 • asorokin@hawk.iit.edu • alegresor.github.io
aleksei-sorokin • alegresor • akk3XSEAAAAJ

Research Scientific Machine Learning, Gaussian Processes, Quasi-Monte Carlo, Probabilistic Numerics

Programming Python (PyTorch, GPyTorch, Pandas, Matplotlib), Julia, C, MATLAB, R, SQL, Wolfram

Tools AWS (SageMaker, EC2), GitHub (general, actions, pages), LaTeX, Docker

Education

2021 - 2026 **PhD in Applied Math.** Illinois Institute of Technology (IIT). GPA $3.89/4$. Advisor *Fred J Hickernell*.

2017 - 2021 **Master of Data Science.** IIT. Summa Cum Laude. GPA $3.94/4$.

2017 - 2021 **B.S. in Applied Math, Minor in Computer Science.** IIT. Summa Cum Laude. GPA $3.94/4$.

Experiences

Jan - Dec 2025 **DOE SCGSR Fellow in Applied Mathematics** at **Sandia National Laboratory** in Livermore, CA. I am researching Gaussian process based scientific ML models for machine precision operator learning. I am also developing fast, scalable multi-task Gaussian processes for multi-fidelity modeling. We are preparing publications and open-source software with scalable GPU support e.g. see `FastGPs` below.

Summer 2024 **Scientific Machine Learning Researcher** at **FM (Factory Mutual Insurance Company).** I built SciML models, including Physics Informed Neural Networks (PINNs) and Deep Operator Networks (DeepONets), for solving Radiative Transport Equations (RTEs) used to speed up CFD fire dynamics simulations. Resulted in publication of "A neural surrogate solver for radiation transfer".

Summer 2023 **Graduate Intern** at **Los Alamos National Laboratory.** I modeled the solution processes of PDEs with random coefficients using efficient and error aware Gaussian processes. Resulted in publication of "Computationally efficient and error aware surrogate construction for numerical solutions of subsurface flow through porous media".

Summer 2022 **Givens Associate Intern** at **Argonne National Laboratory**. I researched methods to efficiently estimate failure probability using Monte Carlo with non-parametric importance sampling. Resulted in publication of *Credible Intervals for Probability of Failure with Gaussian Processes*.

Summer 2021 **ML Engineer Intern** at **SigOpt, an Intel Company**. I developed novel meta-learning techniques for model-aware hyperparameter tuning via Bayesian optimization. In a six person ML engineering team, I contributed production code and learned key elements of the AWS stack. Resulted in publication of "SigOpt Mulch: An intelligent system for AutoML of gradient boosted trees".

2021 - 2024 **Teaching Assistant** at **IIT**. I led reviews for PhD qualifying exams in analysis and computational math.

Open-Source Software

`QMCPy` **Quasi-Monte Carlo Python Software** (qmcsoftware.github.io/QMCSoftware), lead developer. This package provides high quality quasi-random sequence generators, automatic variable transformations, adaptive stopping criteria algorithms, and diverse use cases. Over the past five years, this project has grown to dozens of collaborators and multiple publications [3, 8, 9, 10, 11, 1, 2].

`FastGPs` **Scalable Gaussian Process Regression in Python** (alegresor.github.io/fastgps). Gaussian process regression (GPR) models typically require $\mathcal{O}(n^2)$ storage and $\mathcal{O}(n^3)$ computations. `FastGPs` implements GPR which requires only $\mathcal{O}(n)$ storage and $\mathcal{O}(n \log n)$ computations by pairing certain quasi-random sampling locations with matching kernels to yield structured Gram matrices. We support GPU scaling, batched inference, robust hyperparameter optimization, and multi-task GPR.

`QMCGenerators` **Quasi-Random Sequence Generators in Julia** (alegresor.github.io/QMCGenerators.jl). This package includes routines to generate and randomize quasi-random sequences used in Quasi-Monte Carlo. Supported low discrepancy sequences include lattices with random shifts and digital nets (e.g. Sobol' points) with random digital shifts, linear matrix scrambling, nested uniform scrambling, and higher order construction through digital interlacing. These features are also supported in `QMCPy`.

`AI on HPC` **AI Driven Science on Supercomputers Course** at **Argonne National Laboratory**. Key topics included handling large scale data pipelines and parallel training for neural networks.

Awards

2025 **Karl Menger Student Award for Exceptional Scholarship (Graduate)**, IIT.

2024 **College of Computing Excellence in Dissertation Research**, IIT.

2024 **Teaching Assistant Award**, IIT.

2023 **Outstanding Math Poster**, Los Alamos National Laboratory.

Publications

[1] Fred J. Hickernell, Nathan Kirk, and Aleksei G. Sorokin. *Quasi-Monte Carlo Methods: What, Why, and How?* 2025. arXiv: 2502.03644 [math.NA]. URL: https://arxiv.org/abs/2502.03644.

[2] Aadit Jain, Fred J. Hickernell, Art B. Owen, and Aleksei G. Sorokin. *Empirical Bernstein and betting confidence intervals for randomized quasi-Monte Carlo*. 2025. arXiv: 2504.18677 [math.NA]. URL: https://arxiv.org/abs/2504.18677.

[3] Aleksei Sorokin. *A Unified Implementation of Quasi-Monte Carlo Generators, Randomization Routines, and Fast Kernel Methods*. 2025. arXiv: 2502.14256 [cs.MS]. URL: https://arxiv.org/abs/2502.14256.

[4] Aleksei Sorokin, Xiaoyi Lu, and Yi Wang. "A neural surrogate solver for radiation transfer". In: *NeurIPS 2024 Workshop on Data-driven and Differentiable Simulations, Surrogates, and Solvers*. 2024. URL: https://openreview.net/forum?id=SHidR8UMKo.

[5] Aleksei G. Sorokin, Aleksandra Pachalieva, Daniel O'Malley, James M. Hyman, Fred J. Hickernell, and Nicolas W. Hengartner. "Computationally efficient and error aware surrogate construction for numerical solutions of subsurface flow through porous media". In: *Advances in Water Resources* 193 (2024), p. 104836. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2024.104836. URL: https://www.sciencedirect.com/science/article/pii/S0309170824002239.

[6] Aleksei G. Sorokin and Vishwas Rao. *Credible Intervals for Probability of Failure with Gaussian Processes*. 2023. arXiv: 2311.07733 [stat.ME].

[7] Aleksei G. Sorokin, Xinran Zhu, Eric Hans Lee, and Bolong Cheng. "SigOpt Mulch: An intelligent system for AutoML of gradient boosted trees". In: *Knowledge-Based Systems* (2023), p. 110604. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2023.110604. URL: https://www.sciencedirect.com/science/article/pii/S0950705123003544.

[8] Sou-Cheng T Choi, Yuhan Ding, Fred J Hickernell, Jagadeeswaran Rathinavel, and Aleksei G Sorokin. "Challenges in Developing Great Quasi-Monte Carlo Software". In: *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*. Springer. 2022, pp. 209–222.

[9] Sou-Cheng T. Choi, Fred J. Hickernell, Rathinavel Jagadeeswaran, Michael J. McCourt, and Aleksei G. Sorokin. "Quasi-Monte Carlo Software". In: *Monte Carlo and Quasi-Monte Carlo Methods*. Ed. by Alexander Keller. Cham: Springer International Publishing, 2022, pp. 23–47. ISBN: 978-3-030-98319-2.

[10] Aleksei G Sorokin and Jagadeeswaran Rathinavel. "On Bounding and Approximating Functions of Multiple Expectations Using Quasi-Monte Carlo". In: *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*. Springer. 2022, pp. 583–599.

[11] Aleksei G. Sorokin, Fred J. Hickernell, Sou-Cheng T. Choi, Michael J. McCourt, and Rathinavel Jagadeeswaran. "(Quasi-)Monte Carlo Importance Sampling with QMCPy". In: *IIT Undergraduate Research Journal* (2021), pp. 49–54. URL: http://urj.library.iit.edu/index.php/urj/article/view/48.