Sairam S

sairam.sundararaman2005 | Sairam S | sairam-sundararaman | Sairam S

Education

PES University 2023 – 2027
- B.Tech in Computer Science and Engineering (AI&ML) 8.59/10

Publications

The Final-Stage Bottleneck: A Systematic Dissection of the R-Learner for Network Causal Inference
Sairam S, Sara G, Shivam S
Transactions on Machine Learning Research (TMLR), 2026 | | W&B | OpenReview

Research Experience

Research Intern June 2026 – Present
WSAI, Indian Institute of Technology - Madras Advisor: Prof. Krishna Pillutla
- Designing and analyzing a novel differentially private mechanism for transformer LLMs, targeting tighter privacy-utility tradeoffs than standard DP-SGD baselines.
- Derived sensitivity bounds and Gaussian mechanism noise calibration from first principles; formalized privacy composition under zCDP across layers and generated tokens.
- Implemented and benchmarked on distilgpt2 and Qwen2.5-7B-Instruct.

Interests: Differential Privacy, Robustness, Adversarial ML.

Projects

The Geometry of Bit-Collapse: Quantization Instability Winter '26
- Investigated geometric failure modes of Post-Training Quantization (PTQ) on ResNet-18/CIFAR-10 using a custom **Hessian Spectrum Visualizer** (PyHessian, averaged over 40 batches for variance reduction).
- Diagnosed **"Bit-Collapse"**: Int8 quantization sharply narrows the loss basin and triggers spectral eigenvalue spikes, destabilizing standard SGD training regimes.
- Demonstrated that **SAM** fully restores geometric stability under quantization, recovering a flatter loss basin than the FP32 baseline — suggesting SAM as a principled mitigation for quantization-induced sharpness.

Reproducibility Study: The SAM Optimizer | W&B Fall '25
- Implemented a clean-room version of **Sharpness-Aware Minimization** (ICLR 2021) from first principles, manually engineering the dual-step perturbation-ascent / gradient-descent logic via explicit gradient normalization.
- Reproduced Wide-ResNet/CIFAR-10 accuracy ($\sim$96% vs. paper's $\sim$97%), then identified a structural gap in the original formulation: SAM implicitly outsources model-complexity regularization to weight decay, leaving it unaccounted for in the perturbation objective.
- Characterized a failure regime: the second-order gradient approximation degrades precisely at sharp minima — the regime SAM is designed to escape — revealing a theoretical tension in the algorithm's core assumption.

A First-Principles Dissection of Optimization Pathologies Fall '25
- Built a NumPy-based **Hessian geometry engine** to compute and visualize 2D loss landscapes (Rosenbrock, Beale), diagnosing the oscillatory failure mode of Vanilla SGD in high-curvature valleys.
- Analytically demonstrated how Momentum dampens gradient noise via exponential averaging, accelerating convergence in ill-conditioned optimization basins.

Reverse-Mode AutoDiff Engine Summer '25
- Architected a scalar-valued **autograd engine** from scratch, implementing reverse-mode automatic differentiation via local gradient accumulation on dynamically constructed DAGs with topological sort.
- Verified numerical correctness against PyTorch to $< 10^{-6}$ precision across composite differentiable functions, building ground-up intuition for backpropagation mechanics.

Skills

Languages/Frameworks: Python, PyTorch, NumPy
ML/DP: PyTorch, W&B, HuggingFace Transformers
Other: LaTeX, Git, Linux

Achievements

- Dr. MR Doreswamy Merit Scholarship Awardee 2024
- CN Rao Merit Scholarship Awardee 2024

Co-Curriculars & Volunteering

- Teaching Assistant - Mathematics for Computer Science - Prof. Anitha.N PES University. 2025