Hey there,

Here is your disruptive ML/AI digest for the day. Here's what you folks might find interesting today.

TOP PICKS

Novel Architectures and Objectives

NeuronSoup: Evolving Asynchronous, Shared-Neuron Temporal Graphs without Backpropagation

This paper proposes a genuinely radical departure from standard neural computation: instead of synchronous layer-by-layer processing, signals propagate asynchronously through a pool of shared neurons with delay-mediated interference. The key novelty is that hidden neurons are physically shared across paths, so when two signals traverse the same neuron, the second arrival encounters residual state from the first, producing constructive or destructive interference depending on polarity and timing. The entire architecture -- topology, weights, delays, connectivity -- is co-evolved via a genetic algorithm on a flat genome, with no differentiable computation graph required. The authors argue CMA-ES fails at this scale while genetic algorithms succeed, and demonstrate 85.9% MNIST accuracy with a 115 KB model using frozen ResNet18 features. The conceptual significance is that computation depth becomes per-sample adaptive and lateral interactions between pathways emerge naturally rather than being engineered, addressing limitations that Transformers and GNNs handle only through explicit architectural mechanisms.

Theory and Generalization

Sharp Stability Threshold and Certification for Designing Stable Residual Architectures

This work establishes a necessary and sufficient condition for stable training of residual architectures: the velocity field of each residual block must satisfy a sublinear-growth bound on input magnitude with exponent $q <= 1$. The authors prove this threshold via two independent routes -- classical ODE theory showing global forward flow exists only at $q <= 1$, and optimal-control analysis through the Hamilton-Jacobi-Bellman equation showing the training optimum is bang-bang and blows up for $q > 1$. The practical payoff is an arithmetic of input-magnitude exponents under the five operations that compose residual blocks, enabling certification of stability at the level of architectural primitives rather than empirical trial-and-error. They demonstrate this by reducing the supercritical Mamba block from $q = 5$ to $q = 1$ without layer normalization, and show the stability criterion is the exponent itself, not the presence of normalization. This matters because it provides a principled function space for residual architectures where forward dynamics, adjoint sensitivity, and composition are all well-controlled.

Learning in Infinitesimal Non-Compositional Sketches

This paper reformulates machine learning as a categorical factorization problem: ML models are specified as sketches (graphs with commutativity conditions, limit cones, and colimit cocones), and learning becomes the search for factorizations through quotient sketches. Non-compositionality is defined not as prediction error but as the obstruction to a universal factorization problem. The tangent lift applies the tangent functor to obtain infinitesimal perturbation conditions, and LINCS asks whether these perturbations preserve compositionality constraints. The framework introduces an INC endofunctor that iterates tangent lifts, producing a tower of factorization problems, and ML is cast as finding a coalgebraic fixed point where successive tangent unfoldings stabilize. The authors prove existence of a final INC coalgebra via the Aczel-Mendler theorem. This is a genuine conceptual departure -- it provides a category-theoretic foundation for ML that generalizes scalar loss functions and vector space assumptions, though the experimental validation is deferred to companion papers.

Probabilistic and Statistical Methods

NeuralChaos: Optimal Adapted Approximation of Square Integrable Predictable Processes

This paper introduces a neural operator architecture that approximates predictable square-integrable stochastic processes using only finitely many evaluations of the driving Brownian motion, sidestepping the need for large Wiener-chaos dictionaries and high-order iterated integrals. The authors prove density in the relevant function space and establish best N-term approximation rates for compressible and Malliavin-Sobolev regular processes. A striking negative result shows that finite-dimensional Markovian neural SDE models form a meagre, Gaussian-null subset regardless of discretization, while compressible processes are generic under non-degenerate sub-Gaussian sampling. This has direct implications for continuous-time reinforcement learning and stochastic control, where the standard neural SDE paradigm may be fundamentally limited in expressiveness. The work bridges stochastic analysis and neural approximation theory in a way that could reshape how we think about modeling continuous-time dynamics.

Riesz-Kernel Stein Variational Gradient Descent: Renormalized Entropy and Long-Time Particle Limits

This paper tackles a fundamental problem in Stein variational gradient descent: singular Riesz kernels offer quantitative population-level convergence but suffer from infinite self-interaction at the finite-particle level. The authors remove self-interaction and prove a many-particle, long-time sampling theorem: under a uniform bound on initial relative entropy per particle, the time-averaged empirical-measure law converges weakly to the target point mass as particle count and averaging horizon diverge. Below the logarithmic singularity threshold, they obtain explicit algebraic finite-particle error bounds. These results extend the joint-entropy approach from smooth-kernel SVGD to singular interactions, which is significant because singular kernels are precisely those that can provide the strongest convergence guarantees. The work bridges a gap between the theory of interacting particle systems and practical variational inference.

BY CATEGORY

Novel Architectures and Objectives

- Lyapunov Guidance: A Unified Framework for Stabilizing Generative Flows
- Muse: Representation Geometry of Muon Beyond Normalized Momentum

Theory and Generalization

- Non-vacuous Generalization Bounds for Reinforcement Learning with Verifiable Rewards
- A Minimal Interpretable Architecture for Zero-Shot Reconstruction of Dynamical Systems
- Steering Robustness into World Action Models via Mechanistic Interpretability and Optimal Control
- Adaptive Runge-Kutta Step Control Buys Training Loss, Not Generalization: An Honest Compute-Matched Study of RK-Adam Optimizers

Probabilistic and Statistical Methods

- Optimal Self-Distillation for Rectified Flow via Linear Probing
- Delocalization of bias in unadjusted Hamiltonian Monte Carlo and underdamped Langevin
- Subjective Risk Decomposition: A New View for Uncertainty Quantification
- Tamed Stochastic Gradient Hamiltonian Monte Carlo
- Precise sample covariance spectral norm error -- an RDT view
- Spectral Concentration and Recovery in Sparse High-Dimensional Random Geometric Graphs

- What's in a Smoothness Constant? Tighter Rates for Local SGD with Bounded Second-order Heterogeneity
- GAttNHP: Group Attention Neural Hawkes Process for Extrapolation Reasoning in Temporal Knowledge Graphs
- Post Hoc Inference for Component Attribution in Multivariate Change-Point Detection
- Parsimonious Mixtures of Skewed Bilinear Factor Analyzers

Transferable Methods from Adjacent Fields

- Subgrid-Scale Parameterization in Burgers' Equation Using Structure-Preserving Neural Networks and Entropy Variables
- RTS Smoother-Guided Learning of Physics-Based Neural Differential Models
- Probabilistic Physics-Informed Neural Networks for Estimating Heterogeneous Elastic Properties from Low-Resolution and Noisy Displacement Data

Emergent Behavior and Agentic AI

- Supervised Fine-Tuning vs. In-Context Learning: An Equilibrium Analysis of LLM Personalization under Congestion

---

Until the next announcement!

P.S. In the most recent submission window: cs.LG: 111 papers stat.ML: 17 papers