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

Three-Body Scattering for Generative Modeling (Peng Sun et al.)

This paper introduces **Three-Body Scattering Modeling (TBSM)**, a generative objective that is genuinely not reducible to GANs, diffusion, or autoregressive factorization. The core idea is to turn the energy distance into a constant-size per-projectile interaction: each generated sample is attracted toward one real data point and repelled from one independently generated sample, and the conditional expectation of this interaction equals the 2-Wasserstein gradient-flow velocity of the energy distance. This yields O(B) sample-level losses from a batch of B events, replacing the minibatch-wide all-pairs field used by prior drifting methods, and the conditional expectation can be tracked online to reduce field noise. The method achieves FID 2.23 on ImageNet-256 with a one-step pixel-space generator, which is competitive with far more expensive diffusion distillation pipelines. What matters here is the principled derivation: the objective comes from optimal transport geometry rather than empirical tuning, and the design map connecting diffusion supervision, drift dynamics, and GAN-like objectives gives a clean theoretical scaffold for future work.

Kernelized Linear Attention: Breaking the Capacity Wall with Symmetric Cones (Ayoub Ghriss, Sourav Chakraborty)

This work formulates the associative recall problem in linear attention as a **spherical-packing problem** and derives feature maps from first principles by certifying nonnegative attention weights through a self-dual homogeneous cone. The resulting KATA framework uses rank-one PSD features that offer a favorable capacity-interference tradeoff, characterized through the Welch interference floor, and admits spherical codes with exponentially many keys in the projection dimension. The implementation as fused Triton kernels reaches up to 11x FlashAttention-2 forward throughput at 131k tokens in the exact O(T) chunked-state form. On long-range MQAR and repeated-key overwrite benchmarks, several KATA variants outperform Gated DeltaNet while retaining near-perfect recall at 16x out-of-distribution length. The key conceptual contribution is that the feature geometry is not empirically chosen but derived from a geometric packing principle, which clarifies how positional embeddings, delta rules, and decay gates interact with the representational structure.

Sobek: Streaming Equivariant Tensor Product Convolutions (Vladimir Chorošajev, Cédric Bény)

Sobek addresses a fundamental inefficiency in equivariant GNNs: the conventional wisdom that edge-conditioned tensor-product convolutions require materializing edge-specific weights, messages, and adjoints. The authors show that these edge-sized intermediates are artifacts of the execution schedule, not requirements of the equivariant operator itself. By reassociating radial projection, spherical-harmonic coupling, and graph aggregation, edge-local products can be consumed directly into bounded receiver-side state, yielding a streaming formulation that preserves fully connected multiplicity mixing through forward, backward, and double backward passes. The CUDA backend achieves speedups from 1.2x to 49.7x across all 75 capacity-matched comparisons and reduces peak memory by up to 99%, while executing workloads two orders of magnitude beyond OpenEquivariance's capacity. This is a computational primitive contribution: it changes what is feasible for equivariant architectures without altering the mathematical operator, purely through schedule reassociation.

Probabilistic and Statistical Methods

Twisted Schrödinger Bridge Matching (Maxence Noble et al.)

This paper extends the Schrödinger bridge problem to reference processes that are **twisted Brownian motions** -- Feynman-Kac transforms induced by time-dependent differentiable potentials -- and provides a rigorous extension of the Iterative Markovian Fitting (IMF) scheme to this generalized setting. The resulting Twisted Schrödinger Bridge Matching (TSBM) yields a new bridge-matching loss that depends explicitly on the gradient of the potential and recovers the standard DSBM objective when the potential vanishes. The authors introduce trajectory-based variance-reduction techniques that stabilize optimization and could be useful beyond this specific setting. Empirically, TSBM demonstrates benefits for trajectory inference in crowd navigation and single-cell data. What is novel here is the principled generalization: rather than empirically modifying the bridge objective, the potential gradient term arises naturally from the generalized Schrödinger bridge formulation, opening a design space for incorporating physical or information-theoretic priors into the reference process.

Theory and Generalization

The Geometry of Semantic Space: A Continuous Geometric Framework for the Transformer Architecture (Zhihua Liang)

This paper constructs a continuous geometric framework that models every core Transformer component -- RMSNorm, RoPE, softmax attention, FFN, residual stream, SGD, weight decay -- as an integro-differential equation on a **semantic fiber bundle**. Starting from a single geometric axiom (the token sequence as a discrete 1-manifold with canonical measure lattice), the framework translates attention as entropic optimal transport (a Schrödinger bridge) and SGD as Itô diffusion violating detailed balance. The paper tests six quantitative predictions across five architectures from 124M to 8B parameters, including epsilon^{-1/2} Lipschitz scaling calibration at machine precision (R^2 = 1.000), a thermodynamic context-limit phase transition, and a non-equilibrium steady-state parameter vortex verified across two optimizers. Whether or not every prediction holds up under further scrutiny, the framework provides a predictive descriptive vocabulary for stability limits, context bounds, and optimization dynamics that is grounded in differential geometry and stochastic calculus rather than empirical observation alone.

BY CATEGORY

cs.LG

stat.ML

- An Adjoint-Sensitivity Framework for Lost-in-the-Middle Phenomena in Causal Residual Transformers
- Program Synthesis for Simulation-Based Inference: Joint Model Selection and Parameter Estimation
- The Value of Depth in Message Passing on Sparse Graphs: A Kesten-Stigum Dichotomy
- Dropout and Random Gradient Masking Are Asymptotically Equivalent in Large ResNets
- Backpropagation-Free Trunk Training via the Split Forward Gradients
- Topological Signatures of Context-Level Reliability in TabPFN
- A Deep Second-Order Stochastic Residual Method for Fully Nonlinear Parabolic PDEs
- Optimizing the Preconditioner: A Black-box Online-to-Nonconvex Conversion with Static Regret Minimization Oracles
- Unveiling Invariant and Transferable Latent Factors Across Heterogeneous Environments via ATLAS
- Projective Maximum Entropy: Universality and Acceptance-Region Calibration
- Amortized Inference for Sampling Distributions Where the Bootstrap Fails
- How Fast Do Signatures Learn? Statistical Theory and Applications for Path Regression
- The Resolution of Causal Heterogeneity

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 237 papers stat.ML: 31 papers