Hi all,

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

Optimization Geometrodynamics: A Framework for Dynamic Geometric Optimization

This is one of the clearest theory-first attempts I have seen to recast optimization as a coupled geometric dynamical system rather than "gradient plus heuristic preconditioner." The key novelty is the formal language: parameter trajectory, transported particle distribution, and time-varying Riemannian metric are treated as jointly evolving objects, which lets the paper separate invariant barriers from geometry-induced improvements. The most useful part for your interests is the introduction of *dynamic geometric complexity*, defined as the minimum geometric cost needed to reduce a chosen optimization difficulty observable. On strongly convex quadratics with full metric control, that complexity becomes an exact affine-invariant distance in log-spectrum space, which is a rare case where an optimizer-design intuition is turned into a precise benchmark quantity. I also like that the paper is explicit about what geometry cannot do: positive metrics preserve critical points and Morse indices, so this is not hand-wavy "geometry solves optimization" rhetoric. If this framework holds up, it could become a useful reference language for comparing adaptive optimizers at the level of reachable geometry rather than benchmark folklore.

Intrinsic Green's Learning: Supervised Learning on Manifolds via Inverse PDE

This paper stands out because it does not just add manifold regularization to a standard predictor; it redefines supervised learning as learning a source term of a linear PDE and reconstructing the target through a Green's kernel. That is a meaningful conceptual departure from direct function approximation, and it is especially relevant if you care about geometry-derived objectives and inductive biases. The low-rank tensor decomposition in a learned intrinsic chart is the practical hook: it turns a high-dimensional integral into independent one-dimensional integrals with cost linear in intrinsic dimension. The two-stage training procedure is also notable, because it explicitly separates coordinate discovery from source fitting and claims to avoid the collapse issues of end-to-end joint training. I would treat the MNIST result as secondary, but the automatic intrinsic-dimension recovery is the more interesting signal here. If the method generalizes beyond toy manifolds, this could be a useful bridge between PDE-based modeling and representation learning.

CaLiSym: Learning Symplectic Dynamics of Real-World Systems through Structured Canonical Lifts

What is genuinely new here is not just "another symplectic network," but the decision to impose exact symplectic structure in a *lifted canonical phase space* rather than on the observed dissipative system state itself. That is a smart way around the usual limitation that exact symplectic guarantees apply mainly to closed conservative systems. The lift is explicit and algebraic, and the authors emphasize that they avoid recurrent latent states, implicit solves, and inference-time ODE integration, which makes this more architecturally concrete than many geometric-learning papers. The proposed GRB-SympNet variant also looks like a principled local-function-approximation extension rather than a cosmetic tweak. For your interests, the main relevance is the inductive bias: this is a physically grounded architectural move that changes *where* the invariance is enforced, not just how strongly it is regularized. I would still want to know how sensitive the lifted construction is to modeling choices, but conceptually this is one of the stronger geometry-informed papers in the batch.

THEORY AND GENERALIZATION

The Optimal Sample Complexity of Learning Autoregressive Chain-of-Thought

This is a strong fit if you want formal results on emergent reasoning behavior rather than empirical chain-of-thought anecdotes. The headline claim is sharp: exact-trace learning of full autoregressive CoT has sample complexity matching the local next-token class rate, with no dependence on rollout length. That is surprising enough to matter, because exact-trace loss is unforgiving and one might expect compounding complexity from long rollouts. The technical novelty appears to be the introduction of *parity dimension*, a rollout-stable refinement of Daniely-Shalev-Shwartz dimension that does not blow up under autoregressive composition even when DS dimension can. In other words, the paper is not just proving a bound; it is identifying the right combinatorial object for reasoning traces. If you care about principled accounts of why autoregressive reasoning may or may not be harder to learn than token prediction, this is exactly the sort of theory paper worth reading closely.

Multiplication Beyond Groups: Stratified Fourier Mechanisms in Transformer Circuits

This is the most compelling mechanistic interpretability paper in the list because it pushes beyond the now-familiar "transformers learn group representations" story. The authors study modular multiplication over composite moduli, where zero-divisors make the operation non-invertible, so the standard global group-representation account should fail. Their proposed answer is a *monoid extension*: the model appears to partition inputs into local algebraic regions where group-like Fourier mechanisms still operate, but only piecewise. That is a meaningful conceptual extension of prior circuit analyses, not just another case study. The evidence they highlight, class-sensitive routing, low-rank write directions, and local character features explaining logits, sounds aligned with the kind of structure one would want if transformers are implementing stratified algebraic computation. I would read this as a serious attempt to generalize mechanistic theories of algorithm learning beyond clean group settings.

BY CATEGORY

NOVEL ARCHITECTURES AND OBJECTIVES

- Weight-Space Physics: Interpretable Hypernetworks for Lattice Quantum Field Theories
- Tensor Train Diffusion: Leveraging Low-Rank Structures for High-Dimensional Score-Based Sampling
- Heat-Kernel Entropy Profiles and Geometric Effective Sample Size for Weighted Measures on Manifolds
- Geometric-Nongeometric Optimizer Calculus: A Modular Language for Reachable Gradient Methods
- Restricted Dynamic Geometric Complexity: Certificates for Structured Preconditioning
- Higher-Order Geometric Updates for Levenberg-Marquardt Method via Riemann Normal Coordinates
- Gradient-free Riemannian Langevin Sampler
- Guidance Breaks the Fitted Operator: A Terminal-Fitted Repair for Classifier-Free Guidance
- DiPhon: Diffusion on Graphons for Scalable Graph Generation
- Gauge-Invariant Learnable Spectral Positional Encodings for Directed Graphs via Hermitian Block Krylov Subspaces

THEORY AND GENERALIZATION

- Is Randomness Necessary for Adaptive Data Analysis?
- Fast Rates for Semi-Supervised Learning via Data-Augmentation Graph Regularization
- Avoiding unsafe sets when training with Langevin Dynamics
- Local large deviations for linear-region growth in random piecewise-linear networks
- Finding a stationary point of a stochastic convex problem

- Statistical inverse learning and l1-regularization
- Transfer Learning for Linear Discriminant Analysis with a Shared Classification Signal
- Best-Arm Identification with Generative Proxy

PROBABILISTIC AND STATISTICAL METHODS

- Constrained Decoding for Diffusion Language Models via Efficient Inference over Finite Automata
- Tensorized algorithms and scalable filtering methods for hidden Markov and factorial hidden Markov models

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 132 papers stat.ML: 14 papers