Hi all,

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

Novel Architectures and Objectives

- Gromov-Monge Flow Matching for Equivariant Graph Generation rethinks how symmetry enters flow matching not just through the architecture but through the source-target coupling itself. The key insight is that once graph pairs are compared up to node relabeling, the natural geometry is the Gromov-Monge distance on the quotient space, and symmetrizing the coupling provably yields equivariant flow-matching minimizers. The authors show that quotient couplings can be lifted to aligned representatives at no additional cost, and they construct practical minibatch approximations using Gromov-Wasserstein relaxations. This matters because it changes only the training coupling while remaining compatible with standard permutation-equivariant architectures, yielding substantial sample quality improvements at small integration budgets for both continuous graph and categorical molecular generation.

Theory and Generalization

- Dynamical phase selection controls compute scaling in looped transformers reveals that looped transformers with identical architecture and objective, trained to identical accuracy, can land in distinct dynamical phases determined by initialization. The phases are characterized by different bifurcation mechanisms -- a saddle-node fold versus a Neimark-Sacker transition -- and each phase produces a fundamentally different scaling law for test-time compute. In the fold phase, a one-dimensional normal-form reduction predicts relaxation-time amplitudes from local derivatives of the trained map, yielding the parameter-free relation relating spectral gap to critical slowing down. This is a genuine conceptual departure from the assumption that test-time compute is an architectural property, showing instead it is governed by the dynamical phase selected during training.

- Muon with Finite Newton-Schulz: The Smoothing Benefit in Nonsmooth Nonconvex Optimization flips the conventional wisdom that the finite Newton-Schulz iteration in Muon is merely an approximation to the exact polar factor that can only hurt theoretical guarantees. Through the online-to-nonconvex conversion framework, the authors show that finite Newton-Schulz actually smooths the discontinuous polar map into a Lipschitz map of singular values, and this smoothing is precisely what enables convergence in the nonsmooth nonconvex setting. Crucially, Muon with the exact polar update may fail to converge, while logarithmic Newton-Schulz depth suffices. The resulting sample complexity bounds match best-known guarantees for nonsmooth nonconvex optimization, and the argument extends to general spectral maps with the same smoothing property.

Transferable Methods from Adjacent Fields

- Neural Renormalization Group Flow for Percolation demonstrates a supervised, scale-shared neural architecture that learns real-space renormalization group coarse-graining rules for two-dimensional site percolation. The model recursively applies the same learned rule across scales, trained only on small lattices yet extrapolating to substantially larger systems while recovering finite-size scaling near the critical point. What makes this noteworthy is the finding that the learned latent representation must exhibit critical fluctuations and scale-dependent flows consistent with the RG structure -- the network implicitly discovers the renormalization-group fixed point rather than merely fitting a surrogate. This is a clean example of physics-informed architectural structure enabling extrapolation beyond the training distribution.

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 130 papers stat.ML: 22 papers