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

### Wasserstein Residuals: Learning Gradient Flows from Population Dynamics

This is one of the stronger picks today because it replaces the usual JKO-centric view of Wasserstein gradient flows with a residual objective built directly from the continuity equation. That matters because the standard JKO route is mathematically elegant but algorithmically awkward: it hard-codes time discretization and repeatedly solves expensive OT subproblems. The paper's key move is to define a non-negative loss whose minimum is exactly the target WGF, then combine it with a data-fitting divergence into a single global objective. The resulting "stitching" method is particle-based, simulation-free, and explicitly designed to handle sparse observations with large temporal gaps, which is a meaningful conceptual shift rather than a benchmark tweak. For your interests, the appeal is the combination of **information-geometric / OT structure** with a training objective that looks more like a variational residual principle than a standard generative-model surrogate. The abstract also suggests a useful unification story: several existing methods become special cases of the same residual formulation.

### What Does a Discrete Diffusion Model Learn?

This paper looks especially worth attention if you care about objectives that are actually faithful to the stochastic process being trained. The central claim is unusually sharp: for CTMC discrete diffusion, the negative ELBO is not merely a lower-bound surrogate but exactly the data entropy plus the path KL from the oracle reverse process to the learned one. That "Oracle Distance" theorem gives a much cleaner semantic interpretation of training than the usual denoiser-versus-score folklore, and the paper then shows that denoiser, cavity predictor, and score are just coordinate systems on the same reverse-rate object. The practical relevance is that choosing the wrong coordinate changes the learned law, so this is not just notation cleanup; it is a statement about objective mismatch. It also explains concrete pathologies, including why denoiser parameterization can make the uniform-noise ELBO diverge at initialization while the bridge plug-in remains finite. Bernhard Schoelkopf is on this paper, which adds some weight, but the real reason to read it is that it clarifies the foundations of **diffusion objectives** in a way that could affect how discrete generative models are parameterized and evaluated.

### Constrained Flow Matching via Lagrangian Dual Flows

The novelty here is not "flow matching, but better tuned"; it is the insertion of a primal-dual optimization mechanism directly into the generation dynamics. Instead of enforcing nonlinear constraints by projection, pseudoinverses, or repeated optimization at inference time, the method flows a dual co-state alongside the sample trajectory. That is conceptually appealing because it turns constrained generation into a dynamical system with explicit Lagrangian structure, rather than a post hoc repair step. If the guarantees in the abstract hold as stated, this could be a useful bridge between **generative modeling**, **control**, and **numerical optimization**, especially for robotics and planning settings where constraints are not linear or cheap to project onto. The paper is also relevant to your interests because it proposes a new computational primitive for constrained inference, not just a new conditioning trick for images. I would read this critically for the exact scope of the guarantees, but the core idea is genuinely more structural than most conditional-generation papers.

### THEORY AND GENERALIZATION

#### **Geometric Causal Models**

This is a strong theory-forward paper because it tries to generalize causal inference beyond the i.i.d. setting by making symmetry part of the causal model itself. The interesting move is to use group-theoretic symmetries and ergodic theory for amenable groups to establish identification in dependent data settings such as spatial, graph, and molecular data. That is more ambitious than simply applying equivariant networks to causal tasks; the claim is that symmetry can change what is identifiable and estimable. The estimation side combines geometric deep learning with scalable Bayesian inference, which makes the framework more than purely abstract. For your interests, this sits right at the intersection of **causal representation / causal inference**, **geometric structure**, and **formal theory**, and the DNA example suggests the framework may transfer to scientific ML rather than staying as a toy construction. David Blei is on the paper, and while that is a modest signal, the main draw is the attempt to make symmetry a first-class ingredient in causal identification.

#### **The Map Behind the Flow: Finite-Step Gradient Descent as a Dynamical System**

This is probably the most conceptually provocative optimization paper in the list. Rather than treating gradient descent through the infinitesimal gradient-flow limit, it studies fixed-step GD as a discrete dynamical system and argues that many deep-learning phenomena are properties of the training map itself. The abstract ties edge-of-stability, catapult behavior, balancing, and movement toward flatter representations to bifurcations and attractors of the discrete map, with a universal Ricker-type map emerging in a solvable large-depth limit. If that analysis is sound, it reframes the learning rate as a structural parameter that shapes representation selection, not just a numerical stability knob. This is highly aligned with your interest in **training dynamics**, **implicit regularization**, and mechanistic explanations that go beyond empirical lore. I would prioritize it because it aims to explain several seemingly separate phenomena within one discrete-time dynamical picture, which is exactly the kind of unifying theory that can matter.

## BY CATEGORY

### THEORY AND GENERALIZATION

Tightening the Score Matching Gap for Diffusion Models

Non-asymptotic Convergence of Stochastic Gradient Descent in Score-based Generative Models

Non-Asymptotic Error Bounds for SMC with Biased Proposals: Application to Conditional Diffusion Sampling

Sequential Correlations Change In-Context Learning: Effective Context Length and Architectural Mismatch

A Unified Framework for In-Context Learning with Causal and Masked Language Models

Minimum Block Width for Universal Approximation by Residual Neural Networks with Inner Width One

Fitted Occupancy-Ratio Evaluation without Bellman Completeness

### PROBABILISTIC AND STATISTICAL METHODS

FUSE: FK-Steered Multi-Modal Flow Matching for Efficient Simulation-Based Posterior Estimation

Asymptotic-Preserving A Posteriori Analysis of Diffusion and Flow-Matching Samplers

Reflected Schroedinger Bridge Matching

A Gradient Flow Perspective on Minimum MMD Estimation

Stable Global Weighting of Flow Mixtures using Simplex Exponential Moving Average

## NOVEL ARCHITECTURES AND OBJECTIVES

Foundations of Equivariant Deep Learning: Unifying Graph and Sheaf Neural Networks

ManifoldFlow: SPD-Relaxed Stiefel Layers with Learnable Singular Spectrum

Transformers with Physics-Informed Encodings and Simulation-Based Inference for Robust Detection of Eccentric Binary Black Holes in Pulsar Timing Array Data

## QUANTUM ML AND HYBRID PARADIGMS

Canonical quantization of neurons

Breaking the One-Dimensional Expressibility-Trainability Tradeoff

## EMERGENT BEHAVIOR AND AGENTIC AI

CausalGame: Benchmarking Causal Thinking of LLM Agents in Games

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 250 papers stat.ML: 47 papers