PINS: Proximal Iterations with Sparse Newton and Sinkhorn for Optimal Transport

Di Wu

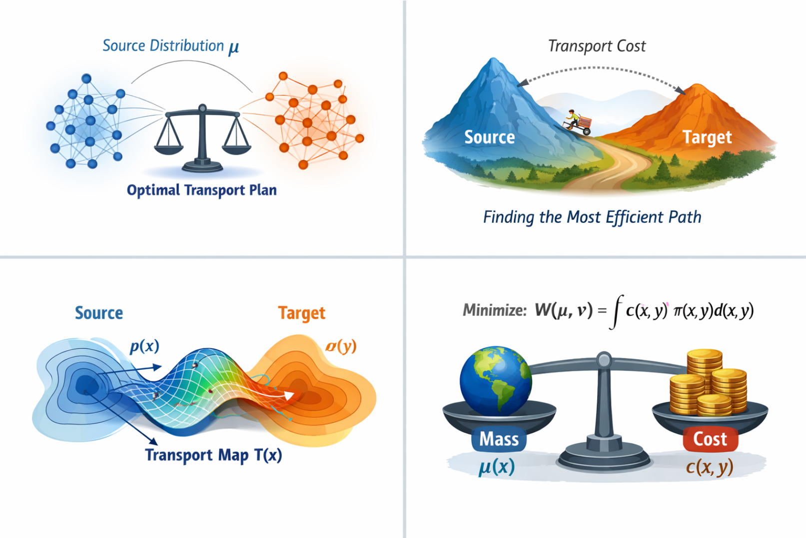
Department of Mathematics, University of Maryland, College Park
Joint work with Prof. Ling Liang, and Prof. Haizhao Yang

Applied Mathematics Seminar
George Washington University

February 27, 2026

Outline

Optimal Transport



Manifold-Constrained Hyper-Connections (mHC)

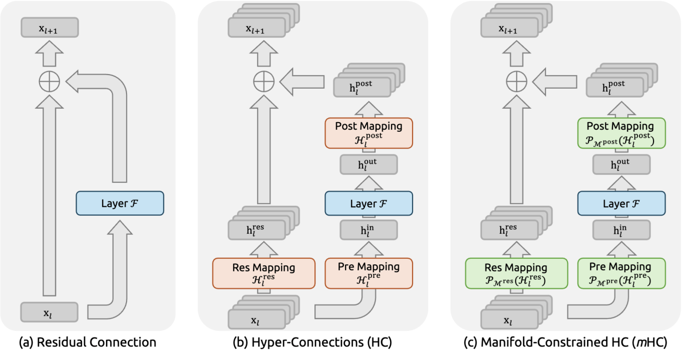


Figure: DeepSeek-AI, January 2026: Contribute to a deeper understanding of topological architecture design and suggest promising directions for the evolution of foundational models.

Introduction to Optimal Transport

- Foundational mathematical framework with broad applications in economics, physics, and machine learning.
- OT seeks to find the most efficient way to transport mass from one probability distribution to another while minimizing the total transportation cost.
- This has led to a surge of interest in developing scalable and efficient algorithms for large-scale OT problems.

Discrete Optimal Transport Formulation

- Classical discrete optimal transport problem:

$$\min_{X \in \mathbb{R}^{m \times n}} \langle C, X \rangle \tag{1}$$

- Subject to the mass conservation constraints:

$$X e_n = a, \quad X^\top e_m = b, \quad X \geq 0 \tag{2}$$

- $C \in \mathbb{R}^{m \times n}$: cost matrix.
- e_m, e_n: vectors of ones.
- $a \in \mathbb{R}^m_+, b \in \mathbb{R}^n_{++}$: marginal distributions satisfying $e_m^\top a = e_n^\top b = 1$.
- X: joint distribution with marginals a and b.

Computational Challenges

- **Significant computational resources**.
- For two distributions with n components each, the decision variable has a dimensionality of $\mathcal{O}(n^2)$.
- Standard linear programming algorithms incurs a computational cost per iteration of approximately $\mathcal{O}(n^6)$.

Entropic Regularization (Schrödinger's Problem)

- A controlled level of smoothness $\Rightarrow$ numerically stable and computationally tractable.
- The regularized problem is defined as:

$$\min_{X \in \mathbb{R}^{m \times n}} \langle C, X \rangle + \eta \sum_{i,j} X_{ij} \log(X_{ij}) \tag{3}$$

- Subject to $Xe_n = a, X^\top e_m = b$, where $\eta > 0$ is the regularization parameter.

The Sinkhorn Algorithm

Algorithm 3.1 The Sinkhorn Algorithm

1: **Inputs:** Cost matrix $C \in \mathbb{R}^{m \times n}$, marginal distributions $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$, initial point $(f^0, g^0) \in \mathbb{R}^{m+n}$, $X^0 \in \mathbb{R}^{m \times n}$, entropy regularization parameter $\eta > 0$.

2: **for** $k \geq 0$ **do**
3: $\quad f^{k+1} \leftarrow f^k + \eta \left(\log(a) - \log\left(X(f^k, g^k) e_n \right) \right).$
4: $\quad g^{k+1} \leftarrow g^k + \eta \left(\log(b) - \log\left(X(f^{k+1}, g^k)^\top e_m \right) \right).$
5: **end for**
6: **Output:** $X(f^{k+1}, g^{k+1})$.

- Operating in the log-domain mitigates some numerical instability and improves robustness for relatively small η.
- The optimal regularized transport plan takes the form:

$$X(f, g) = \exp\left(\frac{1}{\eta} (f e_n^\top + e_m g^\top - C - \eta E) \right) \tag{4}$$

Limitations of Sinkhorn-Type Algorithms

- **Numerical Instability:** The parameter η must not be too small, as small values lead to overflow/underflow due to exponential computations.
- **Approximation Error:** The Sinkhorn algorithm provides only an approximate solution to the original, unregularized OT problem.
- **Slow Convergence:** The algorithm utilizes a first-order alternating maximization approach, which converges slowly, especially when η is small.
- **Key Insight:** The trade-off between approximation accuracy and numerical stability remains a key challenge in real-world large-scale applications.

Exactness Matters

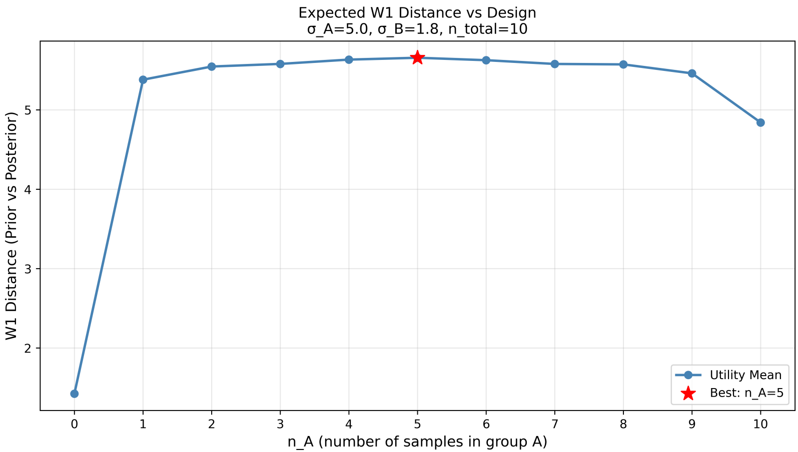


Figure: Our recent work on OED, where the utility is very sensitive to the numerical distance.

Motivation for PINS

- **Question:**

  How to design an efficient, robust, and scalable algorithm that can solve the original optimal transport problems with very high accuracy?

- **Solution - PINS**

- **Ingredients:** Proximal Iterations, sparse Newton's method, the Sinkhorn algorithm

Inexact Proximal Iterations

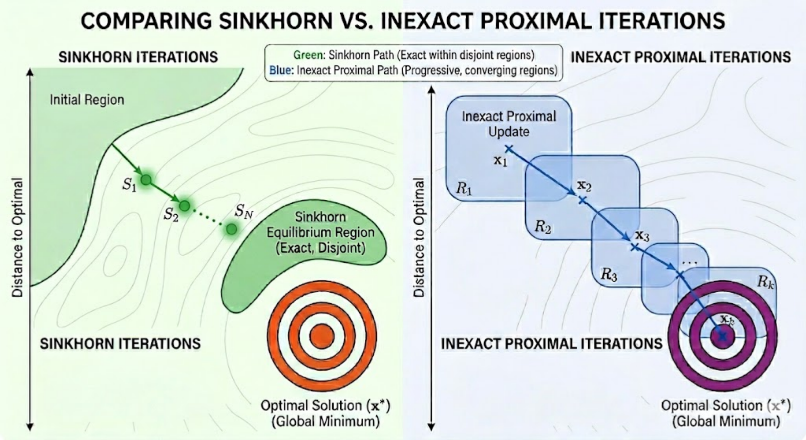


Figure: Sinkhorn v.s. Inexact proximal iteratioins

The Two-Phase Approach

- **Phase 1: Sinkhorn Initialization**
  - Coarse approximate solution to the regularized problem.
  - Mitigating numerical instability and ensuring rapid convergence.
- **Phase 2: Sparse Newton Refinement**
  - Using second-order information for accelerated convergence.
  - A sparsification technique is applied to reduce computational overhead.

Proximal Point Algorithm & Bregman Distance

- PINS utilizes the Boltzmann-Shannon entropy function: $\phi(X) := \sum_{i,j} X_{ij}(\log(X_{ij}) - 1)$.
- The Bregman distance between distributions X and Y is defined as:

$$D_\phi(X, Y) := \phi(X) - \phi(Y) - \langle \nabla\phi(Y), X - Y\rangle \tag{5}$$

- We apply the Entropic Proximal Point Algorithm (EPPA), updating X^{k+1} via:

$$X^{k+1} = \arg\min\{\langle C, X\rangle + \eta D_\phi(X, X^k) : Xe_n = a, X^\top e_m = b\} \tag{6}$$

- The EPPA subproblem is equivalent to:

$$\min_X \langle C - \eta \log(X^k), X\rangle + \eta \sum_{i,j} X_{ij} \log(X_{ij}) \tag{7}$$

Dual Problem Formulation

- To solve this subproblem efficiently, we derive its dual formulation:

$$\max_{f,g} P^k(f,g) = \langle a, f\rangle + \langle b, g\rangle - \eta \left\langle E, \exp\left(\frac{1}{\eta}(fe_n^\top + e_m g^\top - C^k - \eta E)\right)\right\rangle \tag{8}$$

- The gradient and Hessian of $P^k(f,g)$ can be computed explicitly.
- Newton's method requires solving a linear system involving the Hessian.
- To make this practical for large n, we utilize Hessian sparsification.

Sparsified Newton's Method

- **Sparsification:** Given a threshold ρ, PINS truncates the $(1-\rho)n^2$-smallest entries in the Hessian matrix to reduce computational costs.
- **Conjugate Gradient (CG):** The CG algorithm, a Krylov subspace method, is used to efficiently solve the sparse symmetric linear system at each Newton iteration via matrix-vector products.
- **Line Search:** A backtracking line search is employed to determine the optimal step size α^t in each Newton update.

Global Convergence Guarantee & Sparcification Analysis

- **Theorem 1:** Let $\{X^k\}$ be the sequence generated by the EPPA iterations with inexactness. Then $\{X^k\}$ converges to the optimal solution of the original unregularized OT problem.
- **Theorem 2:** We show that the solution computed in the first (Sinkhorn) phase is approximately sparse, independent of iteration k.

- From the formal Theorem 2, we can also find that tuning η remains a central challenge for both numerical considerations and the accuracy of the sparsification procedure.
- PINS offers greater flexibility in selecting the regularization parameter η than the standard Sinkhorn algorithm.

Numerical Setup & Datasets

- PINS was evaluated against standard and accelerated methods on three dataset types:

  1. **Synthetic Data:** Uniformly generated cost matrices with equal marginal constraints ($N = 50$ to 400).
  2. **MNIST:** Grayscale 28×28 images of handwritten digits, utilizing pixel Euclidean distance as cost.
  3. **Augmented MNIST:** $N \times N$ grids of MNIST images creating highly complex, sparse graphs that simulate difficult optimal transport problems.

Exact Optimality

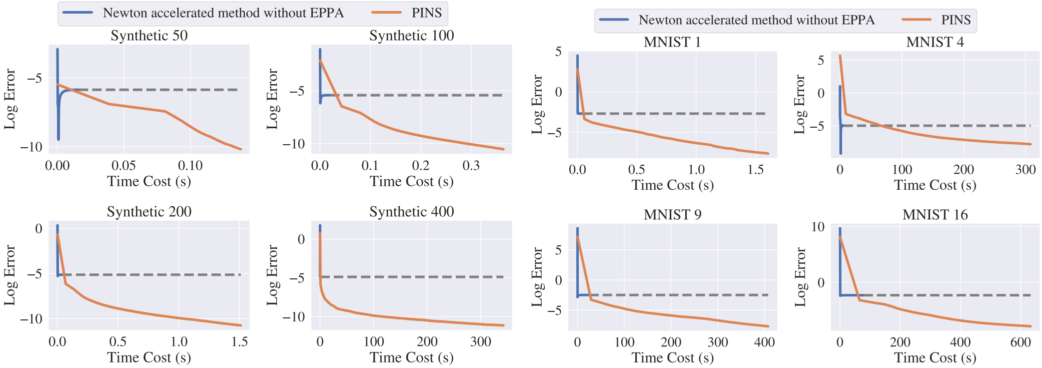


Figure: PINS achieves significantly higher accuracy for the original problem without a substantial loss in time efficiency.

Time Efficiency

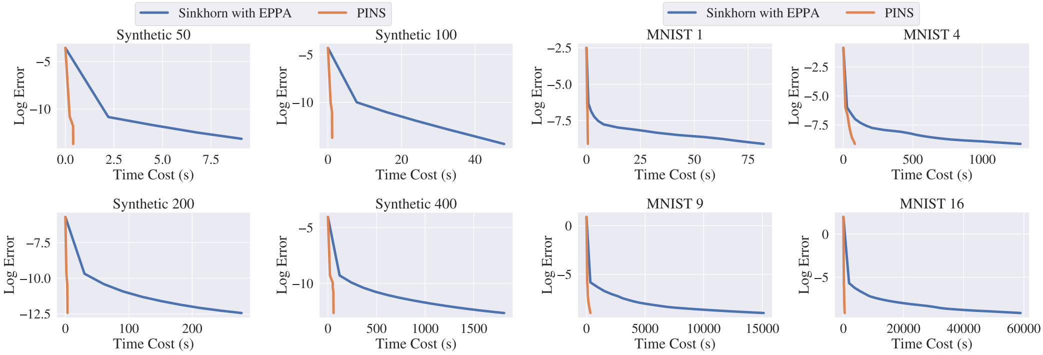


Figure: PINS runs significantly faster than the Sinkhorn algorithm with EPPA, particularly on large-scale datasets.

Regularization Parameter

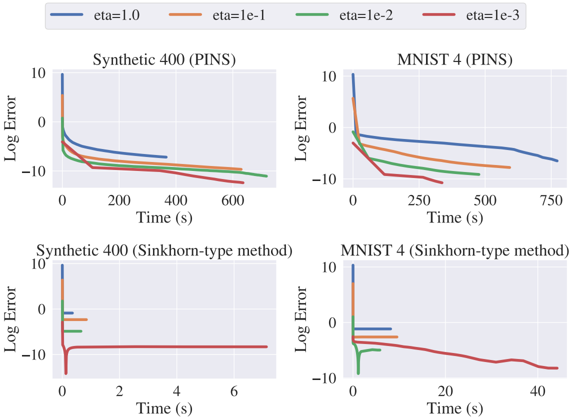


Figure: A small regularization parameter η might bring a more accurate solution to the Sinkhorn-type method, but it also brings some numerical issues

Introduction to Color Transfer

Application: Color Transfer via Optimal Transport

- **Goal:** Map the color palette of a source image to match the color distribution of a target image while preserving the source's structural content.
- **Mathematical Formulation:**
  - **Distribution:** 3-dimensional RGB color space $\Rightarrow$ Clustered into a discrete distribution
  - **Cost Matrix:** Euclidean distances between these color values $\Rightarrow$ The cost matrix C
  - **Color Mapping:** Optimal transport plan $T \Rightarrow$ Barycentric projection
- **Key Insight:** The accuracy of the transport plan directly dictates the visual quality of the output.

Color Transfer: Results

Figure: Illustration of color style transfer using the PINS algorithm. The first row displays the original images, while the second row shows the transferred images.

Color Transfer: Quantitative Comparison

- **SSIM**: Content similarity, higher is better
- **Style Loss**: style similarity, lower is better

Image Pair	**SSIM**		**Style Loss**	
	PINS	**Sinkhorn**	**PINS**	**Sinkhorn**
Autumn–Grafiti	**0.53**	0.29	**0.18**	0.29
Autumn–Comunion	**0.54**	0.37	**0.14**	0.17
Grafiti–Comunion	**0.76**	0.45	**0.34**	0.35

Key Takeaways

1. **Bridging the Speed-Accuracy Gap:** Standard Sinkhorn is fast but approximate; exact LP solvers are accurate but computationally prohibitive for large n. PINS successfully bridges this gap.
2. **The Power of Two Phases:**
   - **Phase 1 (Sinkhorn):** Provides an efficient, well-conditioned initialization.
   - **Phase 2 (Sparse Newton via EPPA):** Leverages second-order information to guarantee rapid convergence to the exact, unregularized solution.
3. **Algorithmic Robustness:** PINS drastically reduces sensitivity to the regularization parameter η, mitigating numerical overflow and eliminating hyperparameter tuning.
4. **Real-World Scalability:** PINS excels on complex, large-scale tasks like Augmented MNIST and color transfer, proving its viability for modern machine learning applications.

Thank You!

Questions?