Progressive Entropic Optimal Transport

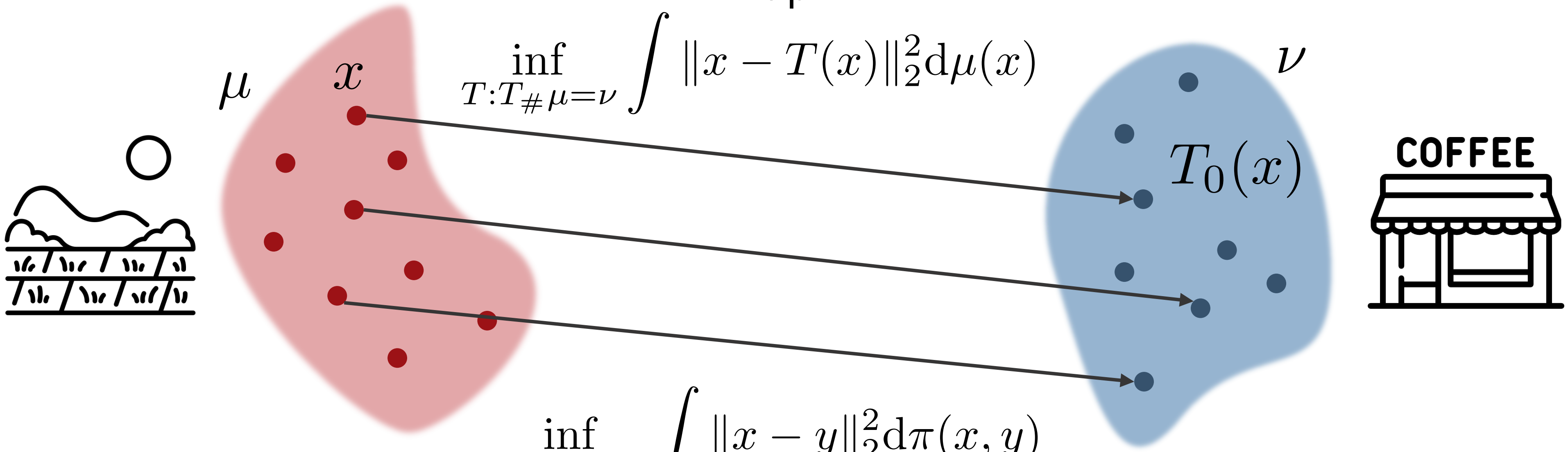
Parnian Kassraie

with Aram Pooladian, Michal Klein, James Thornton
Jon Niles-Weed & Marco Cuturi

Optimal Transport

Map

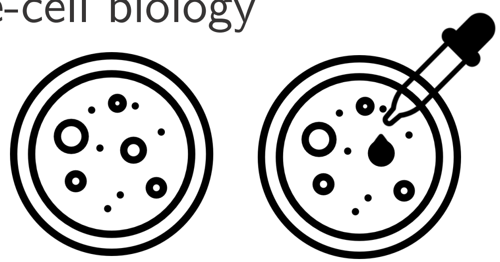
$$\inf_{T:T_{\#}\mu=\nu}\int \|x-T(x)\|_2^2\,\mathrm{d}\mu(x)$$



$$\inf_{\pi\in\Gamma(\mu,\nu)}\int \|x-y\|_2^2\,\mathrm{d}\pi(x,y)$$

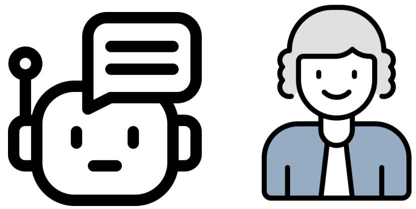
Coupling

Single-cell biology



[1, 2]

RL/Alignment



[3, 4, 5, 6]

[Figures from flaticon.com]

How do you solve it?

In full generality, OT does not have a solution or is very tough to solve.

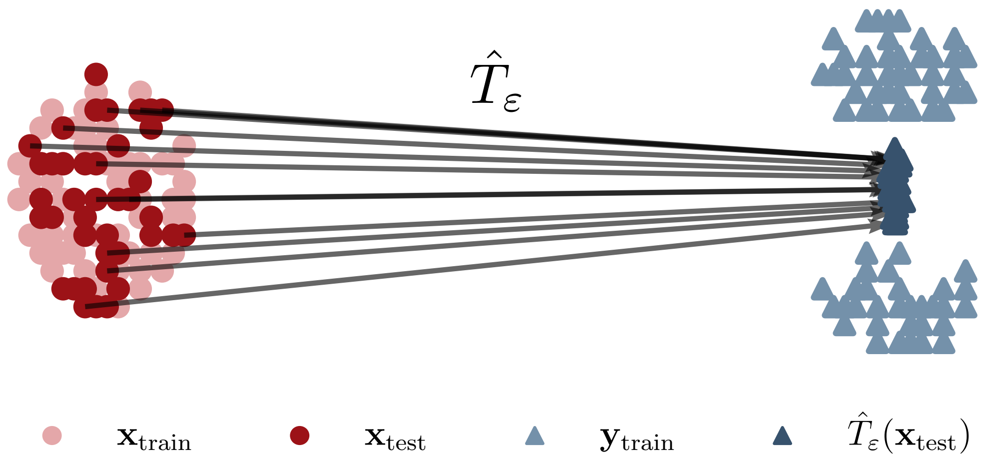
Entropic OT adds a regularization term to make things better:

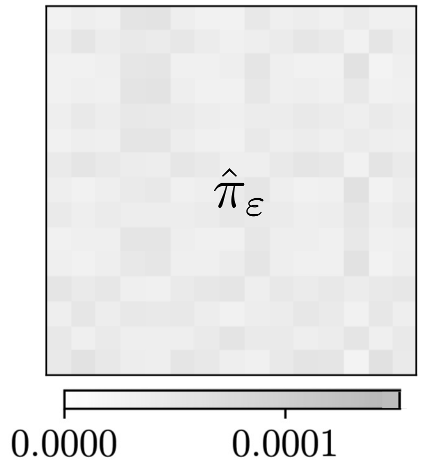
$$\inf_{\pi \in \Gamma(\nu,\mu)} \int \|x-y\|_2^2 \,\mathrm{d}\pi(x,y) + \varepsilon \,\mathrm{D}_{\mathrm{KL}}\left(\pi \| \mu \otimes \nu\right)$$

Given $\hat{\mu}$, $\hat{\nu}$: Sinkhorn's algorithm can solve this and return $\hat{T}_\varepsilon$ and $\hat{\pi}_\varepsilon$

Small ε: the algorithm may not converge

Large ε:

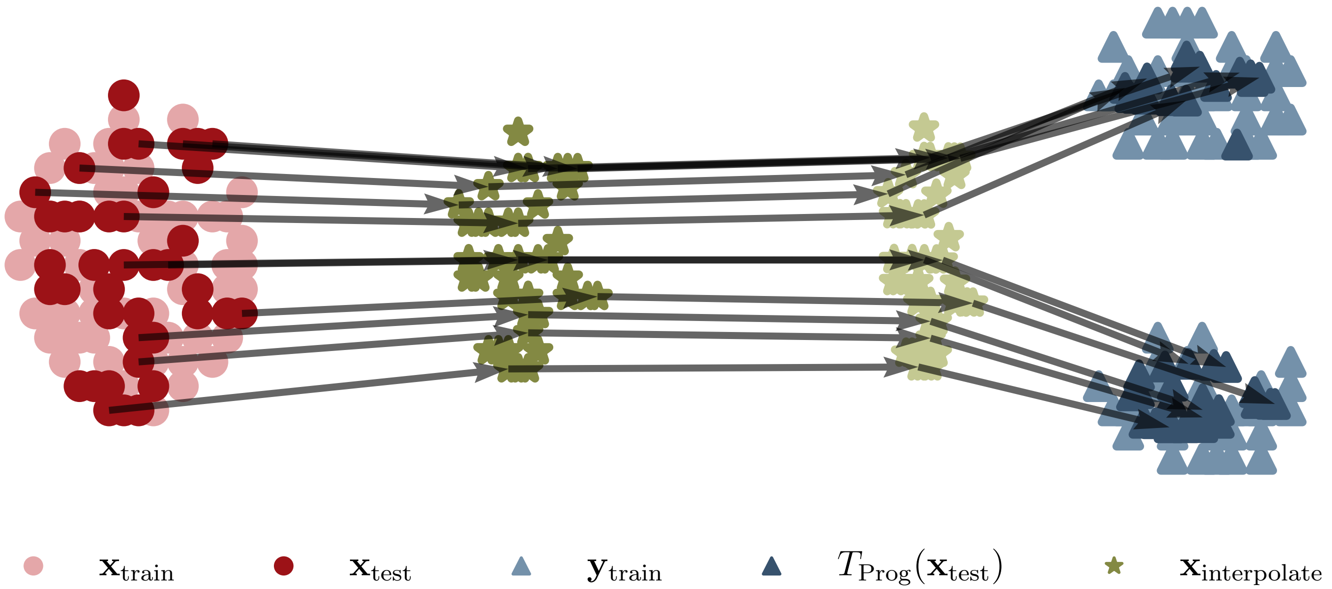


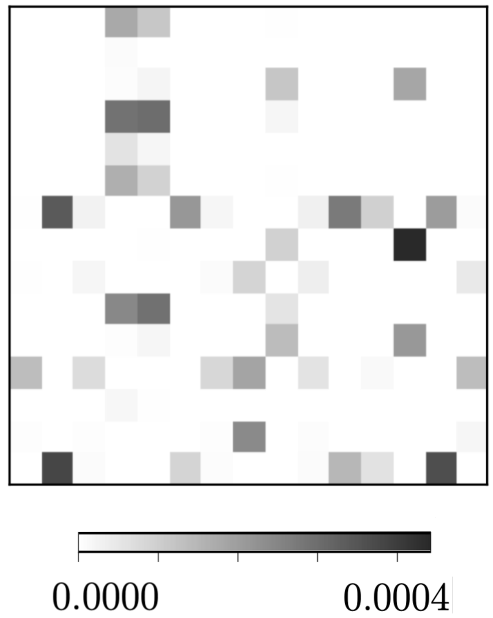


Our solution: ProgOT

Blend the static OT problem with the dynamic perspective.

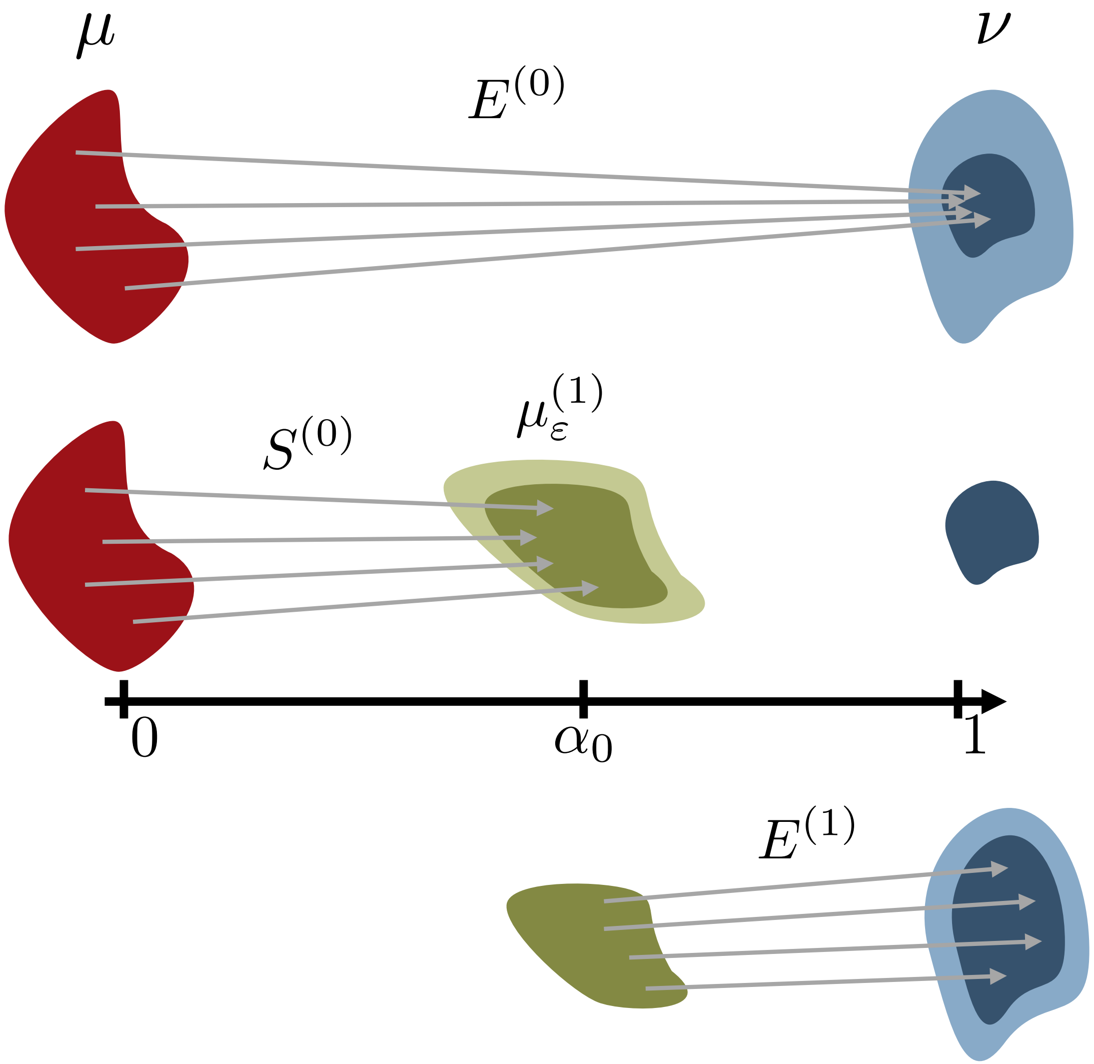
Solve a series of EOT problems, with reduced sensitivity to ε





- Elevates issue of regularization parameter
- Convergences to the ground truth (statistical guarantee)
- Competitive performance, scalable, and computationally light

ProgOT algorithm



Solve Entropic OT with large ε_0

Linearly Interpolate

$$\mu_\varepsilon^{(1)} = \left[(1-\alpha_0)\mathrm{Id} + \alpha_0 E^{(0)}\right]_\# \mu$$

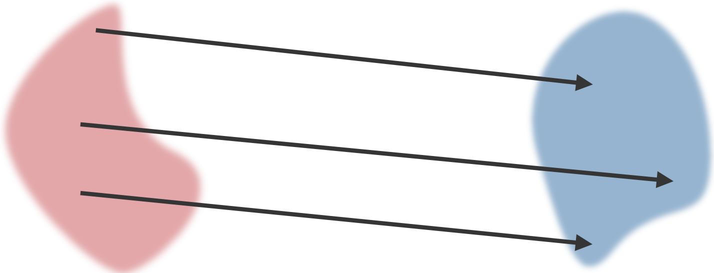
Reduce ε_1 and repeat

$$T_{\mathrm{Prog}}^{(1)} = E^{(1)} \circ S^{(0)}$$

We can repeat this K times to get $T_{\mathrm{Prog}}^{(K)}$

Theoretical guarantee

T_0: OT map between μ & ν



$T_{\text{Prog}}^{(K)}$: ProgOT map between $\hat{\mu}_n$ & $\hat{\nu}_n$

Theorem (Non-Asymptotic Consistency)

Given n i.i.d. samples from μ and ν, for an appropriate choice of $(\varepsilon_k)_k$ and $(\alpha_k)_k$, the K-step progressive map $T_{Prog}^{(K)}$ satisfies

$$\mathbb{E}\left\| T_{Prog}^{(k)} - T_0 \right\|_{L^2(\mu)}^2 \lesssim n^{\boxed{-\frac{1}{d}}},$$

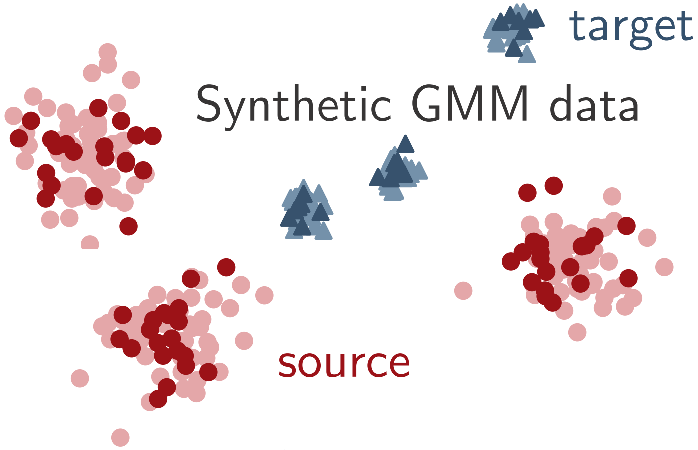
Independent of K!

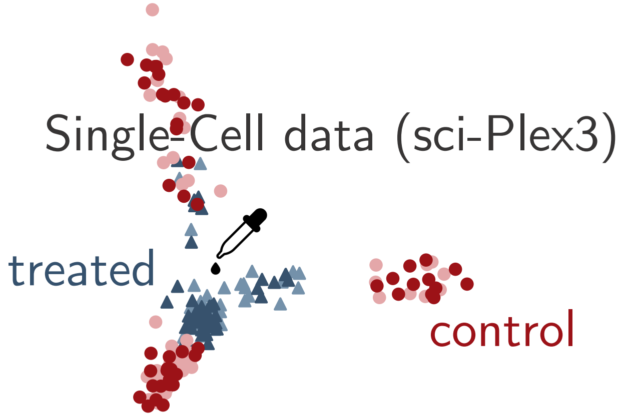
under regularity assumptions on μ, ν, and the true map T_0.

Proof idea: The intermediate steps of ProgOT are on the Wasserstein geodesic.

Map estimation

ProgOT outperforms other map estimators, including neural ones.





	$d = 128$	$d = 256$
ProgOT	**0.099**±0.009	**0.12**±0.01
EOT	0.12±0.01	0.16±0.02
Debiased EOT	0.11±0.01	0.128±0.002
Untuned EOT	0.250±0.023	0.276±0.006
Monge Gap	0.36±0.02	0.273±0.005
ICNN	0.177±0.023	0.117±0.005

Drug	Hesperadin			5-drug rank
d_{PCA}	16	64	256	
ProgOT	**3.7**±0.4	**10.1**±0.4	**23.1**±0.4	**1**
EOT	4.1±0.4	10.4±0.5	26±1.3	2
Debiased EOT	4.0±0.5	15.2±0.6	41±1.1	4
Monge Gap	3.7±0.5	11.0±0.5	36±1.1	3
ICNN	3.9±0.4	14.3±0.5	46±2	5

Ground truth is known: MSE between the maps over test points

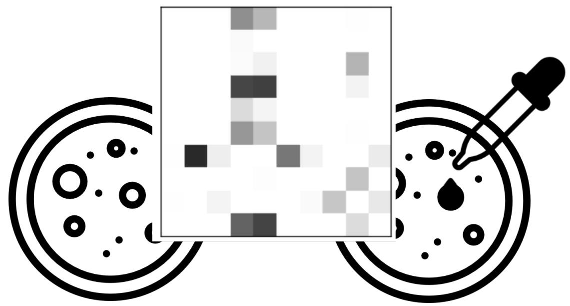
SinkDiv between the predicted target and the test target point cloud

[more in the paper]

Coupling recovery

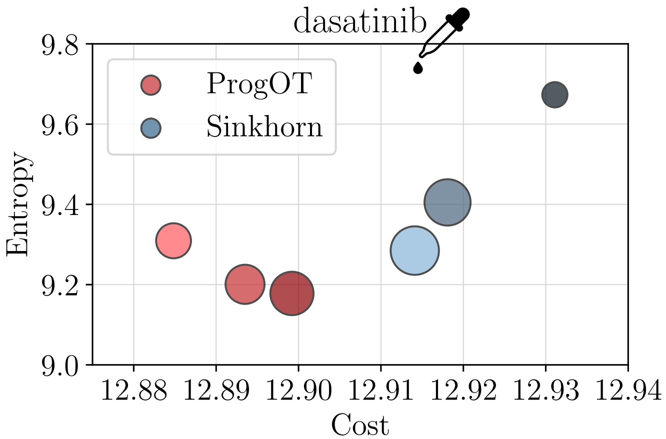
ProgOT attains lower OT cost and lower entropy, at a lower computational cost.

Single-Cell data (4i)

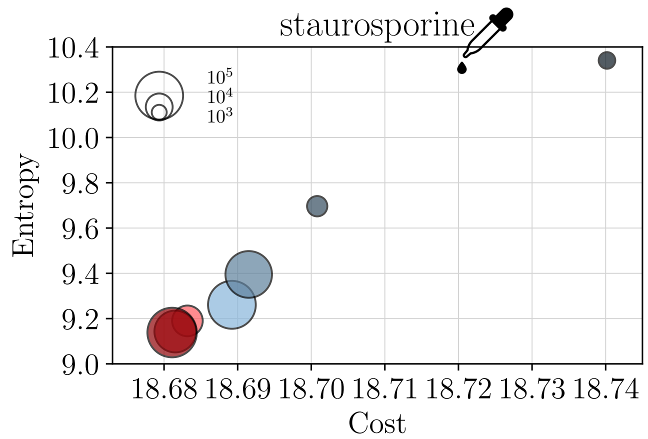


$$\text{Cost} = \sum_{i,j\in[n]} \hat{\pi}_{i,j} h(x_i - y_j)$$

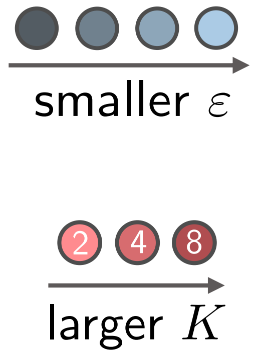
$$\text{Entropy} = \sum_{i,j\in[n]} -\hat{\pi}_{i,j} \log \hat{\pi}_{i,j}$$



$h(\cdot) = \frac{1}{2}\|\cdot\|_2^2$



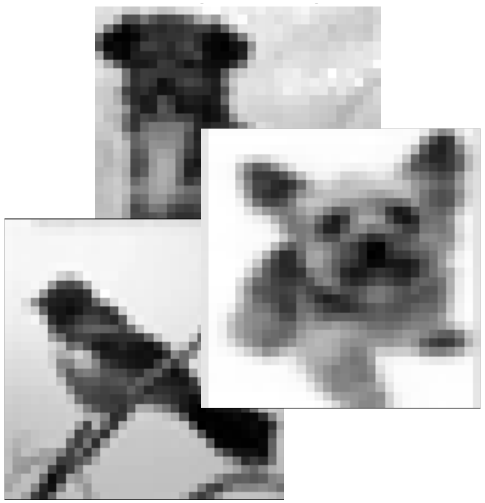
$h(\cdot) = \frac{1}{1.5}\|\cdot\|_{1.5}^{1.5}$

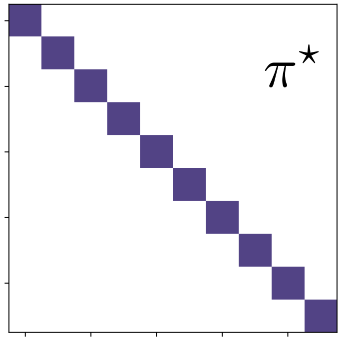


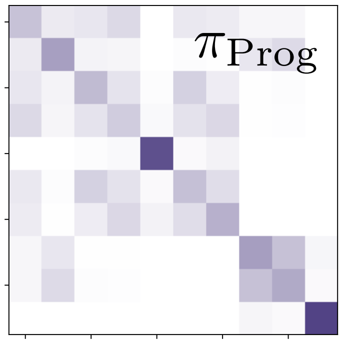
[more in the paper]

Scalability

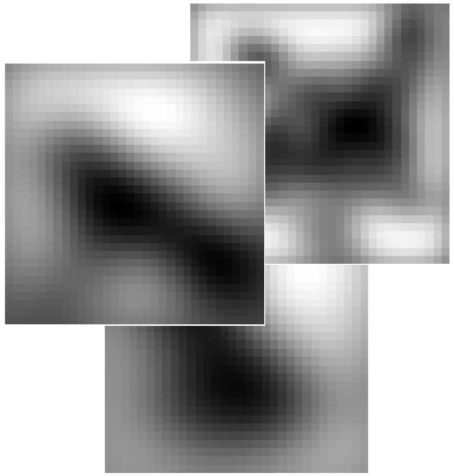
60k CIFAR10 images







blurred CIFAR
(gaussian kernel $\sigma = 4$)



σ		2	4
Sinkhorn	$\mathrm{Tr}(\pi_\varepsilon)$	0.9999	0.9954
	$\mathrm{KL}(\pi^\star \| \pi_\varepsilon)$	0.00008	0.02724
	# iterations	10	2379
ProgOT	$\mathrm{Tr}(\pi_{\mathrm{Prog}})$	1.000	0.9989
	$\mathrm{KL}(\pi^\star \| \pi_{\mathrm{Prog}})$	0.00000	0.00219
	# iterations	40	1590

impossible using neural OT solvers

15' to de-blur CIFAR10
(\w sharding on 8gpus)

The bigger picture

ProgOT

- Light, off-the-shelf, competitive baseline
- Blending static and dynamic views of OT
- [paper], [JAX tutorial]

Follow-ups

- Scaling Limit
- Continuous time extension/implications

OT Applications

- Unbalanced OT
- The Schrödinger Bridge

Other Applications

- Drug purtubations
- Robust generation
- Preference Learning

Thank You

References

1. C. Bunne, et al. Learning single-cell perturbation responses using neural optimal transport. Nature Methods, 2023.
2. C. Bunne, et al. Optimal transport for single-cell and spatial omics. *Nature Reviews Methods Primers* 4, 2024.
3. I. Melnyk, et al. Distributional Preference Alignment of LLMs via Optimal Transport. *arXiv,* 2024.
4. R. Liu, et al. Zero-shot preference learning for offline RL via optimal transport." *arXiv,* 2023.
5. Munos, Rémi, et al. "Nash learning from human feedback." *arXiv arXiv,* 2023.
6. Neu, Gergely, et al. "A unified view of entropy-regularized markov decision processes." *arXiv,* 2017.