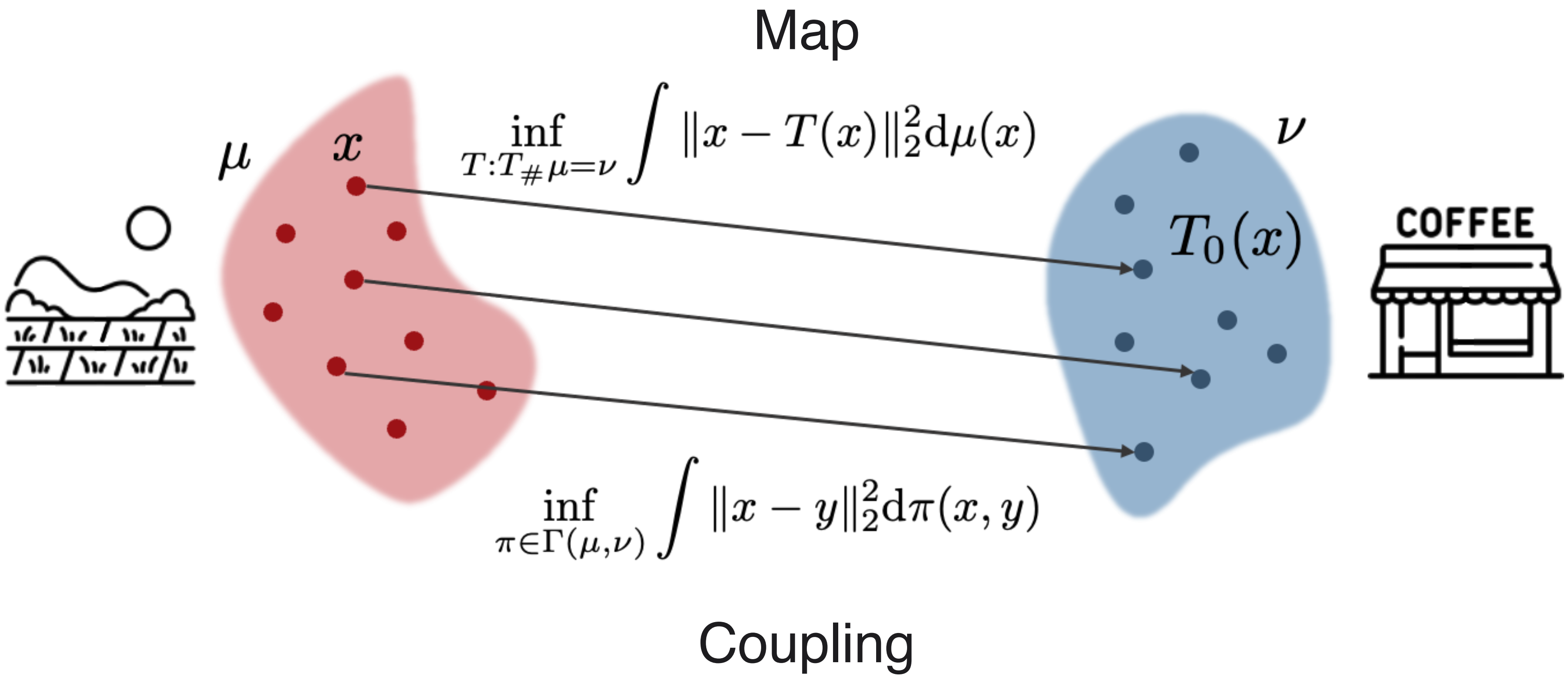
Progressive Entropic Optimal Transport

Parnian Kassraie, Aram-Alexandre Pooladian, Michal Klein, James Thornton, Jonathan Niles-Weed, Marco Cuturi

NeurIPS I Apple I December 2024

Optimal Transport



How do you solve it?

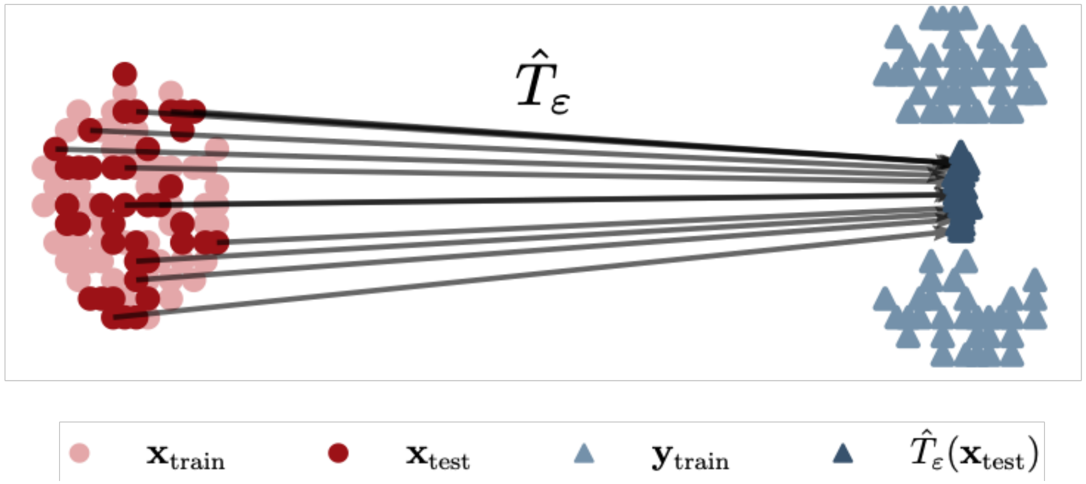
In full generality, OT does not have a solution or is very tough to solve.
Entropic OT adds a regularisation term to make things better.

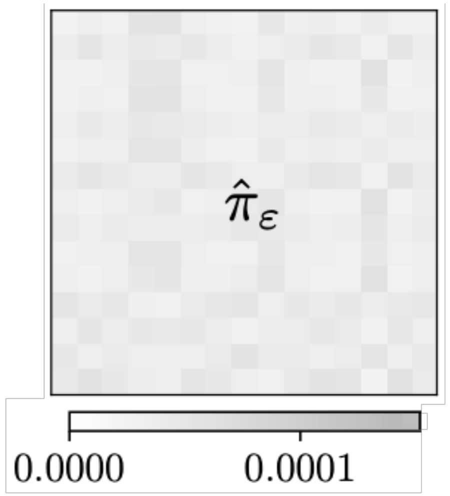
$$\inf_{\pi \in \Gamma(\nu,\mu)} \int \|x-y\|_2^2 \, \mathrm{d}\pi(x,y) + \varepsilon \, \mathrm{D_{KL}} \left(\pi || \mu \otimes \nu\right)$$

Sinkhorn's algorithm solves this and returns a map and a couplings

Small ε : the algorithm does not converge.

Large ε

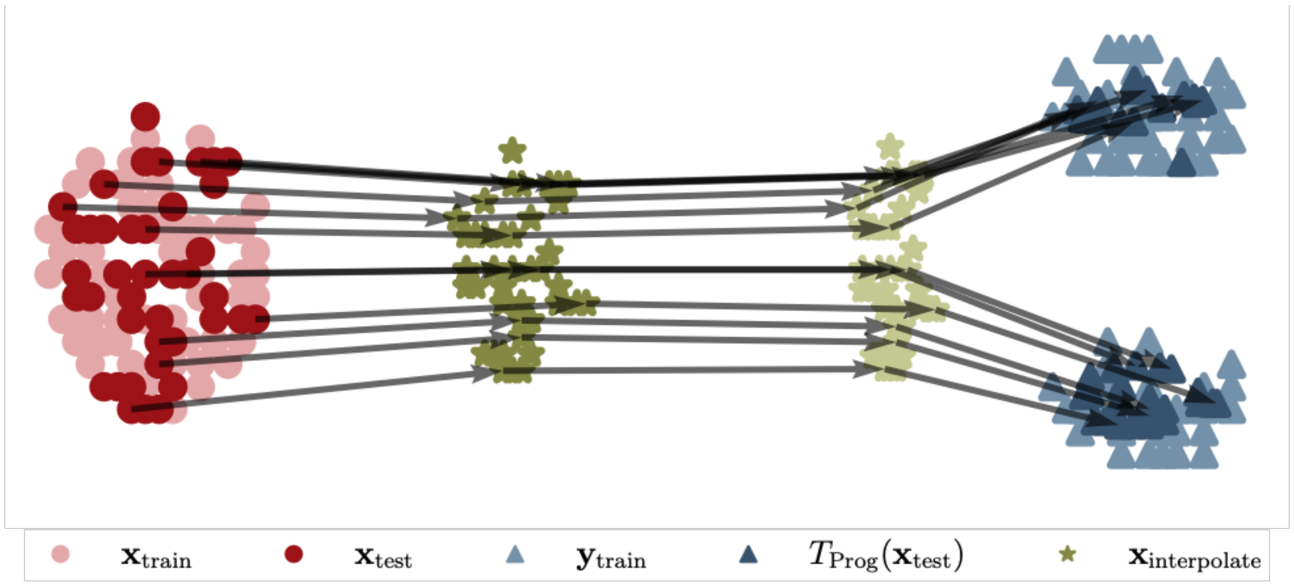


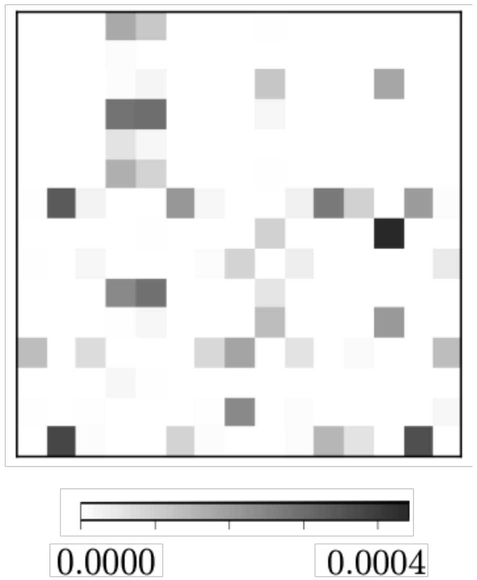


Our solution: ProgOT

Blend the static OT problem with the dynamic perspective

Solve a series of Entropic OT problems, with reduced sensitivity to ε





Theoretical Guarantee

T_0: OT map between μ & ν $\qquad$ $T^{(K)}_{\mathrm{Prog}}$: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 $T^{(K)}_{Prog}$ satisfies

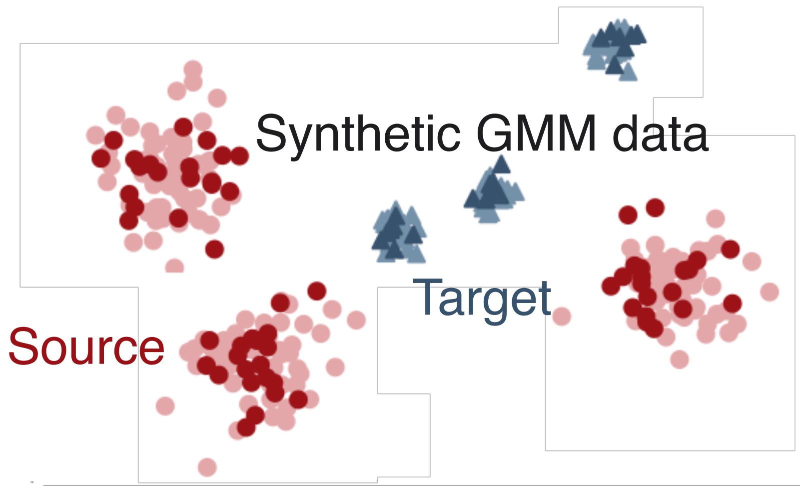
$$\mathbb{E}\left\| T^{(k)}_{Prog} - T_0 \right\|^2_{L^2(\mu)} \lesssim n^{-\frac{1}{d}} ,$$

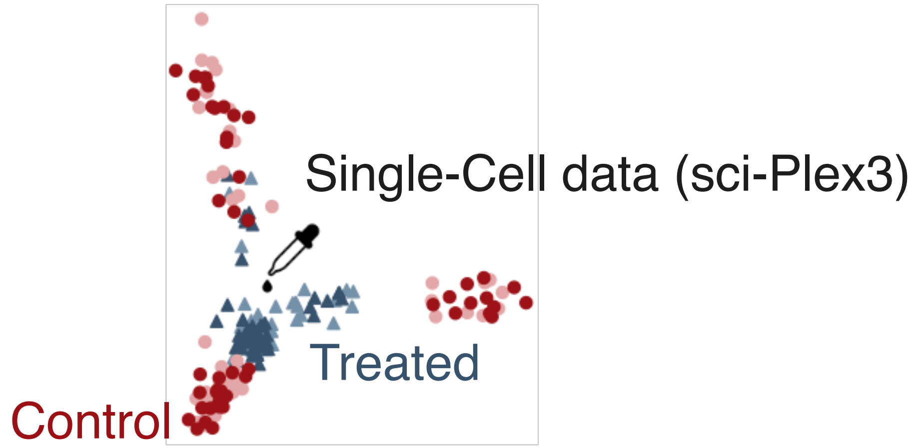
under regularity assumptions on μ, ν, and the true map T_0.

[more in the paper]

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

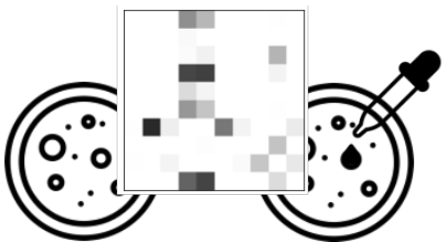
Drug	Hesperadin			5-drug rank
d_{PCA}	16	64	256	
ProgOT	**3.7**±0.4	**10.1**±0.4	**23.1**±0.4	**1**
EOT	4.1±0.4	10.4±0.5	26±1.3	2
Debiased EOT	4.0±0.5	15.2±0.6	41±1.1	4
Monge Gap	3.7±0.5	11.0±0.5	36±1.1	3
ICNN	3.9±0.4	14.3±0.5	46±2	5

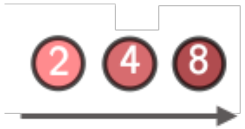
[more in the paper]

Coupling Recovery

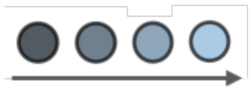
ProgOT attains lower OT cost and lower entropy, at a lower computational cost.

Single-Cell data (4i dataset)

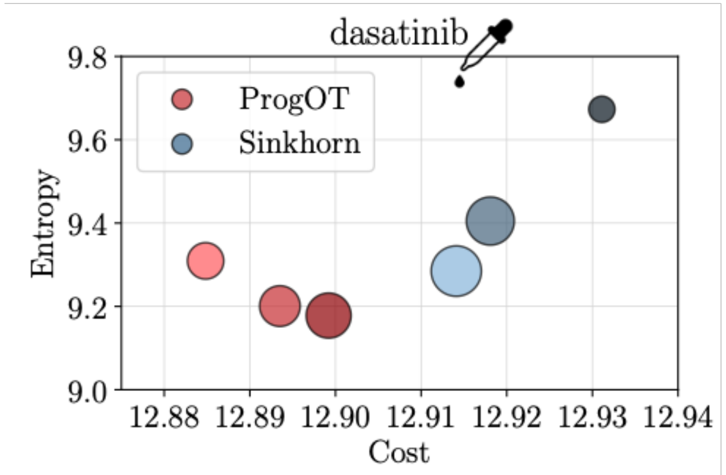




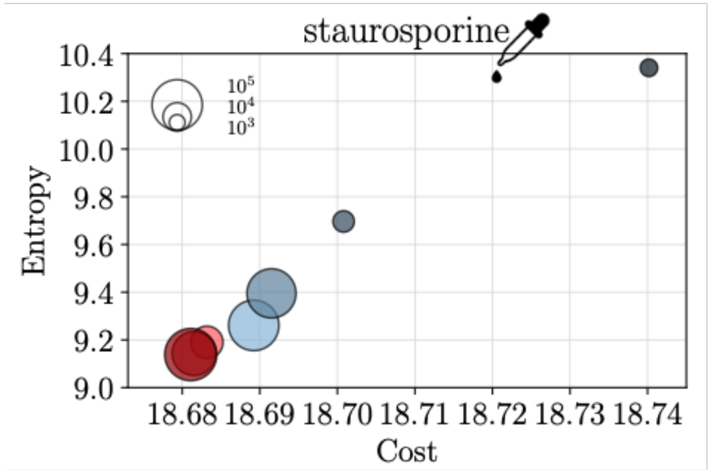
More iterates



Smaller ε



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

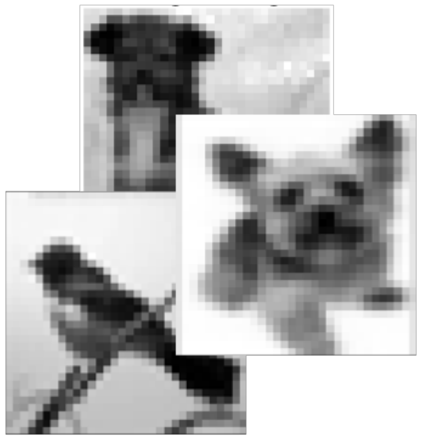


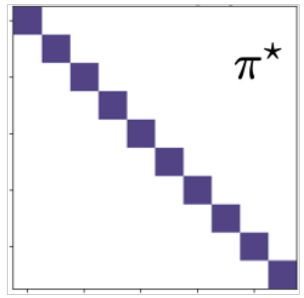
$$h(\cdot) = \frac{1}{1.5}\|\cdot\|_{1.5}^{1.5}$$

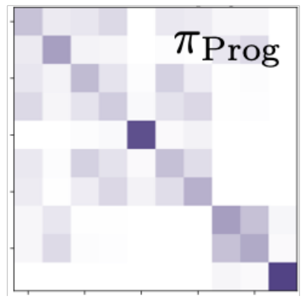
Scalability

ProgOT scales well to large-sample problems in high dimensions.

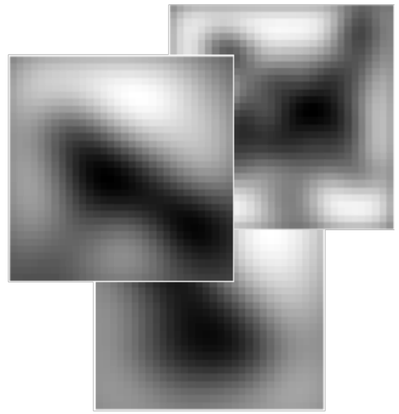
60k CIFAR10 images







Blurred CIFAR



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

15 minutes to de-blur CIFAR10
(with sharding on 8 gpus)

See you at poster session 2!

Wed 11 Dec 16:30-19:30