Neural Optimal Transport with Lagrangian Costs

paper

Aram-Alexandre Pooladian NYU, Meta AI
Carles Domingo-Enrich NYU, Meta AI
Ricky T. Q. Chen Meta AI
Brandon Amos Meta AI

Optimal transport (OT) with Lagrangian costs

📚 *Optimal Transport: Old and New*. Cedric Villani, 2008; *Computational Optimal Transport*. Gabriel Peyré and Marco Cuturi, 2018.

For a cost function $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$, the (dual) **transport problem** is given by

$$\mathrm{OT}_c(\mu,\nu) \coloneqq \sup_{g\in L^1(\nu)} \int g^c(x)\,\mathrm{d}\mu(x) + \int g(y)\,\mathrm{d}\nu(y)$$

c-transform: $g^c(x) \coloneqq \inf_{y\in\mathcal{Y}} J(y;x)$ where $J(y;x) \coloneqq c(x,y) - g(y)$

Lagrangian costs: general-purpose cost function given by an **optimization sub-routine** over curves

$$c(x,y) \coloneqq \inf_{\gamma\in\mathcal{C}(x,y)} E(\gamma;x,y) \quad E(\gamma;x,y) \coloneqq \left\{\int_0^1 \mathcal{L}(\gamma_t,\dot\gamma_t,t)\,\mathrm{d}t\right\}$$

Encompasses ℓ_p norms, barrier functions, non-Euclidean metrics (geodesics), and more

Examples

1) Euclidean kinetic	2) Euclidean kinetic and potential	3) Riemannian kinetic
$\mathcal{L}(\gamma_t,\dot\gamma_t,t) \coloneqq \frac{1}{2}\lVert\dot\gamma_t\rVert^2$,	$\mathcal{L}(\gamma_t,\dot\gamma_t,t) \coloneqq \frac{1}{2}\lVert\dot\gamma_t\rVert^2 - U(\gamma_t)$,	$\mathcal{L}(\gamma_t,\dot\gamma_t,t;A) = \frac{1}{2}\lVert\dot\gamma_t\rVert^2_{A(\gamma_t)}$.
c becomes the squared Euclidean distance		c becomes the squared geodesic distance

OT map for general costs: Our goal is to learn $\hat y(x;c,g) \coloneqq \operatorname{argmin}_{y\in\mathcal{Y}}\{c(x,y) - g(y)\}$.

❗**Challenges:** computing (1) the cost c, (2) the c-transform, (3) the optimal potential g
Our approach: approximate (1), (2), (3) with neural networks (obviously!)

Neural OT with Lagrangian Costs

📚 *Deep generalized Schrödinger bridge*. Liu et al., NeurIPS 2023; *Neural Lagrangian Schrödinger bridge*. Koshizuka and Sato, ICLR 2023; *Optimal transport mapping via input convex neural networks* Makkuva et al., ICML 2020; *Wasserstein-2 Generative Networks*, Korotin et al., ICLR 2021; *On amortizing convex conjugates for optimal transport*. Amos, ICLR 2023; *Tutorial on amortized optimization*. Amos, FnT in ML, 2023.

Parametrization with neural networks: Optimize

(1) Lagrangian path φ_η

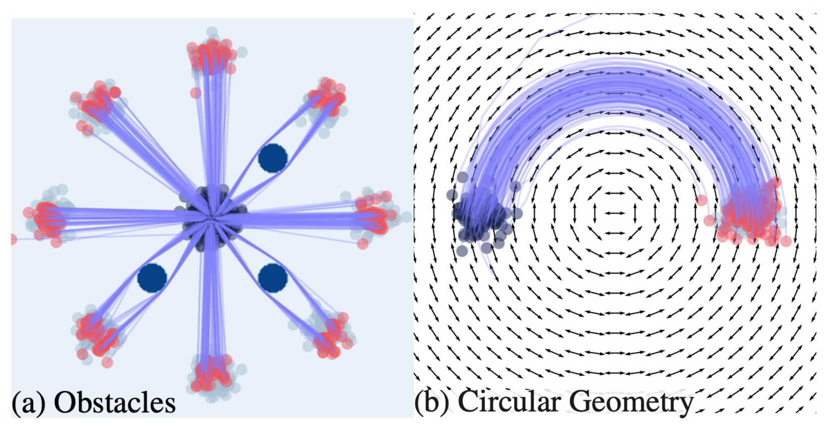
$$\min_\eta \int E(\varphi_\eta;x,\hat y(x))\,\mathrm{d}\mu(x)\ .$$

(2) OT map y_ϕ

$$\min_\phi \int \lVert\hat y(x) - y_\phi(x)\rVert\,\mathrm{d}\mu(x)\ .$$

(3) potential g_θ

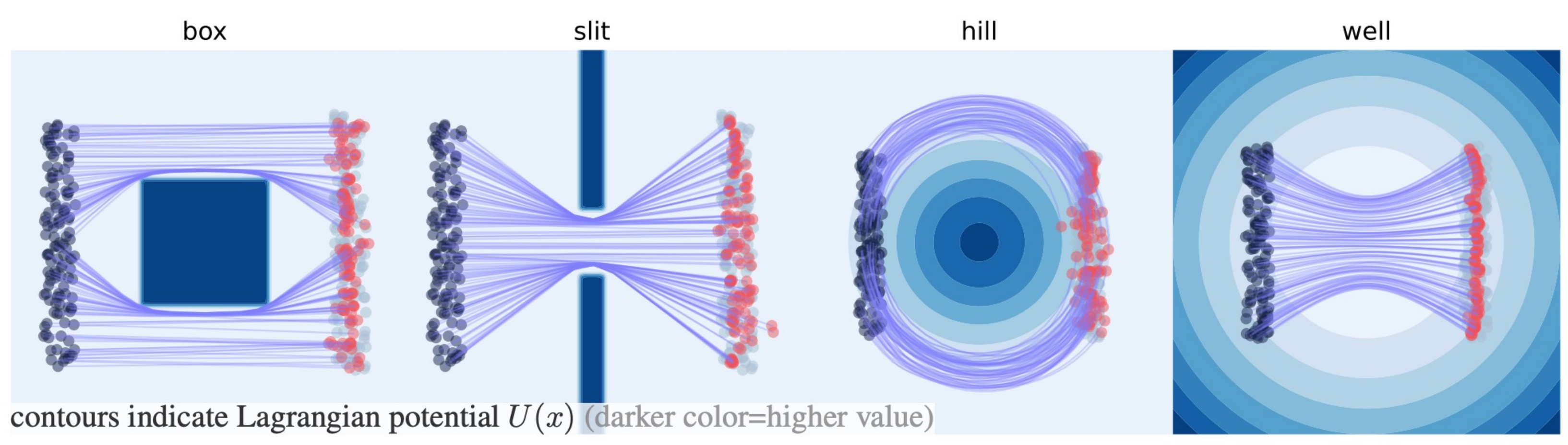
$$\ell_{\mathrm{dual}}(\theta) \coloneqq \int g_\theta^c(x)\,\mathrm{d}\mu(x) + \int g_\theta(y)\,\mathrm{d}\nu(y)$$



samples (■ source ■ target ■ push-forwards) ■ transport paths

Algorithm 1 Neural Lagrangian Optimal Transport

inputs: measures μ and ν, Kantorovich potential g_θ, c-transform predictor y_ϕ, and spline predictor φ_η
while unconverged **do**
  sample batches $\{x_i\}_{i=1}^N \sim \mu$ and $\{y_i\}_{i=1}^N \sim \nu$
  obtain the amortized c-transform predictor $y_\phi(x_i)$ for $i \in [N]$
  fine-tune the c-transform by numerically solving Eq. (9), warm-starting with $y_\phi(x_i)$
  update the potential with gradient estimate of $\nabla_\theta \ell_{\mathrm{dual}}$ (Eq. (18))
  update the c-transform predictor y_ϕ using a gradient estimate of Eq. (20)
  update the spline predictor φ_η using a gradient estimate of Eq. (23)
end while
return optimal parameters θ, ϕ, η



contours indicate Lagrangian potential $U(x)$ (darker color=higher value)

■ source samples ■ target samples ■ push-forward samples ■ transport paths

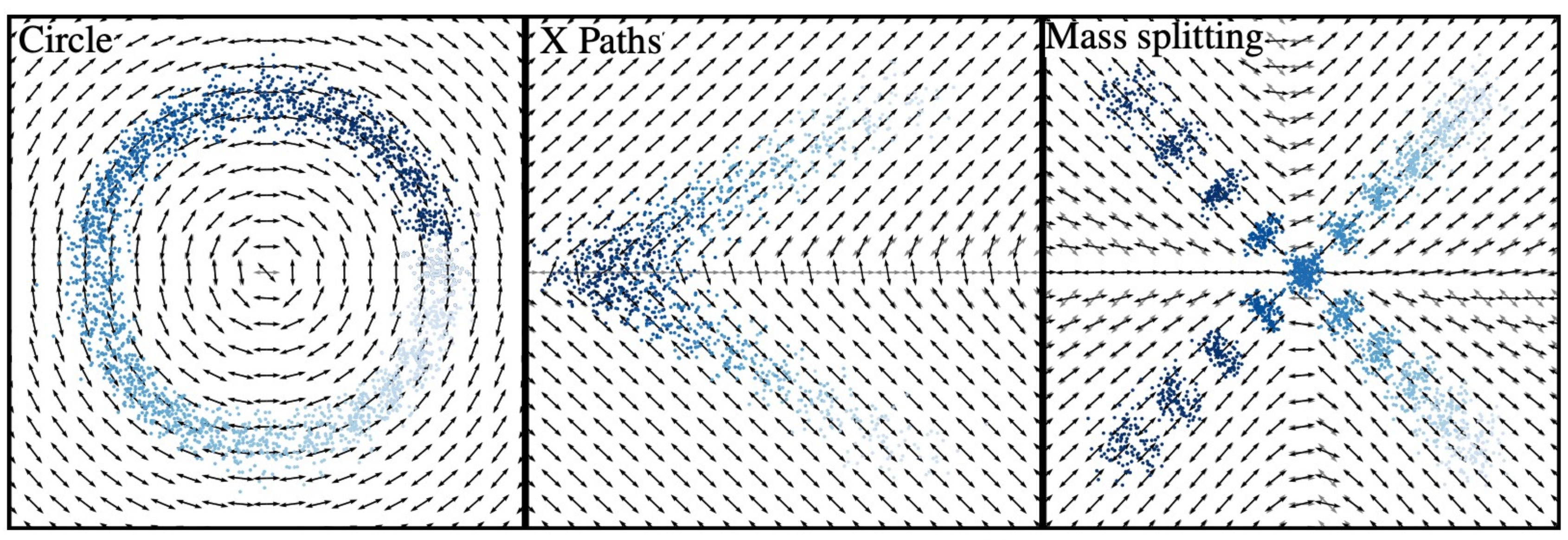
Metric learning with Lagrangian OT

📚 *Riemannian Metric Learning via Optimal Transport*. Scarvelis and Solomon, ICLR 2023.

Task: learn the underlying metric with pairs of probability measures

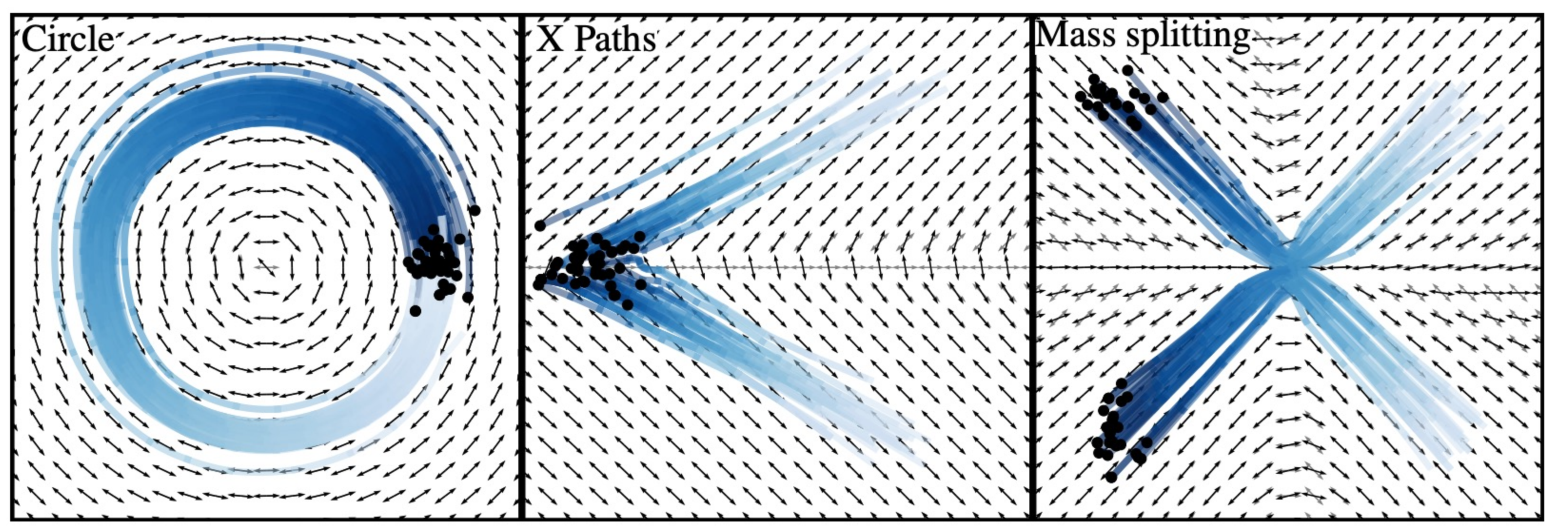
Table 1. Alignment scores ℓ_{align} for metric recovery in Fig. 4. (higher is better)

	Circle	Mass Splitting	X Paths
Scarvelis and Solomon (2023)	0.995	0.839	0.916
Our approach	**0.997 ± 0.002**	**0.986 ± 0.001**	**0.957 ± 0.001**



(smallest) eigenvectors of A (■ learned metric ■ ground-truth) ■ data (lighter colors=later time)

Figure 3. We successfully recover the metrics on the settings from Scarvelis and Solomon (2023).



■ samples from initial measure ■ geodesics (lighter colors=later time)

Figure 4. Our transport geodesics are able to reconstruct continuous versions of the original data that can predict the movement of individual particles given only samples from the first measure.