Improving Neural Optimal Transport via Displacement Interpolation

2024 KIAS CAINS Fall Workshop

JaeWoong Choi

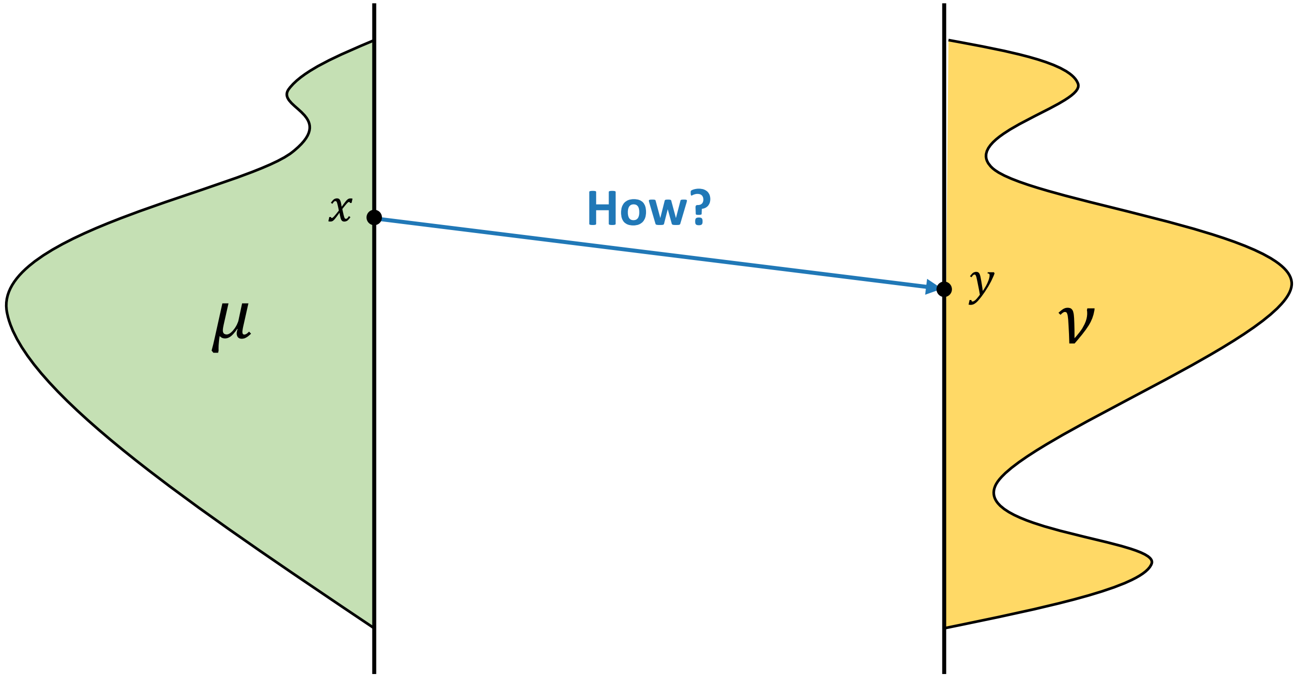
Center for AI and Natural Sciences, KIAS

KOREA INSTITUTE FOR ADVANCED STUDY

CONTENTS

Optimal Transport (OT) Problem

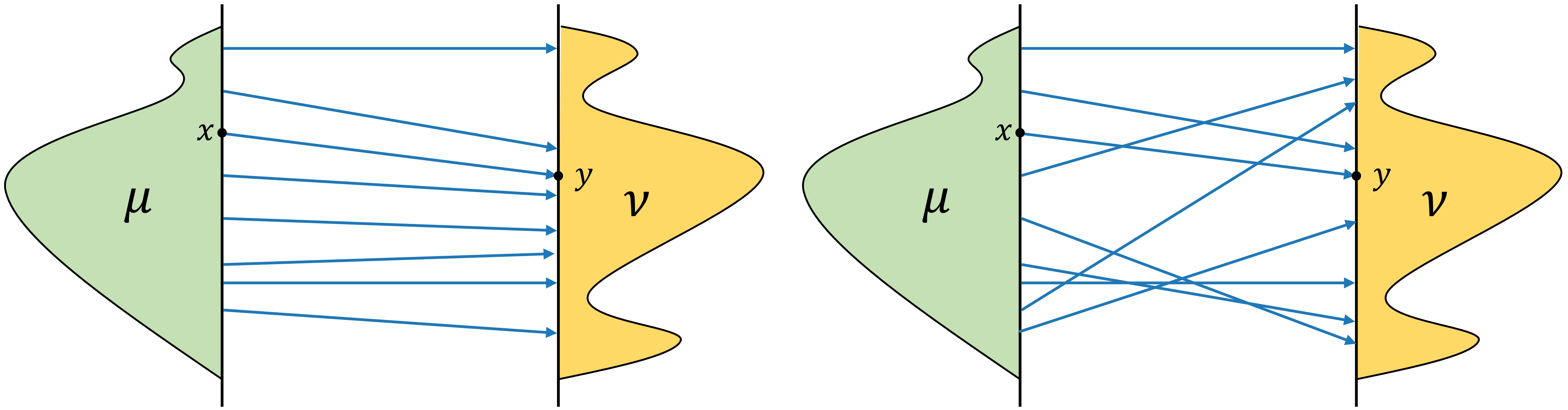
- Optimal Transport refers to the most **cost-minimizing way to transport source distribution μ to target ν**.



[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.

Optimal Transport (OT) Problem

- Optimal Transport refers to the most **cost-minimizing way to transport source distribution μ to target ν**.

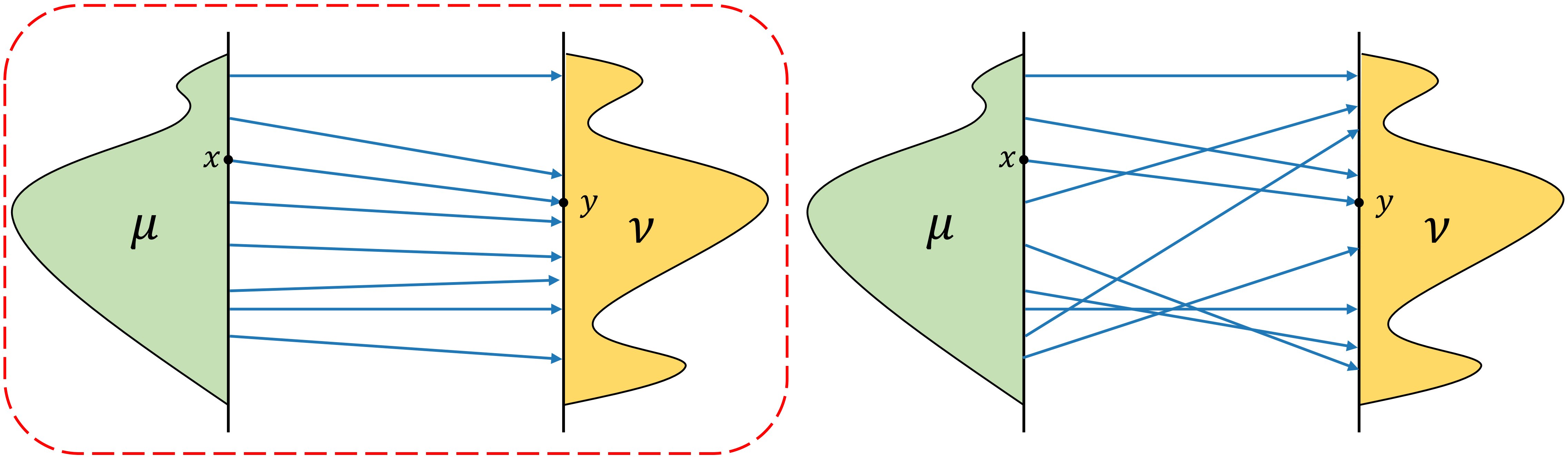


[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.

Optimal Transport (OT) Problem

- Optimal Transport refers to the most **cost-minimizing way to transport source distribution μ to target ν**.

$$c(x, y) = \|x - y\|^2$$



[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.

Optimal Transport (OT) Problem

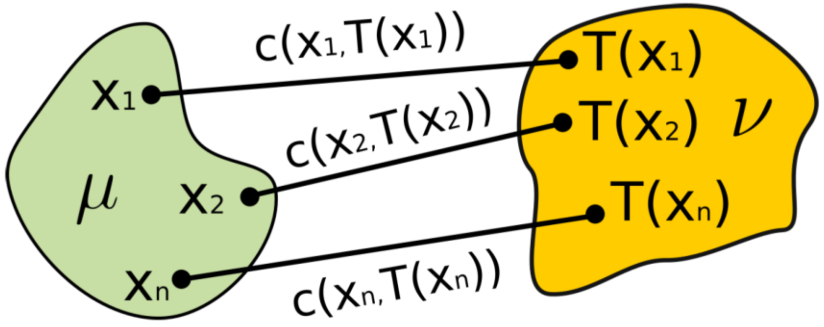
- Optimal Transport refers to the most **cost-minimizing way to transport source distribution μ to target ν**.

- **Monge's Formulation**.

$$C(\mu,\nu) := \inf_{T_{\#}\mu=\nu} \left[\int_{\mathcal{X}} c(x, T(x)) d\mu(x) \right].$$

Transport Map *T*

$$\boldsymbol{x \sim \mu \implies T(x) \sim \nu}$$



Monge's Optimal Transport [2]

[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.

Optimal Transport (OT) Problem

- Optimal Transport refers to the most **cost-minimizing way to transport source distribution μ to target ν**.

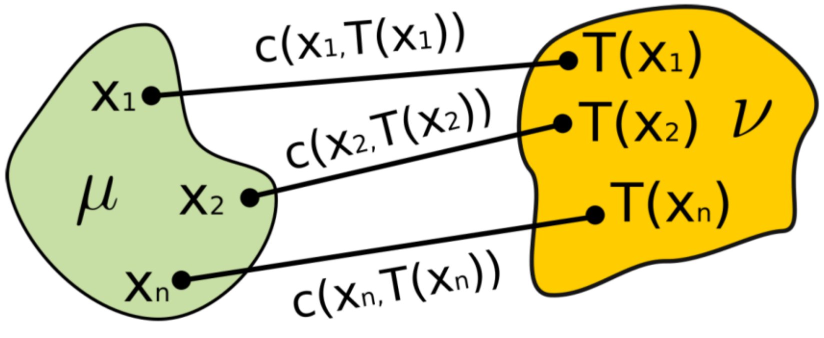
- **Monge's Formulation**.

$$C(\mu,\nu) := \inf_{T_{\#}\mu=\nu} \left[\int_{\mathcal{X}} c(x, T(x)) d\mu(x) \right].$$

Transport Map T

$$\boldsymbol{x \sim \mu \implies T(x) \sim \nu}$$

Optimality by cost-minimizing



Monge's Optimal Transport [2]

[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.

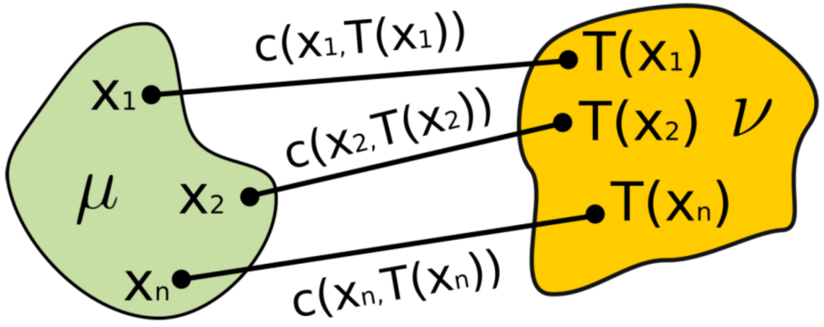
Introduction

Optimal Transport (OT) Problem

- Optimal Transport refers to the most **cost-minimizing way to transport source distribution μ to target ν**.

- **Monge's Formulation**.

$$C(\mu,\nu) := \inf_{T_{\#}\mu=\nu} \left[\int_{\mathcal{X}} c(x, T(x)) d\mu(x) \right].$$



Monge's Optimal Transport [2]

- **Kantorovich's Relaxation.**

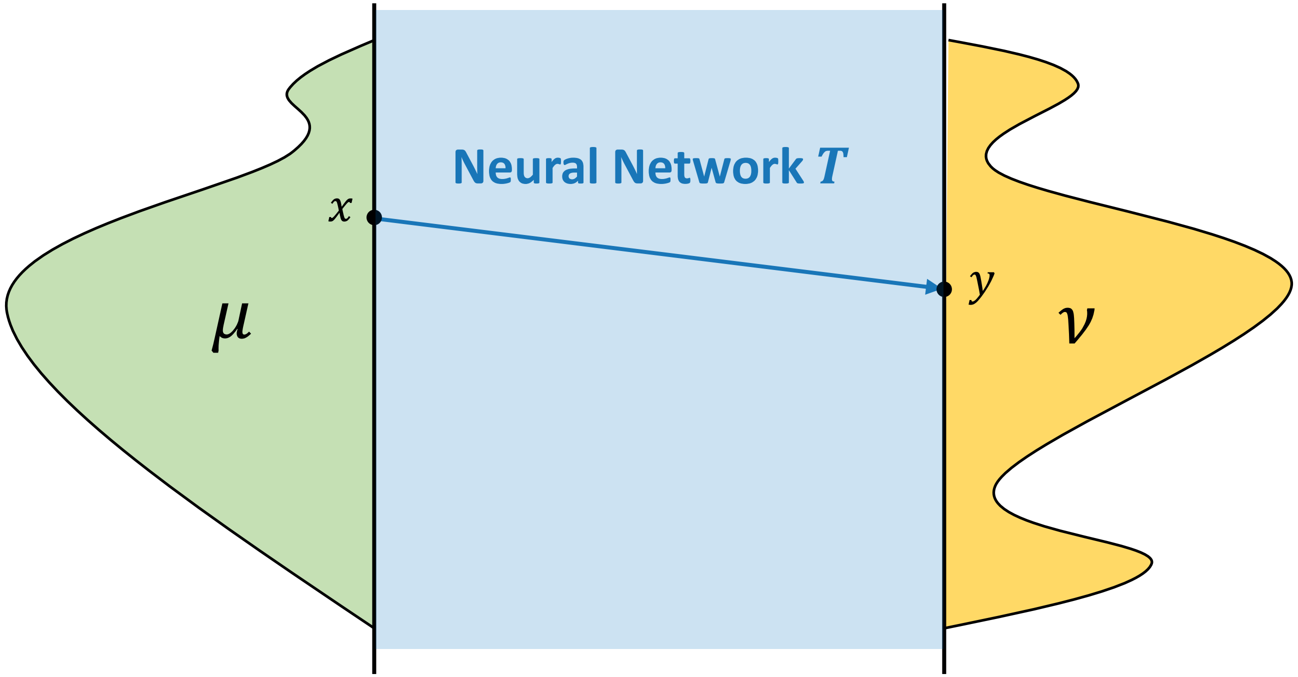
$$C(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \left[\int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi(x, y) \right].$$

where $\Pi(\mu,\nu) \coloneqq$ the set of joint probability distributions on $\mathcal{X} \times \mathcal{Y}$ whose marginals are μ and ν.

[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.

Neural Optimal Transport

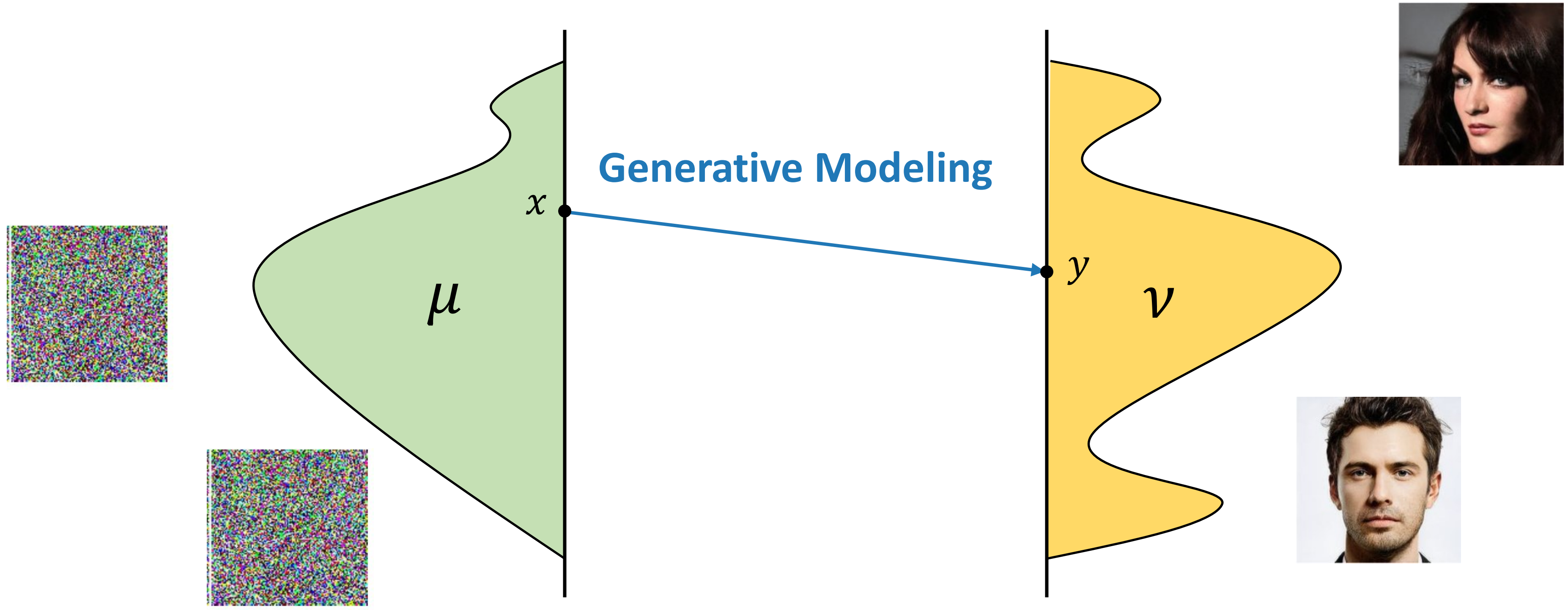
- Today, we focus on approaches for **learning (static) optimal transport maps with neural networks.**



[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps.“ *ICLR*, 2022.

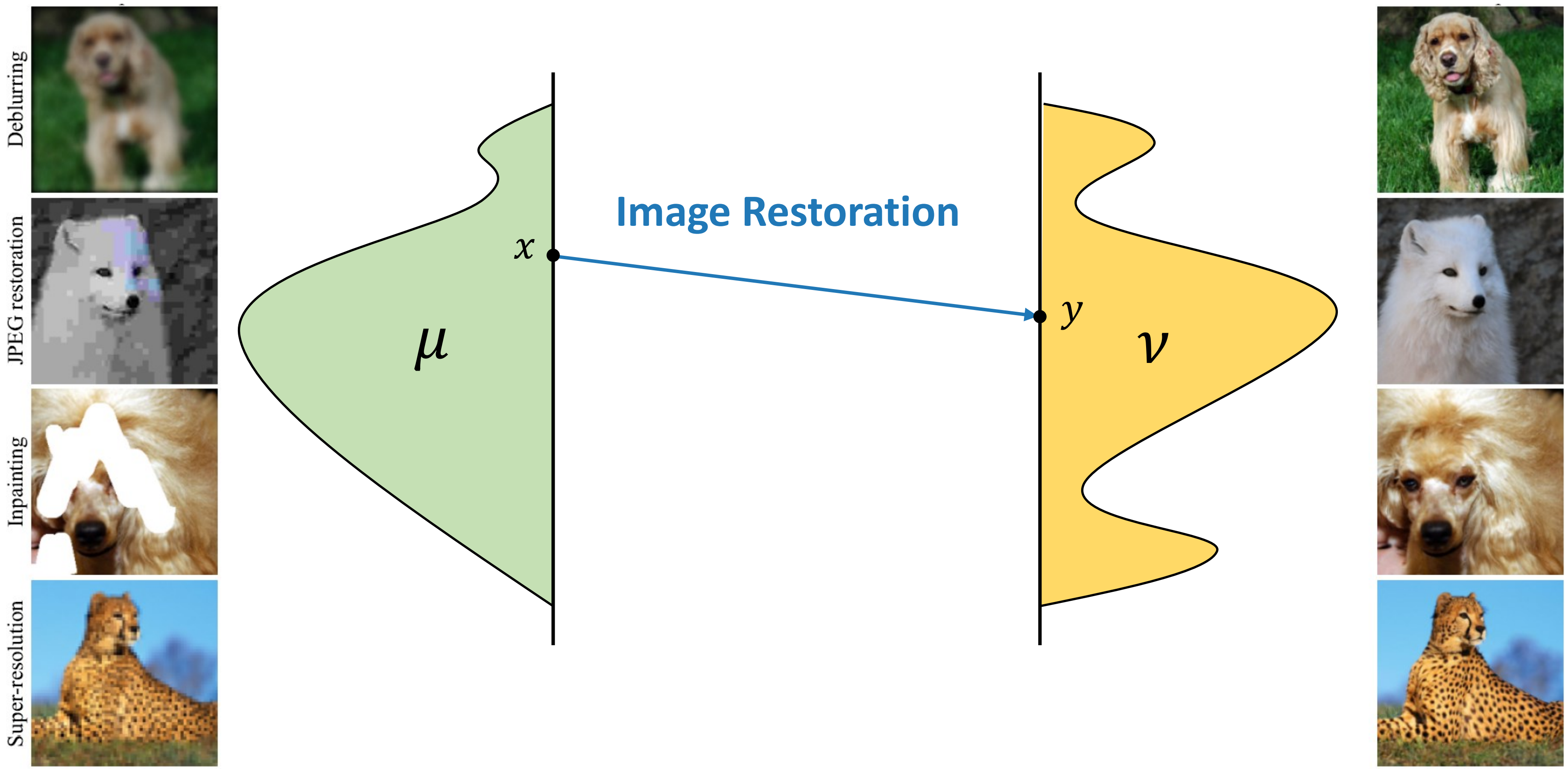
Neural Optimal Transport Applications

- The Neural Optimal Transport can be applied to **any distribution transport applications**.



Neural Optimal Transport Applications

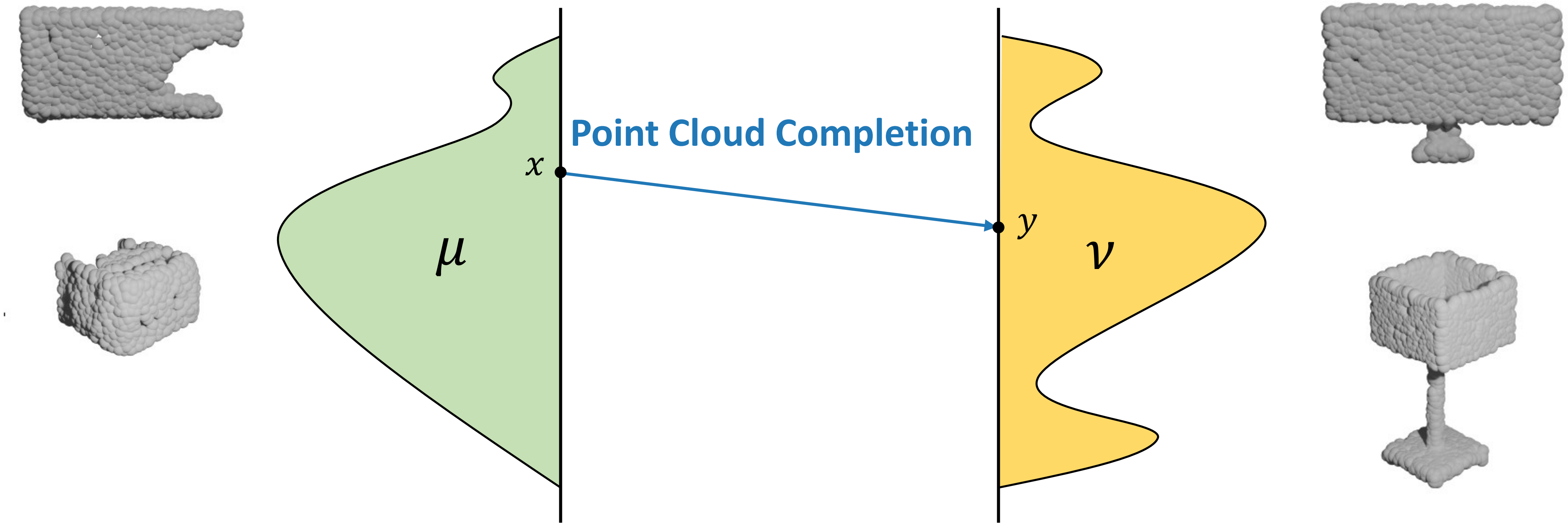
- For example, **Computer Vision Applications**, such as **Image Restoration**.



Liu, Guan-Horng, et al. "I2SB: Image-to-Image Schrödinger Bridge.", *ICML*, 2023.

Neural Optimal Transport Applications

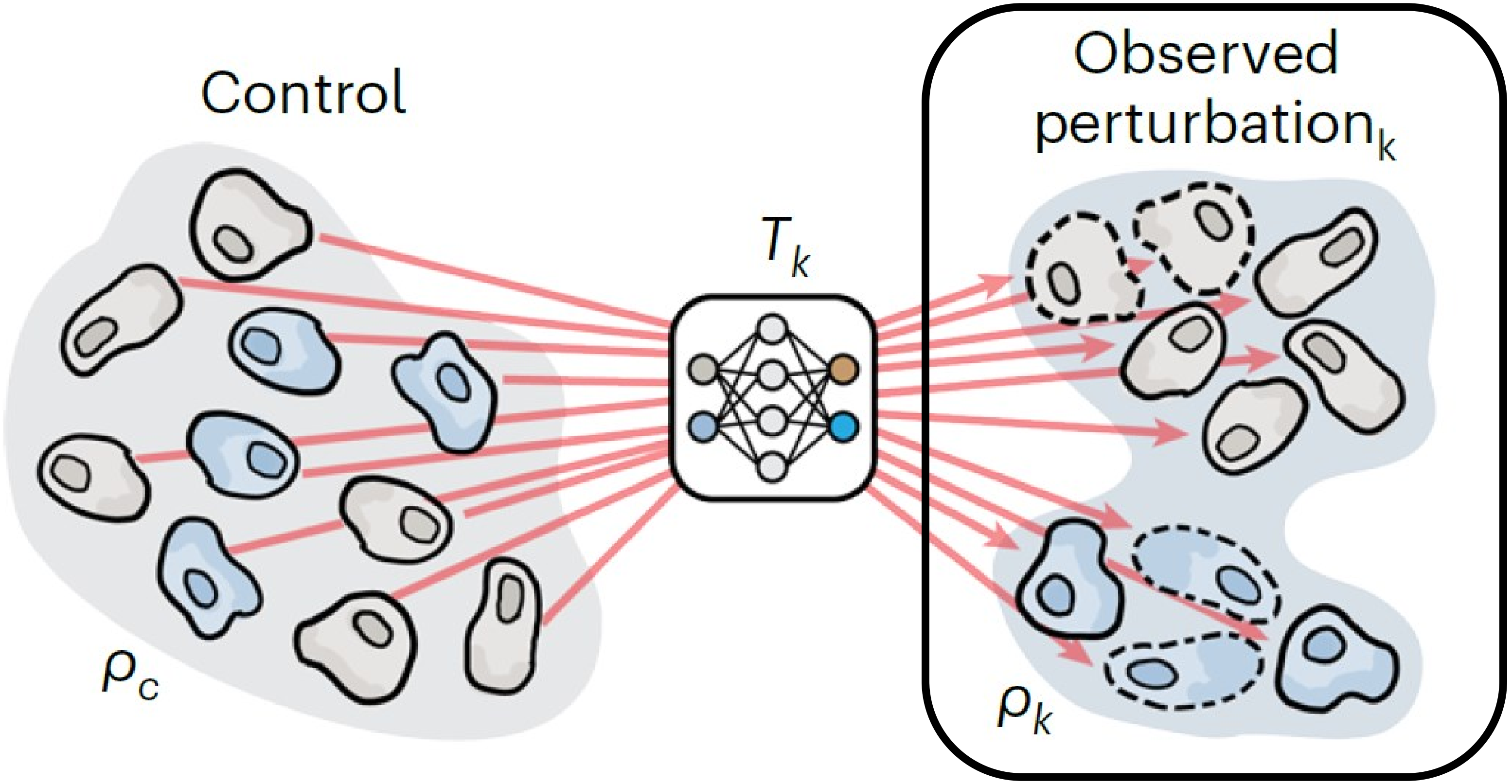
- For example, **Computer Vision Applications**, such as **Point Cloud Completion**.



Lee, Taekyung, et al. "Unsupervised Point Cloud Completion through Unbalanced Optimal Transport.“, *Arxiv*, 2024.

Neural Optimal Transport Applications

- For example, **AI for Science**, such as **Prediction of Single-cell perturbation responses.**
  - What **effect would Drug k have** on these cell populations?

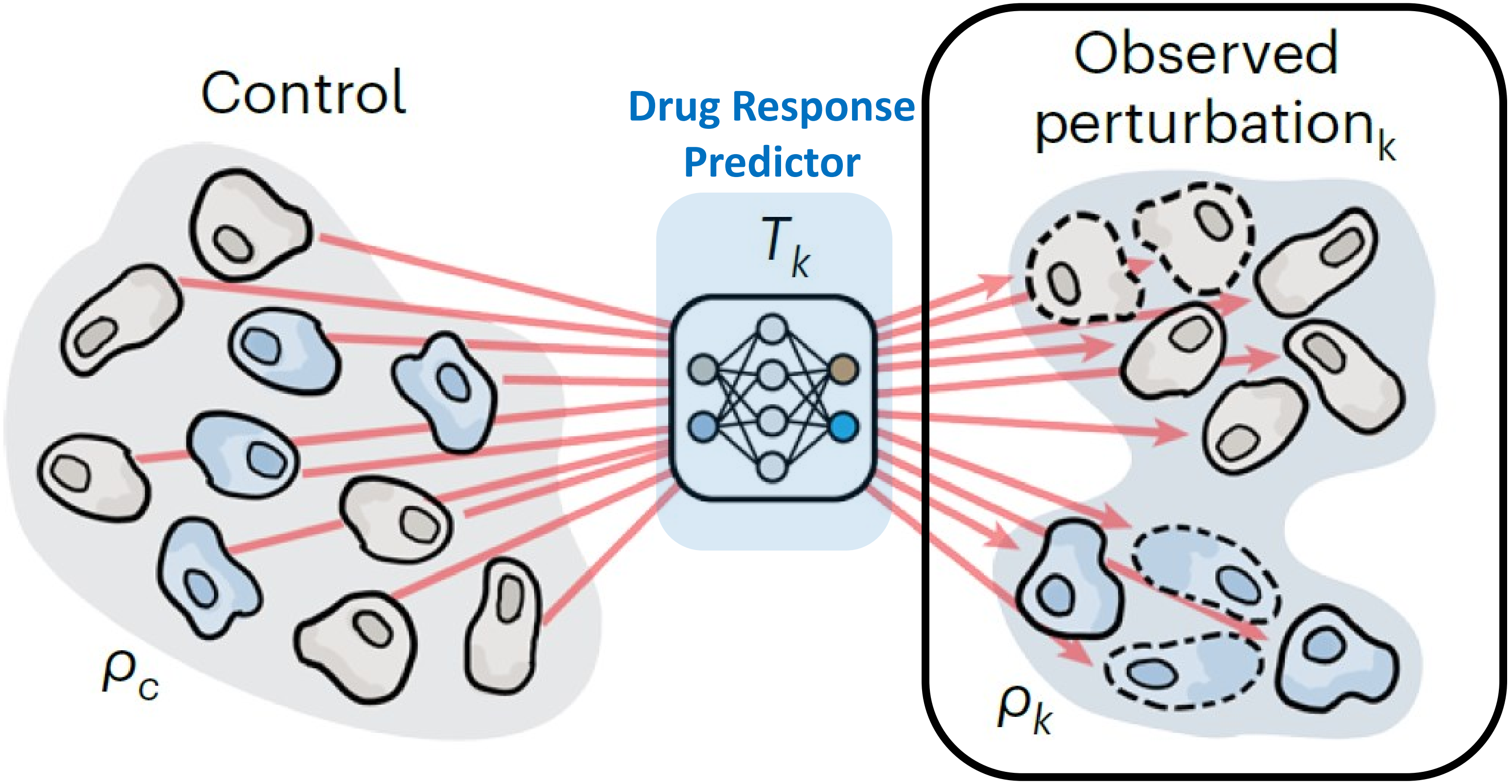


Cells distribution after Drug k Treatment

Bunne, Charlotte, et al. "Learning single-cell perturbation responses using neural optimal transport." *Nature Methods* 20.11 (2023): 1759-1768.

Neural Optimal Transport Applications

- For example, **AI for Science**, such as **Prediction of Single-cell perturbation responses.**
  - What **effect would Drug k have** on these cell populations?



Cells distribution after Drug k Treatment

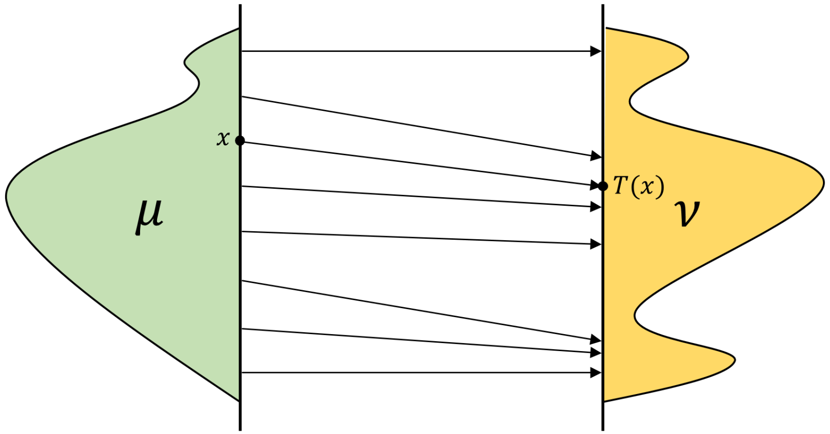
Bunne, Charlotte, et al. "Learning single-cell perturbation responses using neural optimal transport." *Nature Methods* 20.11 (2023): 1759-1768.

Instability Challenges in Neural Static OT

Previous Neural Optimal Transport

Optimal Transport Map (OTM)

- OTM [1, 2] learns the **transport map T** through a max-min formulation.
  - v^c denotes the ***c*-transform** of v, i.e., $v^c(x) = \inf_{y \in \mathcal{Y}} (c(x,y) - v(y))$.



Static Optimal Transport

$$C(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left[\int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\pi(x,y) \right],$$

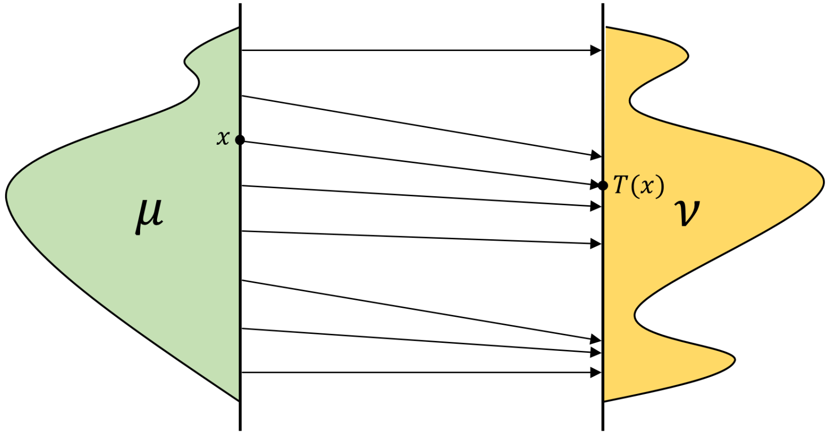
Semi-dual formulation for OT

$$= \sup_{v \in L^1(\nu)} \left[\int_{\mathcal{X}} \underbrace{v^c(x)}_{c\text{-transform of } v} d\mu(x) + \int_{\mathcal{Y}} v(y) d\nu(y) \right]$$

Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.
Fan, Jiaojiao, et al. "Scalable computation of monge maps with general costs." *ICLR Workshop*, 2022.

Optimal Transport Map (OTM)

- OTM [1, 2] learns the **transport map T** through a max-min formulation.
  - v^c denotes the ***c*-transform** of v, i.e., $v^c(x) = \inf_{y \in \mathcal{Y}} (c(x,y) - v(y))$.



Static Optimal Transport

$$C(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left[\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi(x,y) \right],$$

Semi-dual formulation for OT

$$= \sup_{v \in L^1(\nu)} \left[\int_{\mathcal{X}} v^c(x)\, d\mu(x) + \int_{\mathcal{Y}} v(y)\, d\nu(y) \right]$$

***c*-transform of v**

$$\rightarrow \mathcal{L}_{v_\phi, T_\theta} = \sup_{v_\phi} \left[\int_{\mathcal{X}} \inf_{T_\theta} \left[c\left(x, T_\theta(x)\right) - v_\phi\left(T_\theta(x)\right) \right] d\mu(x) + \int_{\mathcal{Y}} v_\phi(y)\, d\nu(y) \right].$$

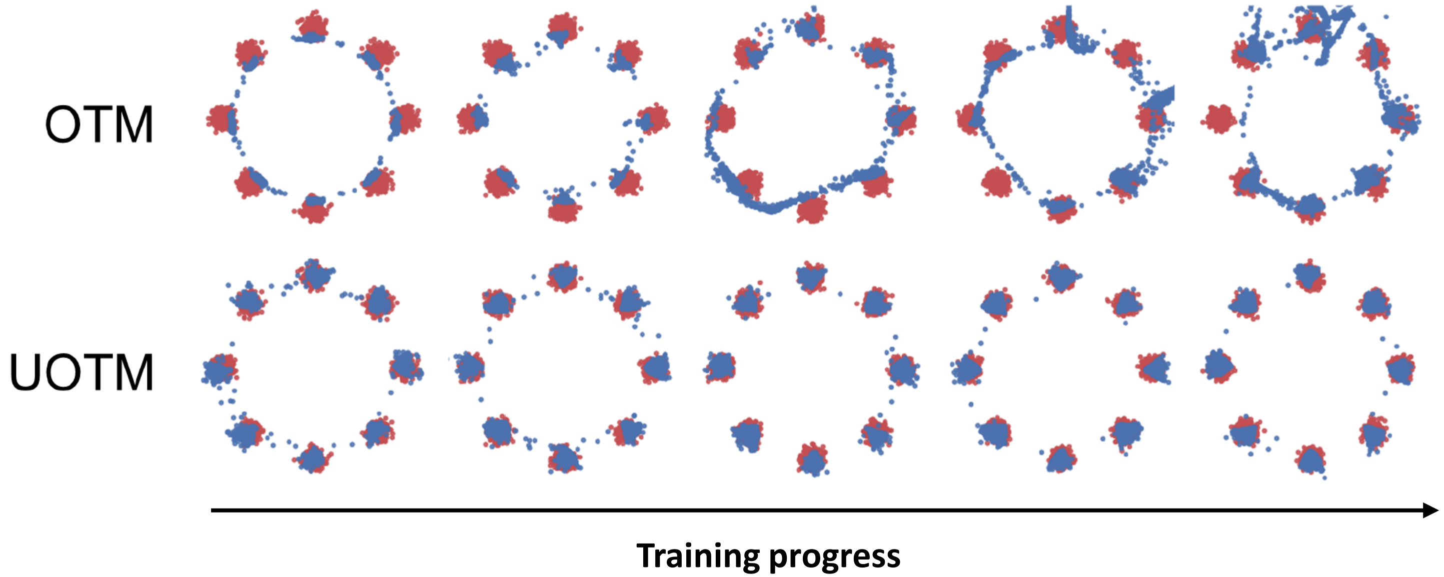
$$T_\theta : \mathcal{X} \to \mathcal{Y},\ x \mapsto \arg\inf_{y \in \mathcal{Y}} \left[c(x,y) - v\left(y\right) \right]$$

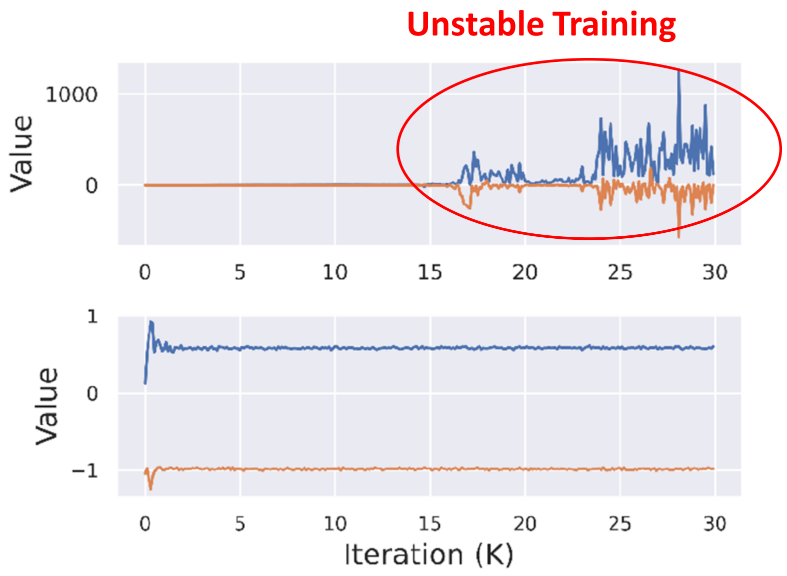
Min-max objective between Transport map T and Potential v

Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.
Fan, Jiaojiao, et al. "Scalable computation of monge maps with general costs." *ICLR Workshop*, 2022.

Instability Challenges in OTM

- **Unstable training dynamics and high sensitivity to hyperparameters in previous approaches**

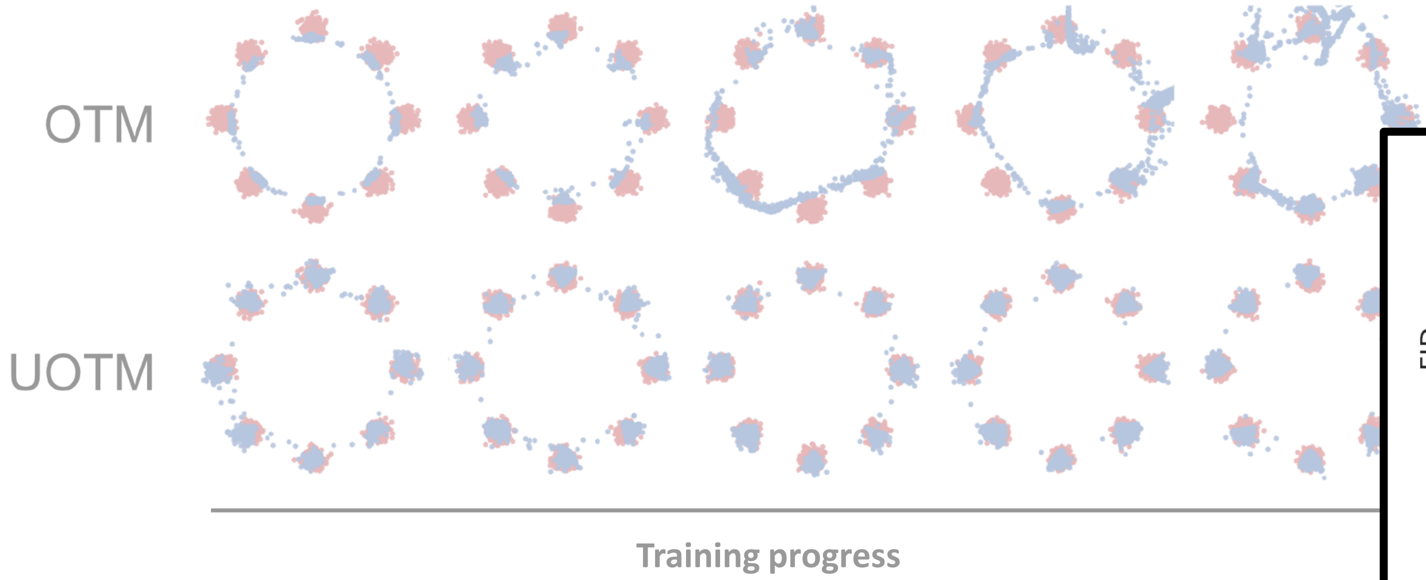




Choi, Jaemoo, Jaewoong Choi, and Myungjoo Kang. "Analyzing and Improving Optimal-Transport-based Adversarial Networks." *ICLR,* 2024.

Instability Challenges in OTM

- **Unstable training dynamics and high sensitivity to hyperparameters in previous approaches**



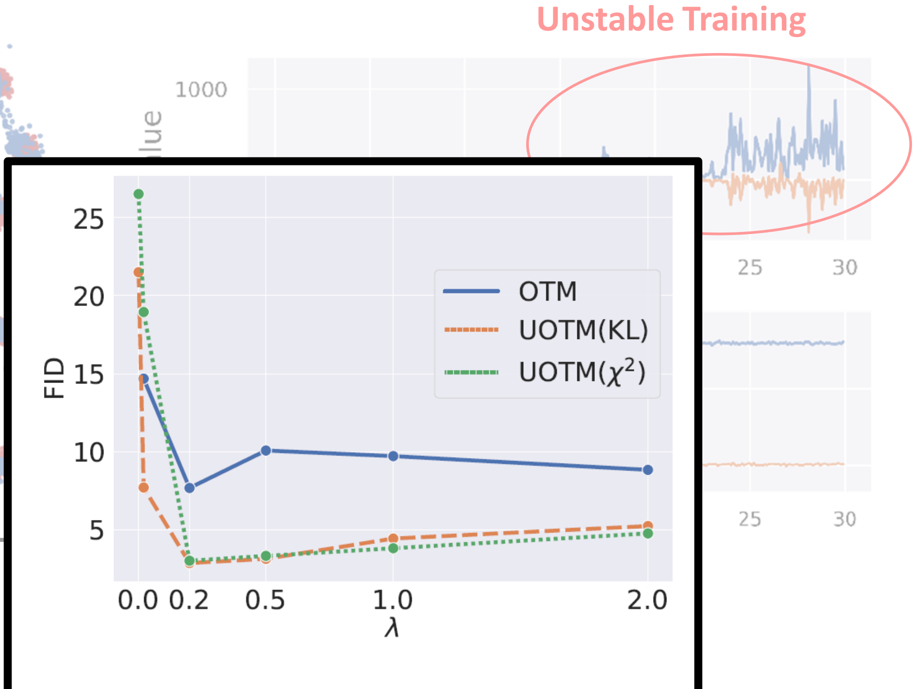


Figure 6: **Ablation Study on Regularizer Intensity λ.**

Choi, Jaemoo, Jaewoong Choi, and Myungjoo Kang. "Analyzing and Improving Optimal-Transport-based Adversarial Networks." *ICLR,* 2024.

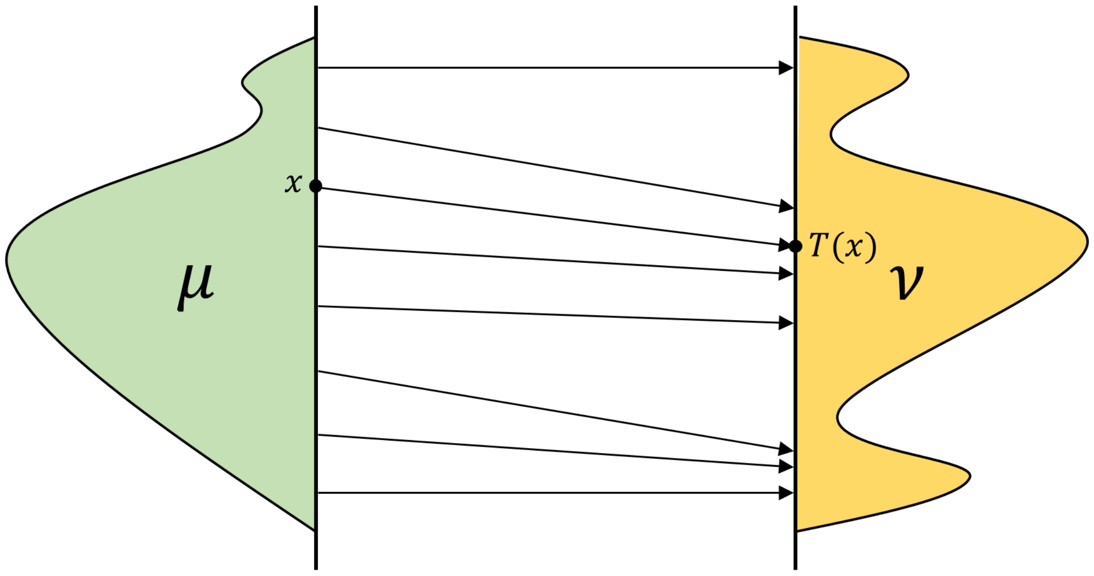
Displacement Interpolation Optimal Transport Model

Dynamic Optimal Transport and Displacement Interpolation

Dynamic Optimal Transport

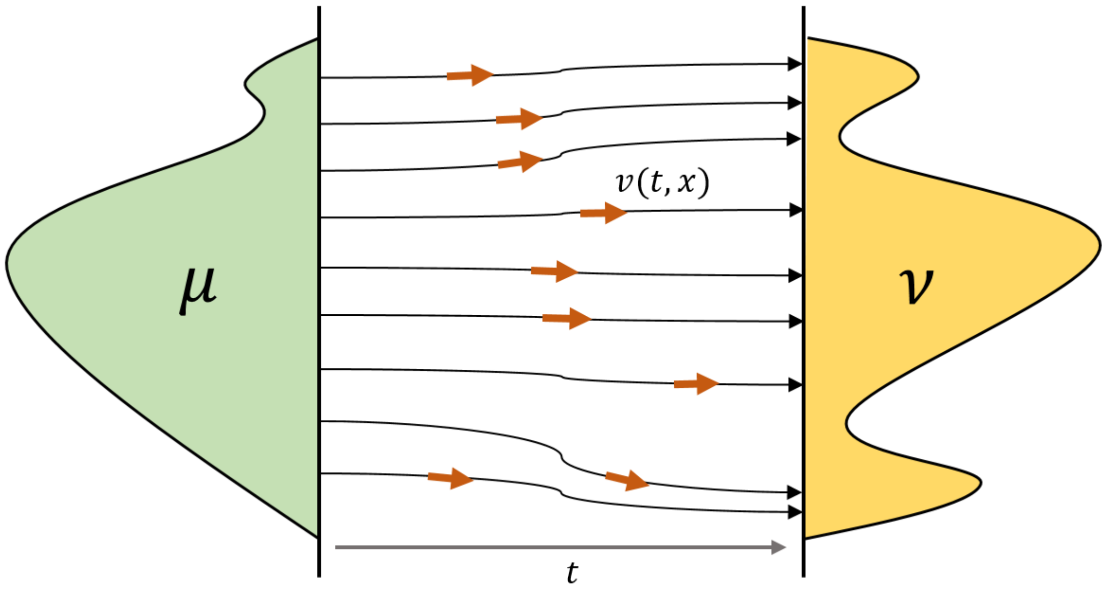
- **The dynamic optimal transport problem tracks the continuous evolution of μ to ν.**

$$T_{\#}\mu = \nu$$



Static Transport Map T from μ to ν

$$\frac{dx}{dt} = v_t(x)$$



Optimal Dynamics from μ to ν

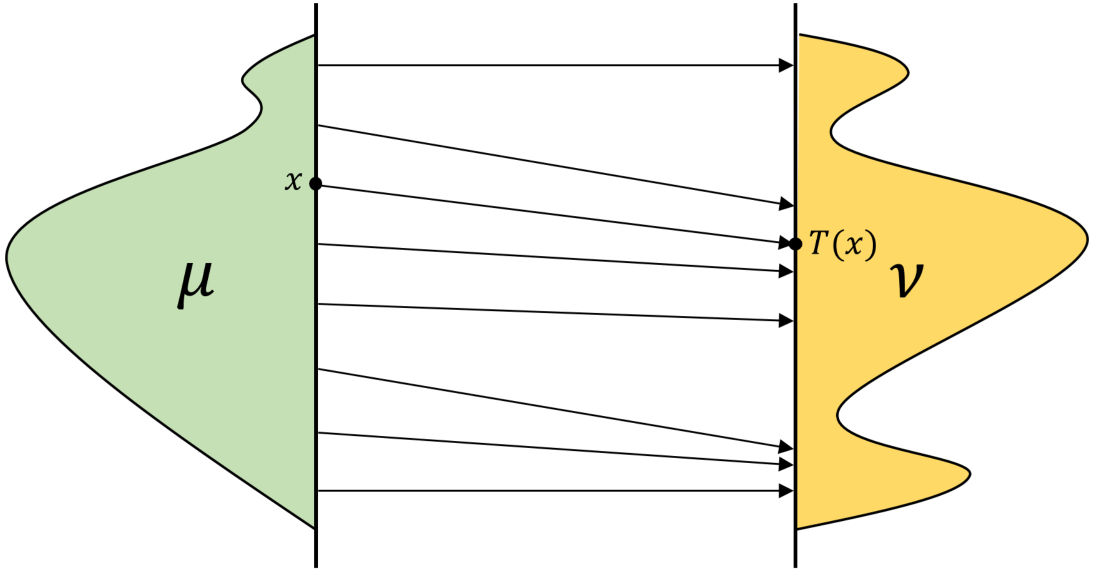
Dynamic Optimal Transport

- **The dynamic optimal transport problem tracks the continuous evolution of μ to ν.**

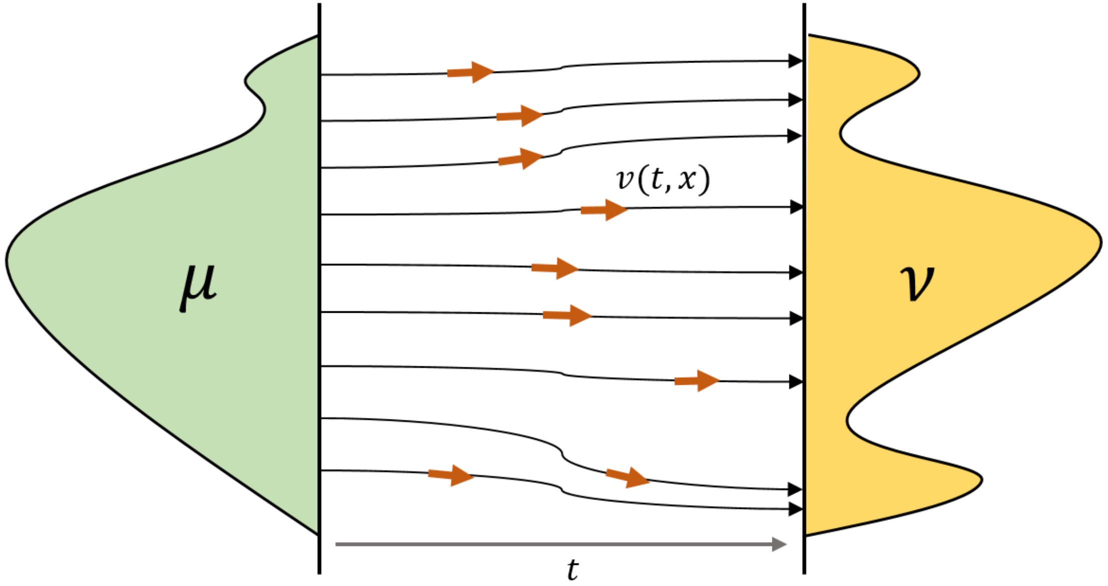
$$\inf_{v:[0,1]\times\mathcal{X}\to\mathcal{X}} \left[\int_0^1 \int_{\mathcal{X}} \alpha \|v_t(x)\|^2 d\rho_t dt; \ \frac{\partial \rho_t}{\partial t} + \nabla\cdot(v_t\rho_t) = 0, \ \rho_0 = \mu, \ \rho_1 = \nu\right].$$

Benamou-Brenier formulation when $c(x,y) = \alpha\|x-y\|^2$.

Continuity Eqn.



Static Transport Map T from μ to ν



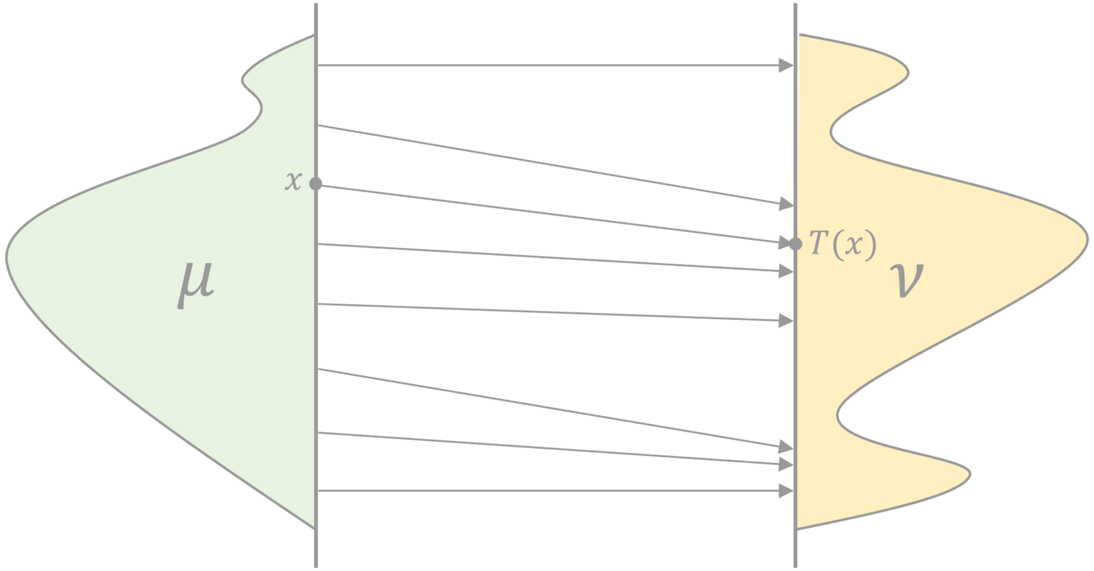
Optimal Dynamics from μ to ν

Dynamic Optimal Transport

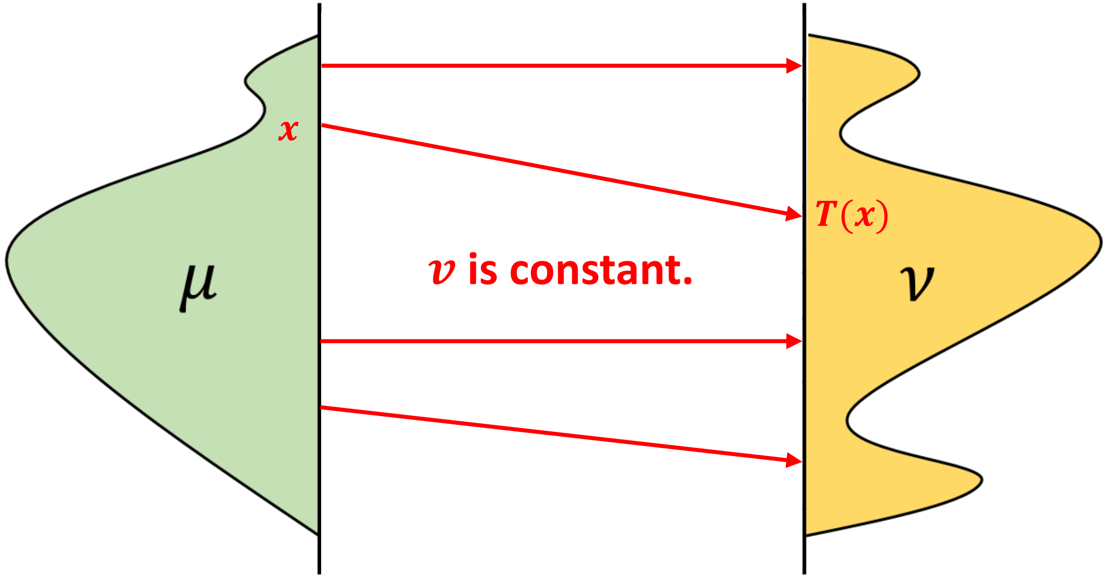
- The dynamic optimal transport problem tracks the continuous evolution of μ to ν.

$$\inf_{v:[0,1]\times\mathcal{X}\to\mathcal{X}} \left[\int_0^1 \int_{\mathcal{X}} \alpha \|v_t(x)\|^2 d\rho_t dt; \; \frac{\partial \rho_t}{\partial t} + \nabla \cdot (v_t \rho_t) = 0, \; \rho_0 = \mu, \; \rho_1 = \nu \right].$$

Benamou-Brenier formulation when $c(x, y) = \alpha\|x - y\|^2$.



Static Transport Map T from μ to ν



Optimal Dynamics from μ to ν

Displacement Interpolation

- **Displacement interpolation (DI) describes the optimal solution to the dynamic OT problem using the optimal transport map $T^\star$ in a simple form.**

Dynamic OT:

$$\inf_{v:[0,1]\times\mathcal{X}\to\mathcal{X}} \left[\int_0^1 \int_{\mathcal{X}} \alpha \|v_t(x)\|^2 d\rho_t dt;\ \frac{\partial \rho_t}{\partial t} + \nabla \cdot (v_t \rho_t) = 0,\ \rho_0 = \mu,\ \rho_1 = \nu\right].$$

Displacement Interp.:

$$\rho_t^{dis} := [(1-t)\cdot Id + t \cdot T^\star]_{\#}\mu \quad \text{and} \quad \rho_t^{dis} = \rho_t^\star \quad \text{for } 0 \le t \le 1.$$

DI = Dynamic OT

Displacement Interpolation and Barycenter

- **Displacement interpolation (DI) describes the optimal solution to the dynamic OT problem using the optimal transport map $T^\star$ in a simple form.**

Dynamic OT:

$$\inf_{v:[0,1]\times\mathcal{X}\to\mathcal{X}} \left[\int_0^1 \int_{\mathcal{X}} \alpha \|v_t(x)\|^2 d\rho_t dt; \ \frac{\partial \rho_t}{\partial t} + \nabla \cdot (v_t \rho_t) = 0, \ \rho_0 = \mu, \ \rho_1 = \nu\right].$$

Displacement Interp.:

$$\rho_t^{dis} := [(1-t)\cdot Id + t \cdot T^\star]_\# \mu \quad \text{and} \quad \rho_t^{dis} = \rho_t^\star \quad \text{for } 0 \le t \le 1.$$

DI = Dynamic OT

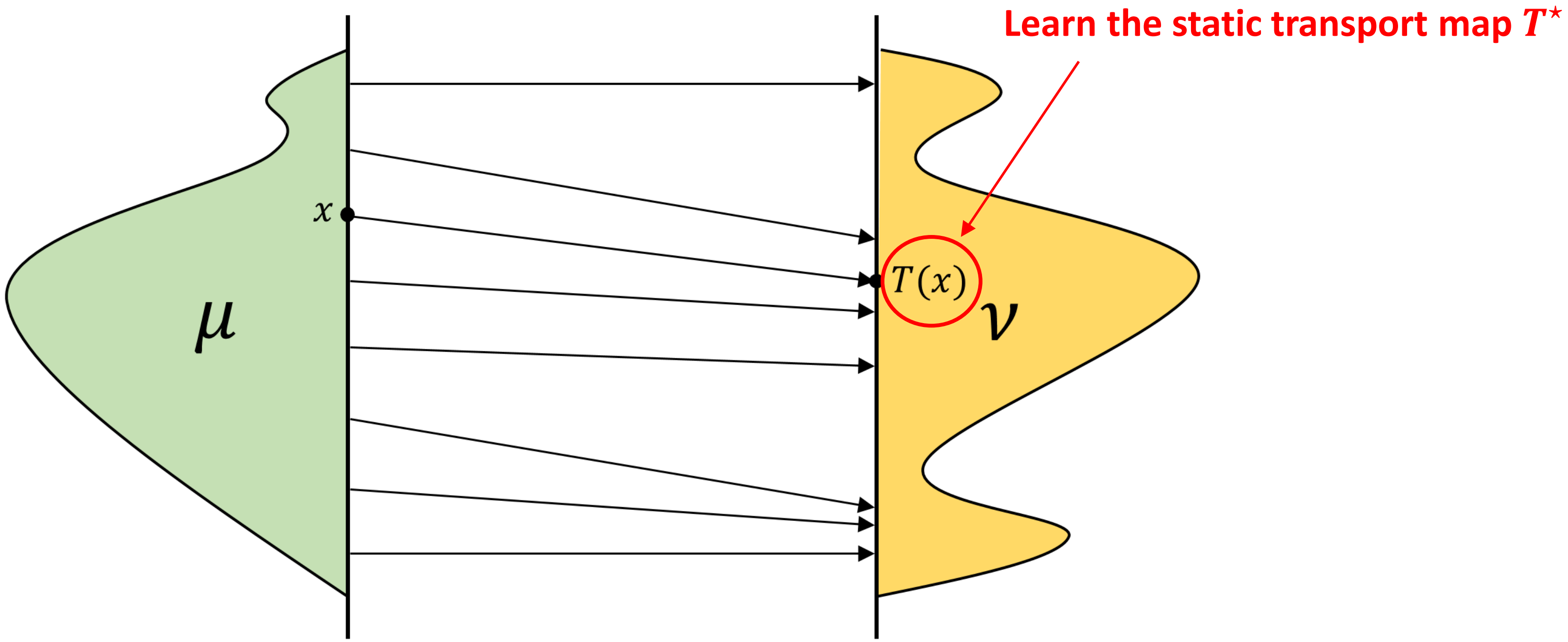
Barycenter:

$$\rho_t^\star = \arg\inf_\rho \mathcal{L}_{DI}(t,\rho) \quad \text{where} \quad \mathcal{L}_{DI}(t,\rho) = (1-t)W_2^2(\mu,\rho) + tW_2^2(\rho,\nu).$$

DI = Wasserstein Barycenter

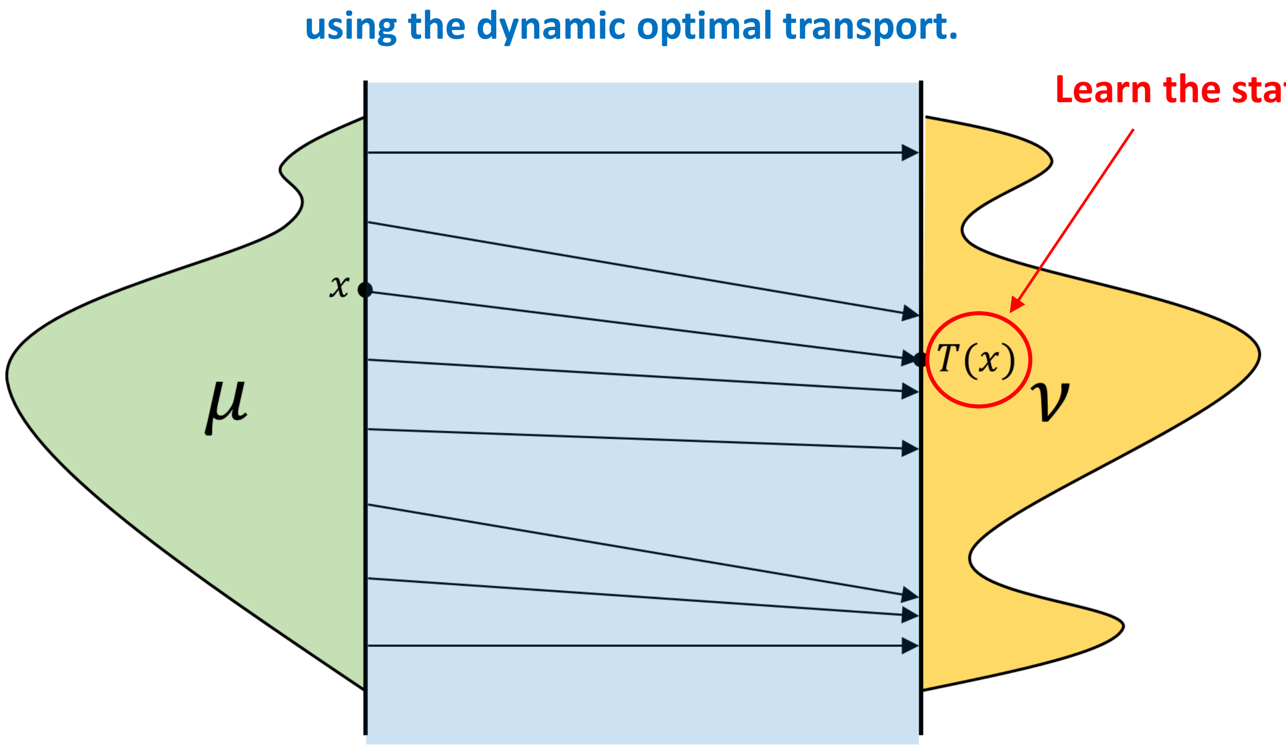
Displacement Interpolation OTM (DIOTM)

- **We proposed a neural optimal transport model T_θ, called *DIOTM*, that leverages the entire trajectory of dynamic optimal transport.**



Displacement Interpolation OTM (DIOTM)

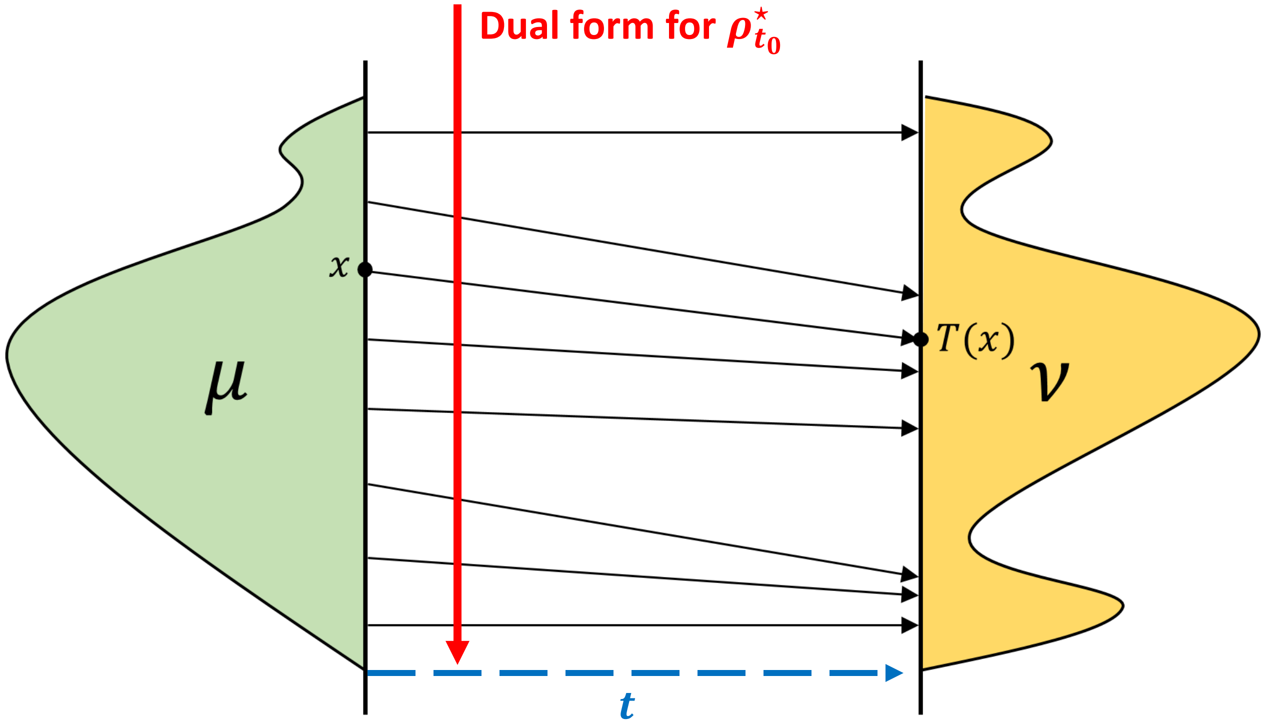
- **We proposed a neural optimal transport model T_θ, called *DIOTM*, that leverages the entire trajectory of dynamic optimal transport.**



DIOTM

Dual form of Displacement Interpolation

- **Begin with the displacement interpolation $\rho_t^\star$ at specific time t.**



DIOTM

Dual form of Displacement Interpolation

- From the Wasserstein barycenter characterization, we derive the following dual form:

Primal : $\rho_t^\star = \arg\inf_\rho \mathcal{L}_{DI}(t,\rho) \quad \text{where} \quad \mathcal{L}_{DI}(t,\rho) = (1-t)W_2^2(\mu,\rho) + tW_2^2(\rho,\nu).$

Dual form of Displacement Interpolation

- From the Wasserstein barycenter characterization, we derive the following dual form:

Primal : $\rho_t^\star = \arg\inf_\rho \mathcal{L}_{DI}(t,\rho) \quad \text{where} \quad \mathcal{L}_{DI}(t,\rho) = (1-t)W_2^2(\mu,\rho) + tW_2^2(\rho,\nu).$

Theorem 3.1. *Given the assumptions in Appendix A, for a given $t \in (0,1)$, the minimization problem $\inf_\rho \mathcal{L}_{DI}(t,\rho)$ (Eq. 8) is equivalent to the following dual problem:*

Dual :

$$\sup_{f_{1,t}, f_{2,t} \text{ with } (1-t)f_{1,t}+tf_{2,t}=0} \left[(1-t)\int_{\mathcal{X}} f_{1,t}^c(x)d\mu(x) + t\int_{\mathcal{Y}} f_{2,t}^c(y)d\nu(y) \right]. \tag{9}$$

where the supremum is taken over two potential functions $f_{1,t}: \mathcal{Y} \to \mathbb{R}$ and $f_{2,t}: \mathcal{X} \to \mathbb{R}$, which satisfy $(1-t)f_{1,t} + tf_{2,t} = 0$.

- Note that v^c denotes the ***c*-transform** of v, i.e., $v^c(x) = \inf_{y \in \mathcal{Y}} (c(x,y) - v(y))$.

DIOTM

Dual form of Displacement Interpolation

- We can combine these two potentials into a single potential V_t.

Dual :

$$\sup_{f_{1,t}, f_{2,t} \text{ with } (1-t)f_{1,t}+tf_{2,t}=0} \left[(1-t) \int_{\mathcal{X}} f_{1,t}^c(x) d\mu(x) + t \int_{\mathcal{Y}} f_{2,t}^c(y) d\nu(y) \right]. \tag{9}$$

Corollary 3.2. *For a given* $t \in (0,1)$, *let* $f_{1,t}(y) = tV_t(y)$ *and* $f_{2,t}(x) = -(1-t)V_t(x)$ *for some value function* $V_t : \mathcal{X} = \mathcal{Y} \to \mathbb{R}$. *Then, the dual formulation of displacement interpolation (Eq. 9) can be rewritten as follows:*

$$\sup_{V_t} \left[\int_{\mathcal{X}} V_t^{c_{0,t}}(x) d\mu(x) + \int_{\mathcal{Y}} (-V_t)^{c_{t,1}}(y) d\nu(y) \right], \tag{14}$$

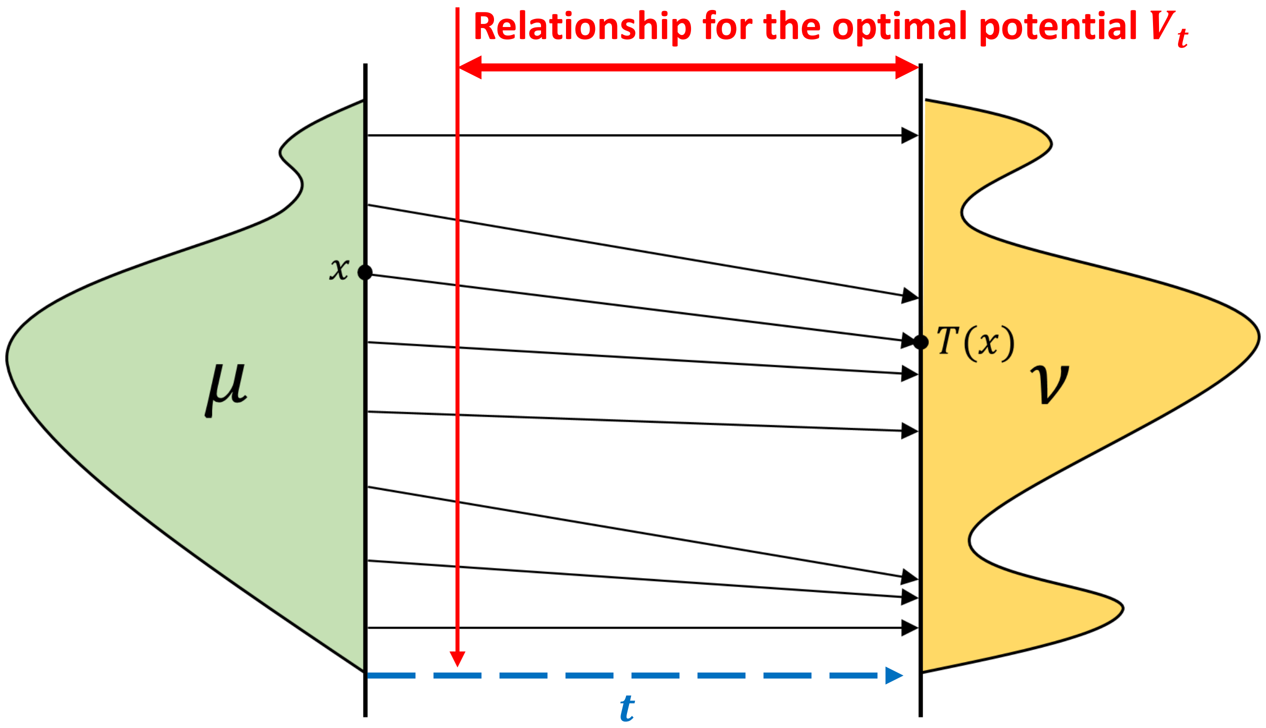
where $c_{s,t}(x,y) = \frac{\alpha\|x-y\|^2}{t-s}$ *for every* $0 \le s < t \le 1$.

Investigate the relationships between V_t across time t.

DIOTM

Optimality Condition for Time-dependent Potential

- We establish the **relationship between the potential functions for each DI** and utilize it to learn the optimal transport map.



Optimality Condition for Time-dependent Potential

- The optimal potential $V_t^\star$ for each $\rho_t^\star$ satisfies the following [1]:

$$-V_s(x) = \inf_y \left(c_{s,t}(x,y) - V_t(y)\right)$$

where $c_{s,t}(x,y) = \frac{\alpha\|x-y\|^2}{t-s}$ for $0 \leq s < t \leq 1$.

- From this, we can derive the following **optimality condition**:

Theorem 3.3. *Given the assumptions in Appendix A, the optimal $V_t^\star$ in Eq. 14 satisfies the following:*

$$V_t^\star = \arg\sup_{V_t} \left[\int_{\mathcal{X}} V_t^{c_{0,t}}(x) d\mu(x) + \int_{\mathcal{Y}} V_t(x) d\rho_t^\star(x) \right], \tag{15}$$

up to constant $\rho^\star$-a.s.. Moreover, there exists $\{V_t^\star\}_{0\leq t\leq 1}$ that satisfies Hamilton-Jacobi-Bellman (HJB) equation, i.e.

Use this term as regularizer

$$\partial_t V_t^\star + \frac{1}{4\alpha}\|\nabla V_t^\star\|^2 = 0, \qquad \rho^\star\text{-a.s.} \tag{16}$$

[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.

DIOTM Model

- Introduce **the c-transform parametrization [1].**

Dual : $$\sup_{V_t} \left[\int_{\mathcal{X}} V_t^{c_{0,t}}(x) d\mu(x) + \int_{\mathcal{Y}} (-V_t)^{c_{t,1}}(y) d\nu(y) \right]$$

- Parametrize $V_t^{c_{0,t}}$ and $V_t^{c_{t,1}}$ as follows:

$$\boxed{\begin{aligned} \overrightarrow{T}_t^\star(x) &\in \mathrm{argmin}_{y\in\mathcal{Y}} \left[c(x,y) - f_{1,t}^\star(y) \right] \\ \overleftarrow{T}_t^\star(y) &\in \mathrm{argmin}_{x\in\mathcal{X}} \left[c(x,y) - f_{2,t}^\star(x) \right] \end{aligned}} \begin{aligned} &\Leftrightarrow \quad V_t^{c_{0,t}}(x) = \frac{1}{t} c\left(x, \overrightarrow{T}_t(x)\right) - V_\phi(t, \overrightarrow{T}_t(x)) \\ &\Leftrightarrow \quad (-V_t)^{c_{t,1}}(y) = \frac{1}{1-t} c\left(\overleftarrow{T}_t(y), y\right) + V_\phi(t, \overleftarrow{T}_t(y)) \end{aligned}$$

$\overrightarrow{T}_t^\star$ **and** $\overleftarrow{T}_t^\star$ **satisfy these conditions.**

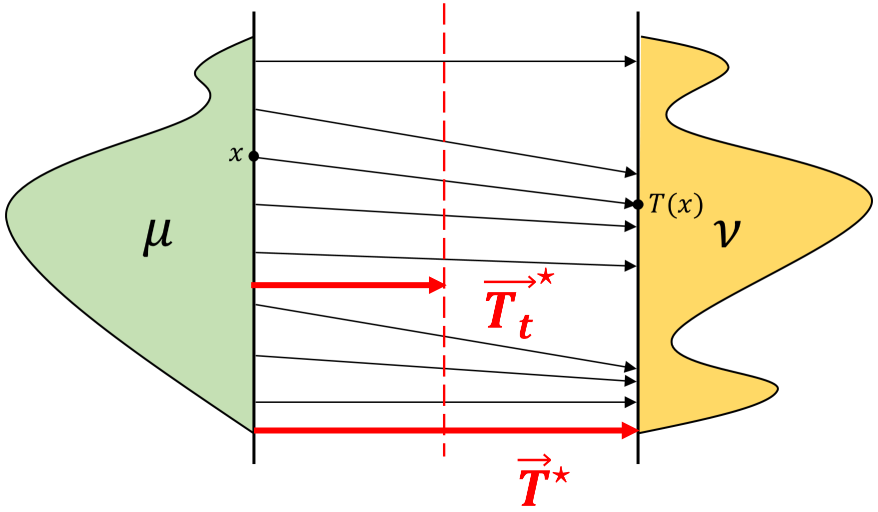
[1] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps." *ICLR*, 2022.

DIOTM Model

- **Using DI, Parametrize interpolant generators through the boundary generators:**

$$\overrightarrow{T}_t(x) = (1-t)x + t\overrightarrow{T}_\theta(x), \quad \overleftarrow{T}_t(y) = (1-t)y + t\overleftarrow{T}_\theta(y) \quad \text{for } t \in (0,1).$$

  - Then, we have the following learning objective:



$$\mathcal{L}^{DI}_{\phi,\theta} = \sup_{V_\phi} \int_{\mathcal{X}} \inf_{\overrightarrow{T}_\theta} \mathbb{E}_t \left[\alpha t \|x - \overrightarrow{T}_\theta(x)\|^2 - V_\phi(t, \overrightarrow{T}_t(x))\right] d\mu(x)$$
$$+ \int_{\mathcal{Y}} \inf_{\overleftarrow{T}_\theta} \mathbb{E}_t \left[\alpha(1-t)\|\overleftarrow{T}_\theta(y) - y\|^2 + V_\phi(t, \overleftarrow{T}_t(y))\right] d\nu(y).$$

DIOTM Model

- **Parametrize interpolant generators through the boundary generators using DI.**

$$\overrightarrow{T}_t(x) = (1-t)x + t\overrightarrow{T}_\theta(x), \quad \overleftarrow{T}_t(y) = (1-t)y + t\overleftarrow{T}_\theta(y) \quad \text{for } t \in (0,1).$$

$$\mathcal{L}_{\phi,\theta}^{DI} = \sup_{V_\phi} \int_{\mathcal{X}} \inf_{\overrightarrow{T}_\theta} \mathbb{E}_t \left[\alpha t \|x - \overrightarrow{T}_\theta(x)\|^2 - V_\phi(t, \overrightarrow{T}_t(x)) \right] d\mu(x) + \int_{\mathcal{Y}} \inf_{\overleftarrow{T}_\theta} \mathbb{E}_t \left[\alpha(1-t)\|\overleftarrow{T}_\theta(y) - y\|^2 + V_\phi(t, \overleftarrow{T}_t(y)) \right] d\nu(y).$$

- **HJB regularizer**

$$\mathcal{R}(V_\phi) = \mathbb{E}_{t, x\sim\rho_t} \left| 2\alpha\, \partial_t V_\phi(t,x) + \frac{1}{2}\|\nabla V_\phi(t,x)\|^2 \right|.$$

- **The total learning objective is as follows:**

$$\mathcal{L}_{\phi,\theta} = \mathcal{L}_{\phi,\theta}^{DI} + \lambda \mathcal{R}(V_\phi(t,x)).$$

Experiments

Experiments

Optimal Transport Map Evaluation

- **Our model learns a more accurate optimal transport map compared to previous methods.**

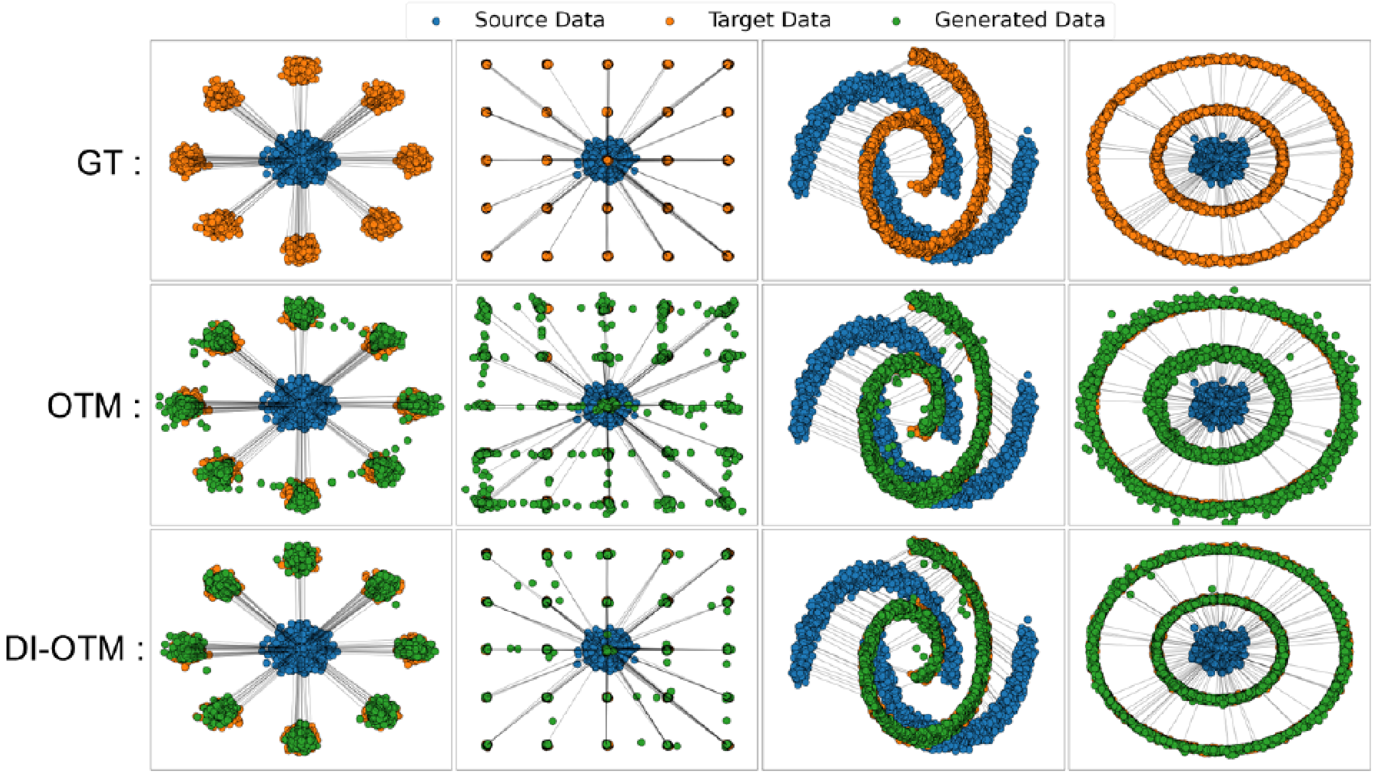


Figure 1: **Visualization of transport maps T on synthetic datasets.** The transport map is visualized as a black line connecting each source sample x to its corresponding generated data $T(x)$.

Experiments

Optimal Transport Map Evaluation

- **Our model learns a more accurate optimal transport map compared to previous methods.**
  - $\boldsymbol{W_2 = W_2\left(\overrightarrow{T_\theta}_{\#}\,\mu, \nu\right)}$: Error between the generated and target distributions.
  - $\boldsymbol{L_2 = \int_{\mathcal{X}} \left\|\overrightarrow{T_\theta}(x) - T^\star(x)\right\|_2^2 d\mu_{test}}$: Error between the transport maps.

Table 1: **Evaluation of optimal transport map** on the synthetic datasets between OTM (Rout et al., 2022) and ours, based on the 2-Wasserstein distance W_2 and the L2 distance between transport maps.

Metric	G→8G		G→25G		Moon→Spiral		G→Circles	
	OTM	DIOTM	OTM	DIOTM	OTM	DIOTM	OTM	DIOTM
W_2 (↓)	4.93	**3.72**	10.09	**6.49**	**0.40**	0.55	3.96	**2.34**
L2 (↓)	6.48	**4.38**	13.09	**10.00**	1.77	**1.67**	6.23	**5.44**

Scalability Comparison on I2I Translation Tasks

- **We assessed our model on several Image-to-Image (I2I) translation benchmarks.**
  - Optimal transport map serves as a generator for the target distribution, which maps each input x to its cost-minimizing counterpart y.

Source Images (Male)

Translated Images (Female)

Generates the target distribution while minimizing $c\big(x, T(x)\big)$.

Scalability Comparison on I2I Translation Tasks

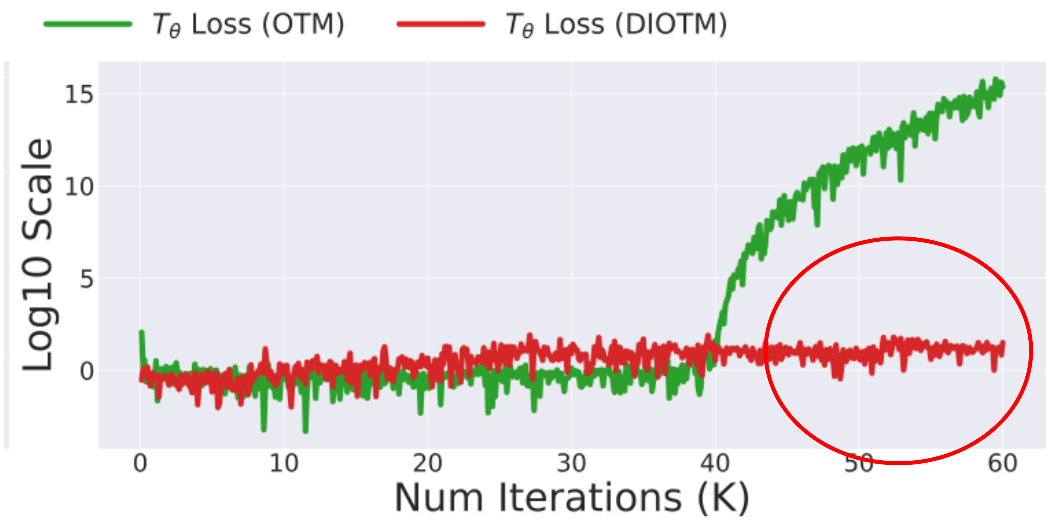
- **We assessed our model on several Image-to-Image (I2I) translation benchmarks.**
  - Optimal transport map serves as a generator for the target distribution, which maps each input x to its cost-minimizing counterpart y.

Data	Model	FID (↓)
Male→Female (64x64)	CycleGAN Zhu et al. (2017)	12.94
	NOT Korotin et al. (2023)	11.96
	OTM† Fan et al. (2022)	6.42
	DIOTM† (Ours)	**5.27**
Wild→Cat (64x64)	DSBM Shi et al. (2024a)	20+
	OTM† Fan et al. (2022)	12.42
	DIOTM† (Ours)	**10.72**
Male→Female (128x128)	DSBM Shi et al. (2024a)	37.8
	ASBM Gushchin et al. (2024)	16.08
	OTM† Fan et al. (2022)	7.55
	DIOTM† (Ours)	**7.40**

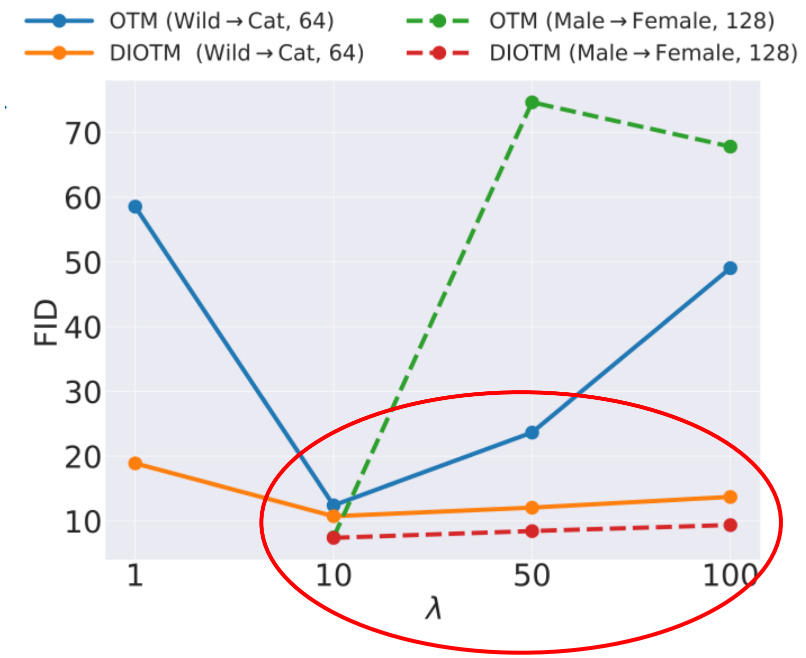
Image-to-Image translation benchmarks compared to existing Neural (Entropic) OT Models

Stability Comparison to Previous Approaches

- **Our model is more robust to hyperparameters and exhibits stable training dynamics.**



Stable training dynamics of DIOTM



Robustness to hyperparameters of DIOTM

Conclusion

- We introduced **DIOTM**, a static neural optimal transport model based on displacement interpolation.
- We derived the **dual formulation of displacement interpolation** and investigated the optimal relationships between potential functions.
- We proposed the **HJB regularizer**, which is derived from the optimality condition of the potential function.
- Our model improves the training stability and accuracy of existing OT Map models that leverage min-max objectives.

Thank you!

Appendix

Training Algorithm

Algorithm 1 Training algorithm of DIOTM

Require: The source distribution μ and the target distribution ν. Transport networks $\overrightarrow{T}_\theta$, $\overleftarrow{T}_\theta$ and the discriminator network V_ϕ. Total iteration number K, and regularization hyperparameter λ.

1: **for** $k = 0, 1, 2, \ldots, K$ **do**

2: Sample a batch $x \sim \mu$, $y \sim \nu$, $t \sim U[0, 1]$.

3: $x_t \leftarrow (1-t)x + t\overrightarrow{T}_\theta(x)$, $y_t \leftarrow (1-t)\overleftarrow{T}_\theta(y) + ty$.

4: Update V_ϕ to increase the $\mathcal{L}_\phi$

$$\mathcal{L}_\phi = -V_\phi(t, x_t) + V_\phi(t, y_t) - \lambda\mathcal{R}(V_\phi(t, x_t)) - \lambda\mathcal{R}(V_\phi(t, y_t)).$$

5: Update $\overrightarrow{T}_\theta$ to decrease the loss: $c_{0,t}(x, x_t) - V_\phi(t, x_t)$.

6: Update $\overleftarrow{T}_\theta$ to decrease the loss: $c_{t,1}(y_t, y) + V_\phi(t, y_t)$.

7: **end for**

Appendix

Stability Comparison to Previous Approaches

- **Our HJB regularizer outperforms other regularizers.**
  - Our HJB regularizer is the only regularizer that incorporates the time derivative.
    - $\mathcal{R}_{\mathrm{OTM}}(t, x_t, y_t) = \|\nabla_y \left(c_{0,t}(x, x_t) - V_\phi(t, x_t)\right)\| + \|\nabla_y \left(c_{t,1}(y, y_t) + V_\phi(t, y_t)\right)\|.$
    - $\mathcal{R}_{\mathrm{R1}}(t, x_t, y_t) = \|\nabla_y V_\phi(t, x_t)\|^2 + \|\nabla_y V_\phi(t, y_t)\|^2.$
    - $\mathcal{R}(t, x_t, y_t)_{\mathrm{HJB}} = \left|2\alpha\, \partial_t V_\phi(t, x_t) + \frac{1}{2}\|\nabla V_\phi(t, x_t)\|^2\right| + \left|2\alpha\, \partial_t V_\phi(t, y_t) + \frac{1}{2}\|\nabla V_\phi(t, y_t)\|^2\right|.$

Table 3: **Comparison of our HJB regularizer with the OTM and R1 regularizers** on the DIOTM model. Our HJB regularizer exhibits superior performance and stability to λ.

	Model	G→8G				G→25G				Moon→Spiral			
	λ	0.1	0.2	1.0	10	0.1	0.2	1.0	10	0.1	0.2	1.0	10
W_2 (↓)	OTM	22.08	22.90	DIV	31.35	68.01	89.62	DIV	81.02	19.99	14.19	15.66	33.80
	R1	3.59	5.01	3.29	4.42	9.20	**9.94**	11.78	DIV	1.91	2.08	1.05	2.74
	HJB	**1.93**	**2.69**	**2.92**	**3.21**	**7.19**	14.64	**7.99**	**12.38**	**0.54**	**0.59**	**0.30**	**1.31**
L2 (↓)	OTM	27.41	28.21	DIV	34.47	96.89	97.98	DIV	87.05	20.96	15.01	34.31	33.80
	R1	4.49	5.39	3.87	5.14	86.05	17.64	19.52	DIV	2.88	3.56	2.36	3.74
	HJB	**3.05**	**3.44**	**3.36**	**3.98**	**16.51**	**15.82**	**11.11**	**15.64**	**1.42**	**2.25**	**1.13**	**2.27**