Overcoming Fake Solutions in Neural Optimal Transport: A Smoothing Approach for Learning the Optimal Transport Plan

2025 KIAS CAINS Workshop

JaeWoong Choi

(with Jaemoo Choi and Dohyun Kwon)

SKKU Statistics

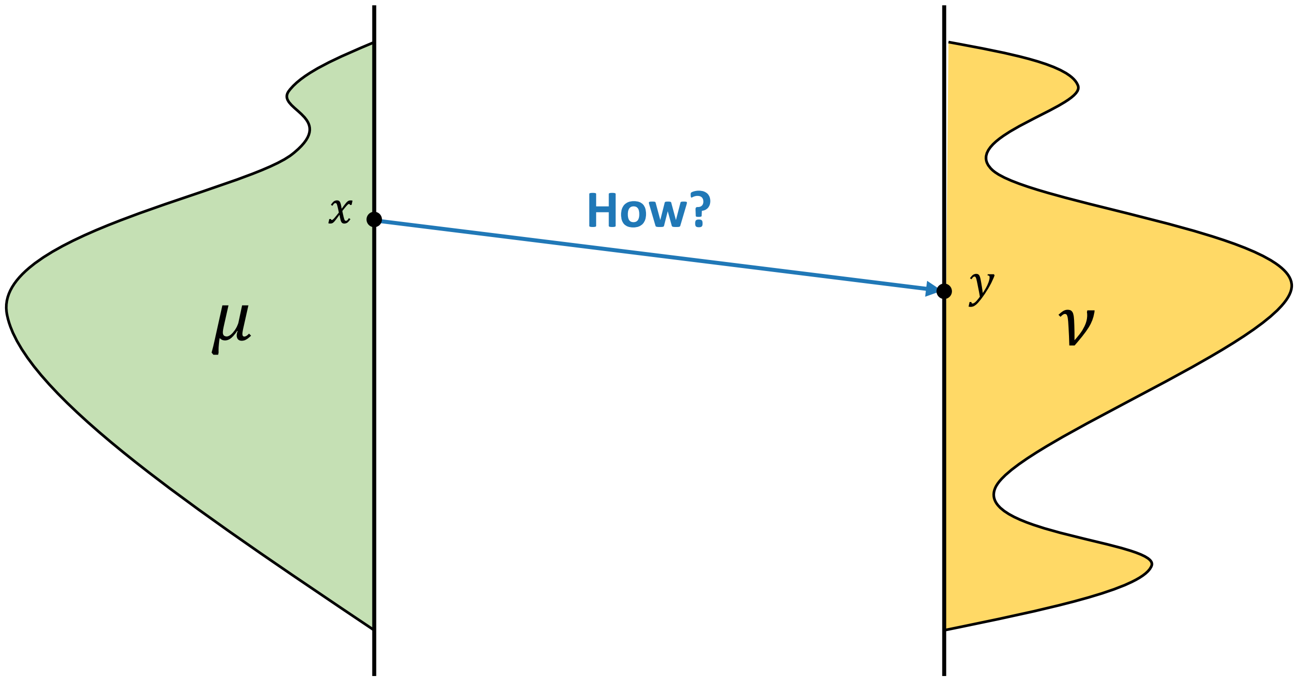
CONTENTS

Introduction

Introduction

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

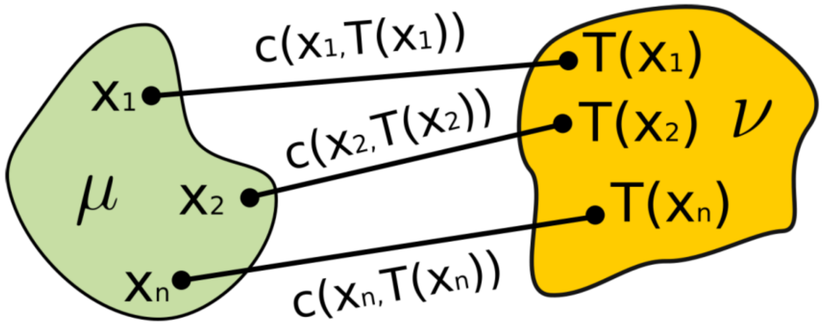
- Optimal Transport refers to the most **cost-minimizing way to transport source distribution μ to target ν**.

- **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

$$x \sim \mu \implies T(x) \sim \nu$$



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.

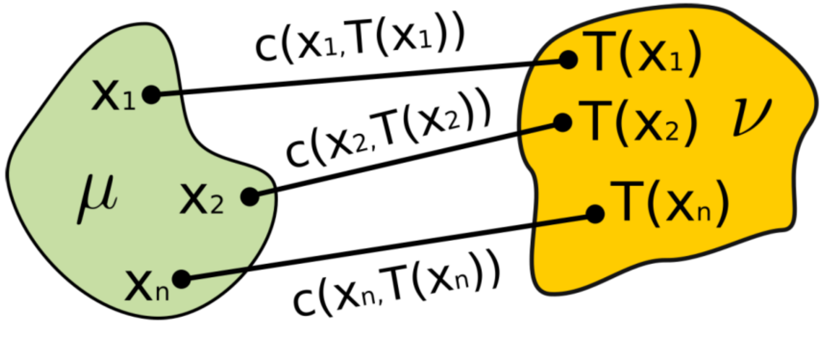
Introduction

Optimal Transport (OT) Problem

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

- **Monge's Formulation**.

$$C(\mu,\nu) := \underbrace{\inf_{T_{\#}\mu=\nu}}_{\substack{\text{Transport Map } T \\ x\sim\mu \implies T(x)\sim\nu}} \left[\int_{\mathcal{X}} \underbrace{c(x,T(x))}_{\text{Optimality by cost-minimizing}} d\mu(x)\right].$$



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.

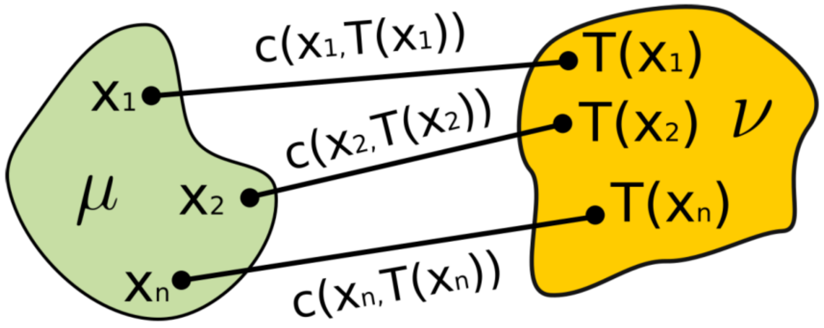
Introduction

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.**

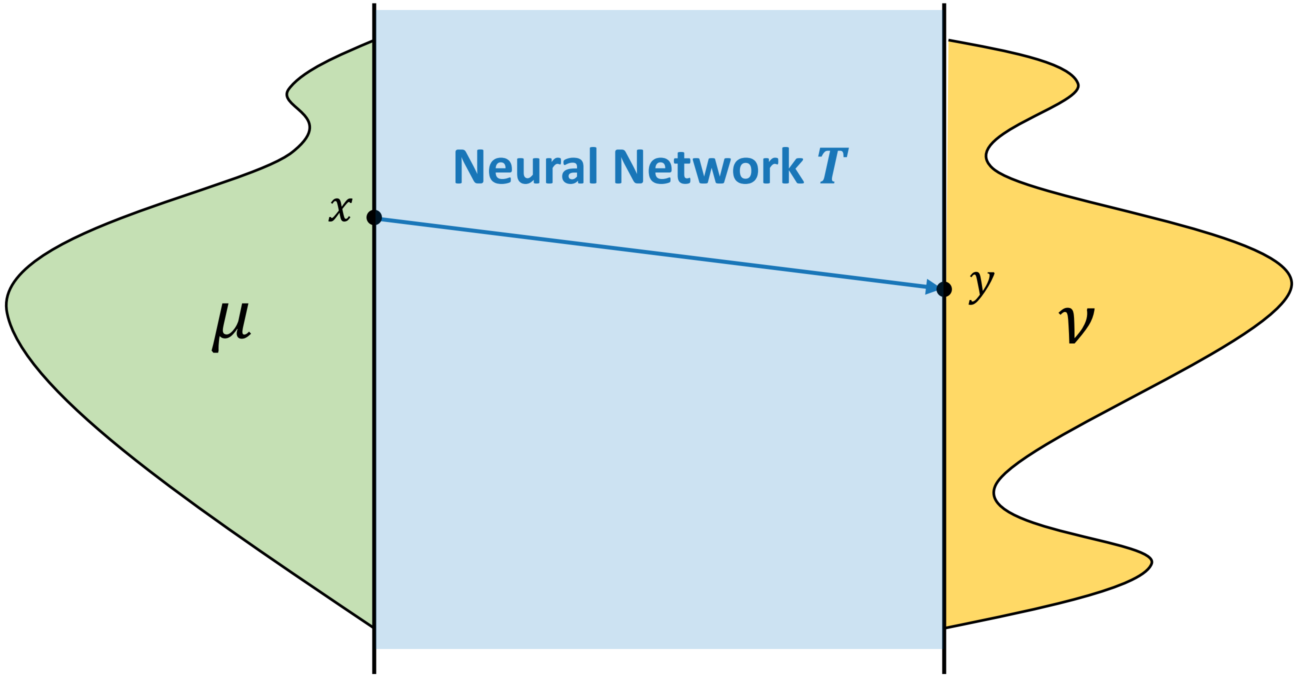
$$C(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left[\int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi(x, y) \right].$$

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 optimal transport maps with neural networks.**

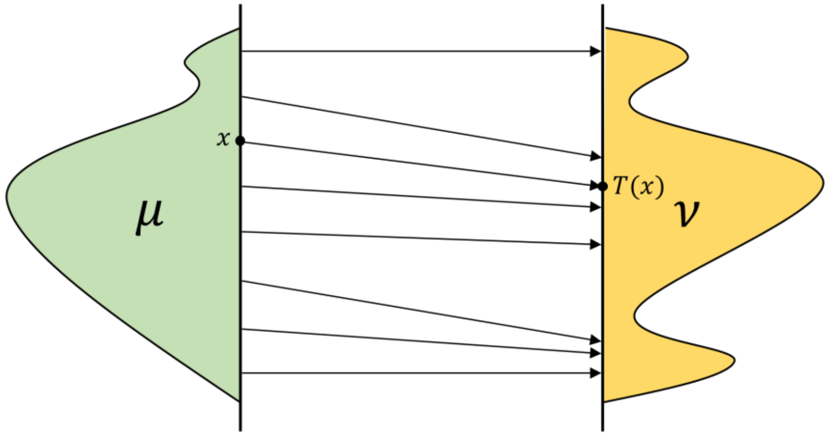


[1] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2009.
[2] Rout, Litu, Alexander Korotin, and Evgeny Burnaev. "Generative modeling with optimal transport maps.“ *ICLR*, 2022.

Fake Solutions in Neural OT

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

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

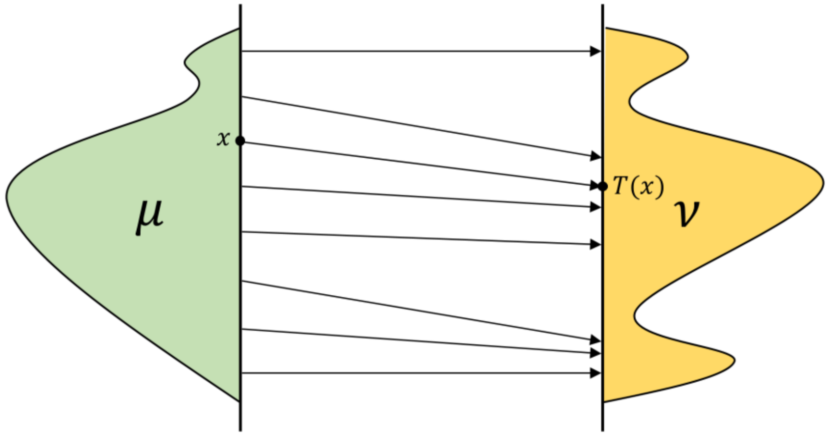
Semi-dual formulation for OT

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

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

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

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

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

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.

Fake Solutions in Semi-dual Neural OT (SNOT)

- The semi-dual approach is known to have the **fake solutions**.

$$\sup_{V} \inf_{T:\mathcal{X}\to\mathcal{Y}} \mathcal{L}(V,T) \quad \text{where} \quad \mathcal{L}(V,T) := \int_{\mathcal{X}} c(x,T(x)) - V(T(x))d\mu(x) + \int_{\mathcal{Y}} V(y)d\nu(y).$$

- When the optimal potential $V^\star$ and the transport map $T^\star$ exists, it is well known that [1, 2]

$$T^\star \in \arg\min_{T} \mathcal{L}(V^\star, T)$$

- Thus, the pair $(\boldsymbol{V}^\star, \boldsymbol{T}^\star)$ **is the solution to the max-min problem**.

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.

Fake Solutions in Semi-dual Neural OT (SNOT)

- The semi-dual approach is known to have the **fake solutions**.

$$\sup_{V} \inf_{T:\mathcal{X}\to\mathcal{Y}} \mathcal{L}(V,T) \quad \text{where} \quad \mathcal{L}(V,T) := \int_{\mathcal{X}} c(x,T(x)) - V(T(x))d\mu(x) + \int_{\mathcal{Y}} V(y)d\nu(y).$$

  - When the optimal potential $V^{\star}$ and the transport map $T^{\star}$ exists, it is well known that [1, 2]

$$T^{\star} \in \arg\min_{T} \mathcal{L}(V^{\star}, T)$$

  - Thus, the pair $(\boldsymbol{V^{\star}}, \boldsymbol{T^{\star}})$ **is the solution to the max-min problem**.

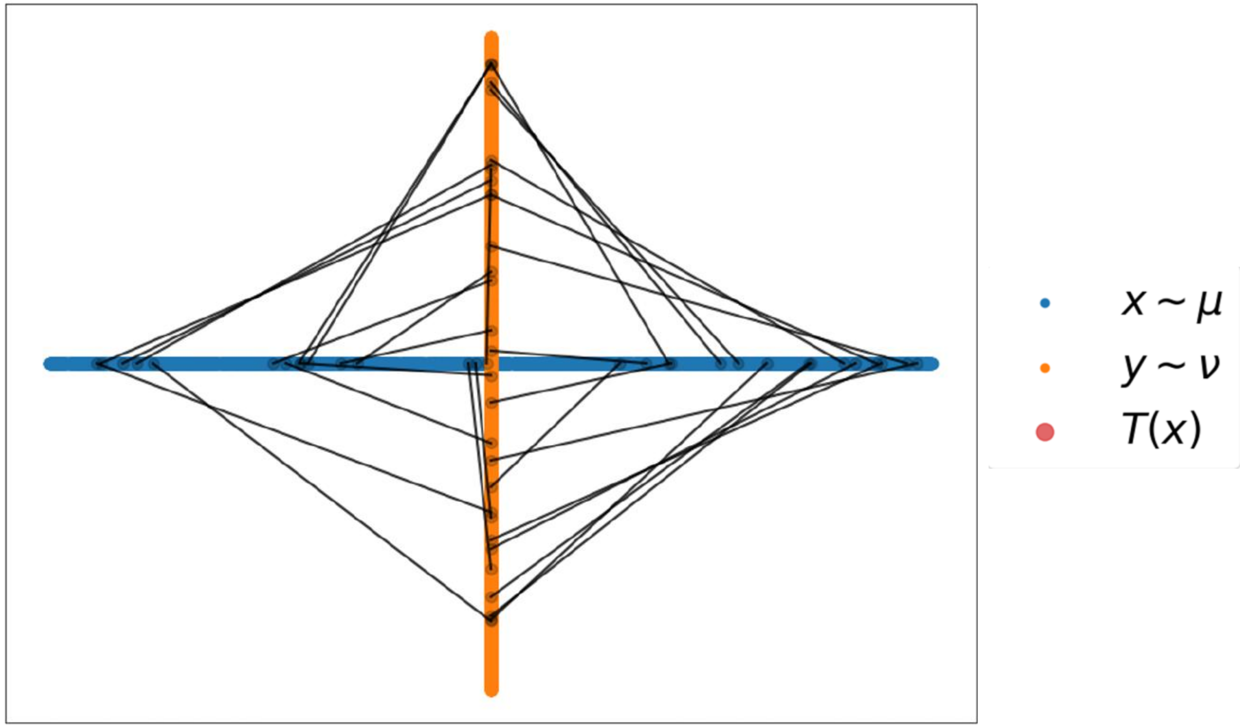
- However, **not all solutions** of the max-min problem correspond to the **true optimal potential and transport map pair**.
- The optimal solution of SNOT framework may fail to recover the correct optimal transport map.

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.

Fake Solution Example

- Assume that the **source μ** and **target ν** distributions are **uniformly supported** on

$$A = [-1, 1] \times \{0\} \quad \text{and} \quad B = \{0\} \times [-1, 1].$$



Fake Solution Example

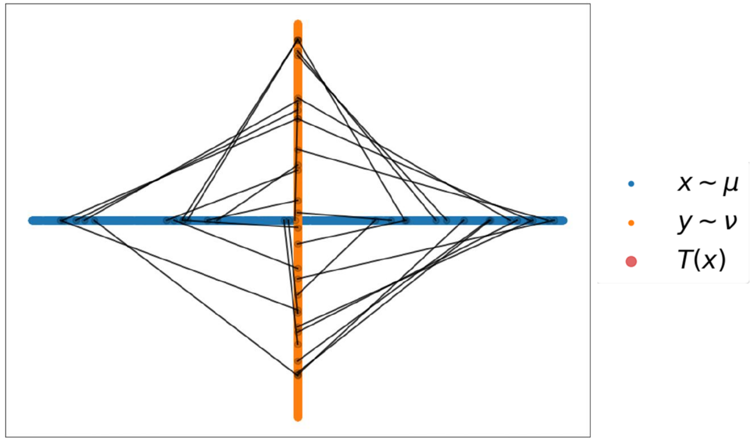
- Assume that the **source $\boldsymbol{\mu}$** and **target $\boldsymbol{\nu}$** distributions are **uniformly supported** on

$$A = [-1, 1] \times \{0\} \quad \text{and} \quad B = \{0\} \times [-1, 1]$$

- All transport map achieves the **same transport cost**:

$$\int_{\mathcal{X}} c(x, T(x)) d\mu(x) = \frac{1}{2} \int_{-1}^{1} x_1^2 d\mu(x) - \int_{\mathcal{X}} \cancel{\langle x, T(x) \rangle} \, d\mu(x) + \frac{1}{2} \int_{-1}^{1} y_1^2 d\nu(y) = \frac{2}{3}.$$

- Since all transport maps yield the same cost, **any transport map becomes an OT Map $\boldsymbol{T^\star}$**.



Fake Solution Example

- To analyze the fake solution, we first show that $V^\star$ is the optimal Kantorovich potential

$$V^\star(y) = \frac{1}{2}\|y\|^2.$$

Dual Prob $$S(\mu, \nu) := \sup_{V \in S_c} \left[\int_{\mathcal{X}} V^c(x) d\mu(x) + \int_{\mathcal{Y}} V(y) d\nu(y) \right]$$

Fake Solutions in Semi-dual Neural OT

Fake Solution Example

- To analyze the fake solution, we first show that $V^\star$ is the optimal Kantorovich potential

$$V^\star(y) = \frac{1}{2}\|y\|^2.$$

- This result follows from solving the **inner problem** of SNOT (dual OT problem) as follows:

$$\sup_V \inf_{T:\mathcal{X}\to\mathcal{Y}} \mathcal{L}(V,T) \quad \text{where} \quad \mathcal{L}(V,T) := \int_{\mathcal{X}} c(x,T(x)) - V(T(x))d\mu(x) + \int_{\mathcal{Y}} V(y)d\nu(y).$$

$$\inf_T \int_{\mathcal{X}} \frac{1}{2}\|x\|^2 - \cancel{\langle x, T(x)\rangle} d\mu(x) + \int_{\mathcal{Y}} \frac{1}{2}\|y\|^2 d\nu(y) = \frac{2}{3}.$$

where the cost function is the quadratic cost $c(x,y) = \frac{1}{2}\|x-y\|^2$.

- Since $V^\star$ achieves the optimal transport cost, $V^\star$ is the **Kantorovich potential.**

Fake Solution Example

- The max-min solution of the SNOT problem is defined as the **minimizer of inner problem**.
- This implies that **any measurable map $T_\theta : A \to B$ can be a max-min solution of SNOT**.

$$T_\theta : x \mapsto \arg\min_{y \in \mathcal{Y}} \left[c(x,y) - V_\phi\left(y\right) \right]$$

$$\inf_T \int_{\mathcal{X}} \frac{1}{2}\|x\|^2 - \cancel{\langle x, T(x) \rangle} d\mu(x) + \int_{\mathcal{Y}} \frac{1}{2}\|y\|^2 d\nu(y) = \frac{2}{3}.$$

No dependency on T

Fake Solution Example

- The max-min solution of the SNOT problem is defined as the **minimizer of inner problem**.
- This implies that **any measurable map $T_\theta : A \to B$ can be a max-min solution of SNOT**.

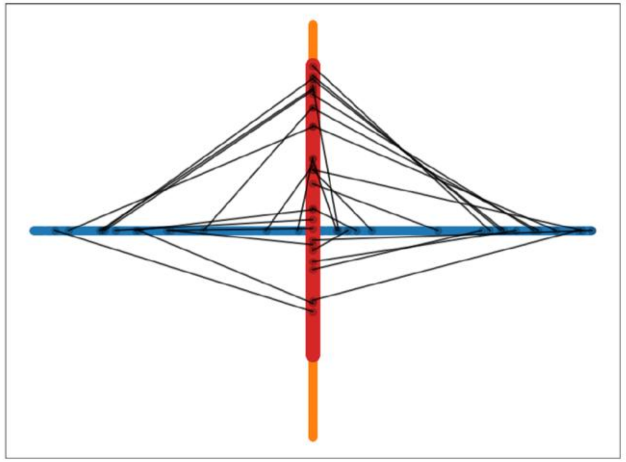
$$T_\theta : x \mapsto \arg\min_{y \in \mathcal{Y}} \left[c(x,y) - V_\phi(y) \right]$$

$$\inf_T \int_{\mathcal{X}} \frac{1}{2}\|x\|^2 - \cancel{\langle x, T(x) \rangle} d\mu(x) + \int_{\mathcal{Y}} \frac{1}{2}\|y\|^2 d\nu(y) = \frac{2}{3}.$$

No dependency on T

- There is no **constraint ensuring that $T_\theta \mu = \nu$.**

Fake Solution Visualization



Analytical results for Semi-dual Neural OT

Sufficient Conditions for Ensuring True Solution

- The max-min solution $(V^\dagger, T^\dagger)$ and the optimal solution $(V^\star, T^\star)$ satisfy:

$$\{(V^\star, T^\star)\} \subsetneq \{(V^\dagger, T^\dagger)\}.$$

- **The sufficient condition to prevent fake solution** is as follows:

Theorem 3.1. *Let $\mu \in \mathcal{P}_2(\mathcal{X})$, $\nu \in \mathcal{P}_2(\mathcal{Y})$, and $c(x, y) = \frac{1}{2}\|x - y\|^2$. Assume that μ does not give mass to the measurable sets of Hausdorff dimension at most d − 1 dimension.*

(1) Then, there exists a unique OT Map $T^\star$ in (Eq. 3) and the Kantorovich potential $V^\star \in S_c$ in (Eq. 5).

(2) For the Kantorovich potential $V^\star \in S_c$, the minimization problem,

$$\mathcal{D}_x := \arg\min_{y \in \mathcal{Y}} \left[c(x, y) - V^\star(y) \right], \tag{10}$$

is uniquely determined μ-a.s., i.e. $\mathcal{D}_x = \{y_x\}$ for μ-a.s $x \in \mathcal{X}$. In particular, a map $x \mapsto y_x \in \mathcal{D}_x$ is a unique OT Map $T^\star$ in law.

Condition on μ

Ex) Absolutely continuous

T_θ-parametrization

Analytical results for Semi-dual Neural OT

Sufficient Conditions for Unique max-min Solution

- The max-min solution $(V^\dagger, T^\dagger)$ and the optimal solution (V^*, T^*) satisfy:

$$\{(V^*, T^*)\} \subset \{(V^\dagger, T^\dagger)\}$$

Unique

- For completeness, an additional condition for ν is required to ensure **the uniqueness of the mix-min solution.**

Corollary 3.3. *Suppose $\mathcal{Y} \subset \mathbb{R}^d$ is a closure of a bounded open set. Suppose $\mu \in \mathcal{P}_2(\mathcal{X})$ and $\nu \in \mathcal{P}_2(\mathcal{Y})$ are absolutely continuous distributions that have positive density functions on their domain. Then, the solution $(V^\star, T^\star)$ of equation (9) is unique. In other words, $V^\star \in S_c$ is unique up to constant, and $T^\star$ is a deterministic OT Map.*

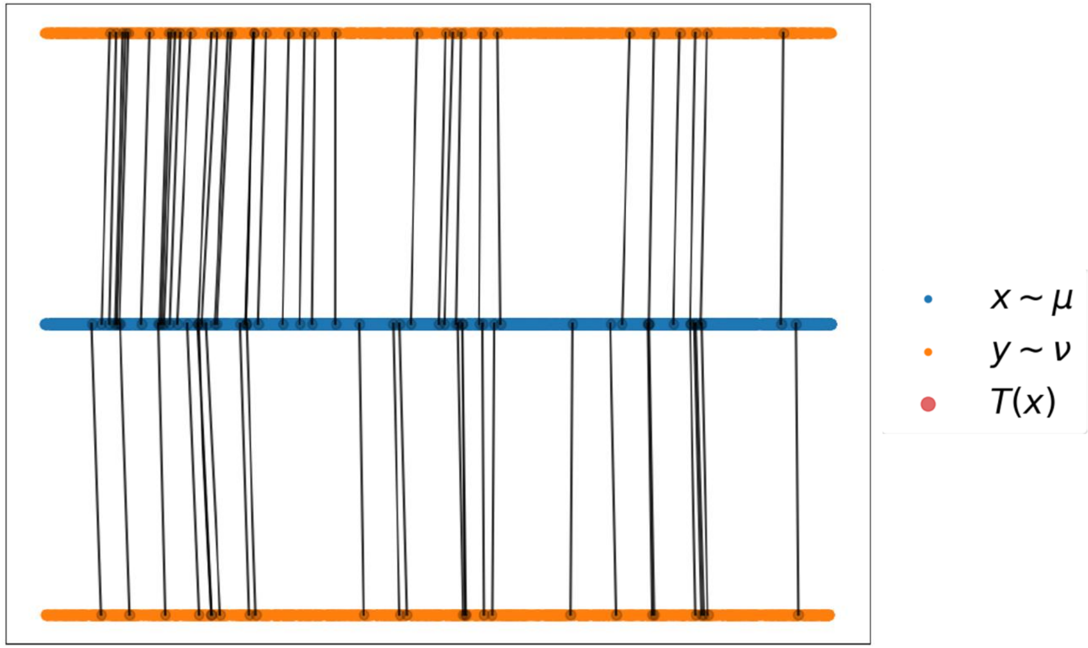
Failure Cases When Our Conditions Is Not Met

When only Stochastic Transport $\pi^\star$ exists

- Assume that the **source μ** and **target ν** distributions are uniformly supported on

$$A = [0,1] \times \{0\} \quad \text{and} \quad B = [0,1] \times \{1\} \cup [0,1] \times \{-1\}.$$

- The **standard SNOT** parameterizes the transport map with as **deterministic function** T_θ.
    - **When only an OT Plan π^* exists, the SNOT cannot accurately represent the stochastic OT Map.**



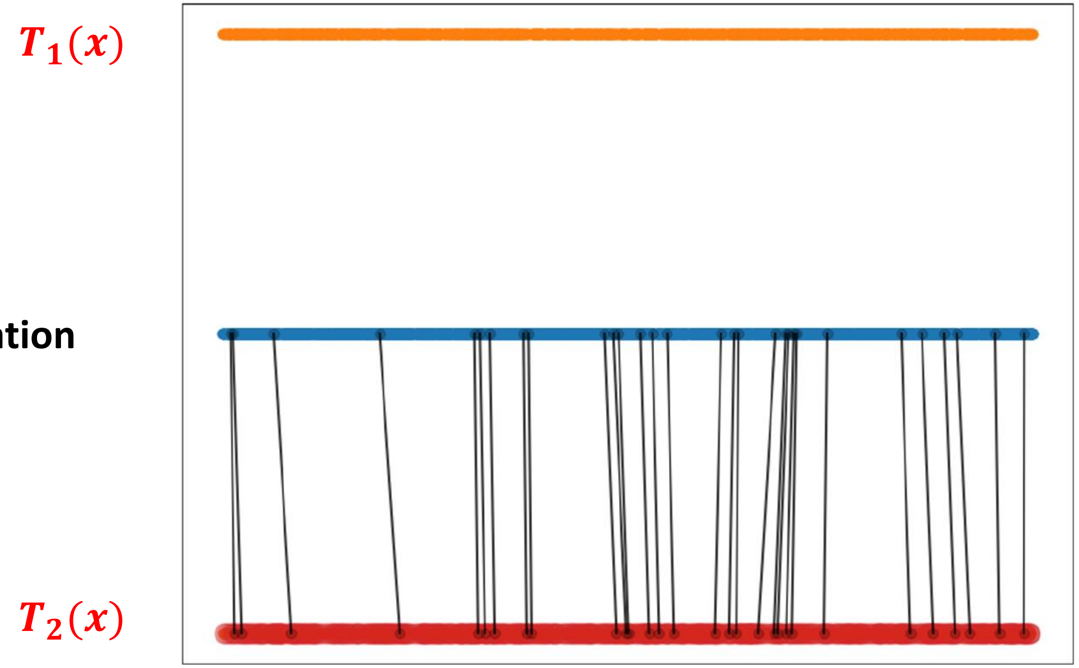
Failure Cases When Our Conditions Is Not Met

When only Stochastic Transport π^* exists

- Any function $\boldsymbol{T(x) = T_1(x)}$ **or** $\boldsymbol{T_2(x)}$ becomes a max-min solution of SNOT.

$$T_1(x) = (x_1, 1) \quad \text{and} \quad T_2(x) = (x_1, -1).$$

Fake Solution Visualization



When only Stochastic Transport π^* exists

- The **naïve stochastic parametrization** of OT Map **cannot address this problem**.
  - Given a stochastic variable $z \sim N(0, I)$, the stochastic parametrization is given as:

$$T_\theta(x, z) \in \arg\min_{y \in \mathcal{Y}} \{c(x, y) - V^\star(y)\}$$

Failure Cases When Our Conditions Is Not Met

When only Stochastic Transport π^* exists

- The **naïve stochastic parametrization** of OT Map **cannot address this problem**.
  - Given a stochastic variable $z \sim N(0, I)$, the stochastic parametrization is given as:

$$T_\theta(x, z) \in \arg\min_{y \in \mathcal{Y}} \{c(x, y) - V^\star(y)\}$$

- However, even stochastic parametrization does not guarantee the true solution.
  - It only ensures that the **transport plan $\pi^\dagger$ is supported on the subdifferential**, i.e., $\pi^\dagger(\partial_c V^\star) = 1$.

Proposition 3.4 (Informal). *Assume that the stochastic parametrization of $T_\theta(x, z)$ is ideally trained as in equation (17) for $(\mu, \mathcal{N})$-a.s. $\mathcal{D}_x$ in Eq. 10 may not uniquely determined and $T_\theta(x, z)$ may contain fake solutions.*

Failure Cases Visualization

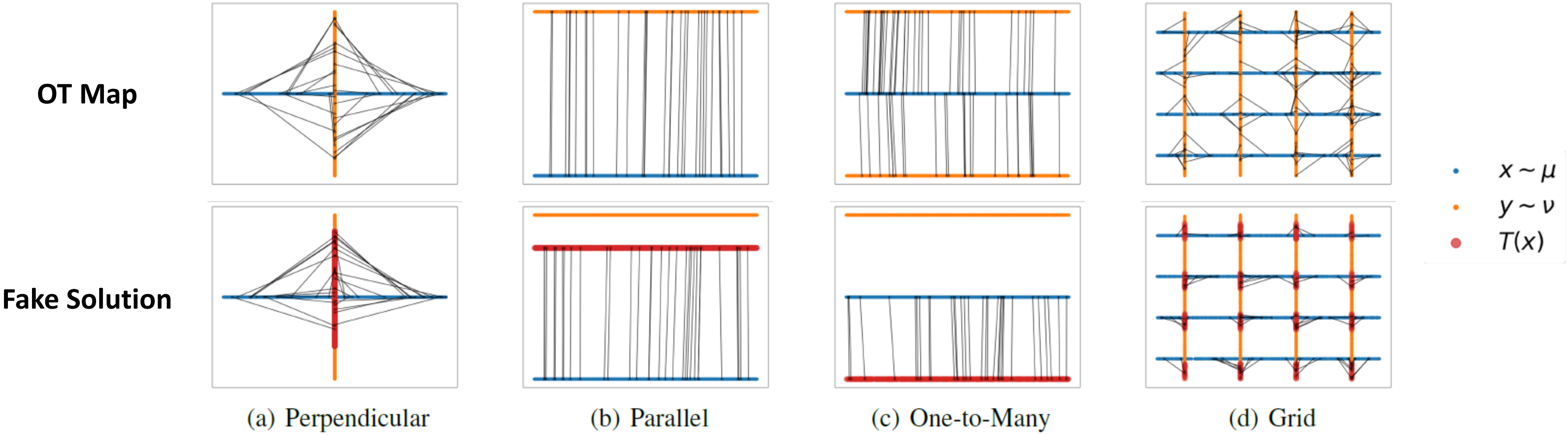


Figure 1. **Visualization of failure cases** by comparing the Optimal Transport map (**1st row**) and the max-min solution (**2nd row**) of Semi-dual Neural OT in the failure cases. The source data $x \sim \mu$, target data $y \sim \nu$, and generated data $T(x)$ are represented in Blue, Orange, and Red. The max-min solution fails to recover the correct OT Map.

Optimal Transport Plan (OTP) Model

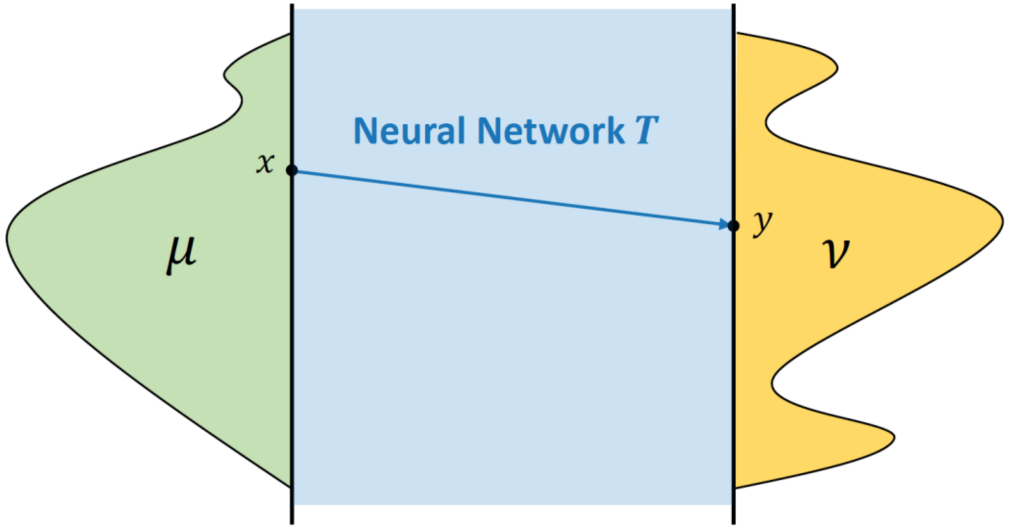
Optimal Transport Plan (OTP)

Proposed Method

- Our goal is to learn the **OT Plan $\pi^\star$** between the source distribution μ and the target distribution ν.

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

- **Why the stochastic OT Plan?**
  - **Thm 3.1 is the inherent property of μ.** When this condition is not satisfied, the existence of OT Map $T^\star$ is not guaranteed.



Optimal Transport Plan (OTP)

Proposed Method

- Our goal is to learn the **OT Plan $\pi^\star$** between the source distribution μ and the target distribution ν.

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

- **Why the stochastic OT Plan?**
  - **Thm 3.1 is the inherent property of μ**. When this condition is not satisfied, the existence of OT Map $T^\star$ is not guaranteed.

- **Our method consists of two steps:**
  1. Introduce the **smoothed version** of the source distribution μ_ϵ.
  2. **Gradually adjust μ_ϵ** back to the original source measure μ.

Optimal Transport Plan (OTP)

OTP Model

- We require **two conditions on the smoothed measure $\boldsymbol{\mu_\epsilon}$**:
  - (c1) μ_ϵ does not give mass to the measurable sets of Hausdorff dim at most $d-1$ dim.
  - (c2) μ_{ϵ_k} weakly converges to μ as $k \to \infty$.

  → *Allows the SNOT to recover the true OT Plan* and *ensures the convergence of the OT Plan*.

- **Two options for the smoothing distribution.**
  - Gaussian convolution $\mu_{\epsilon_k} = \mu * N(0, \epsilon_k I)$
  - Variance-preserving convolution $\mu_{\epsilon_k} = \left(\sqrt{1-\epsilon_k} Id\right)_\# \mu * N(0, \epsilon_k I)$

  with a predefined noise schedule $\epsilon_k \searrow 0$

  - For noise-level scheduling, we follow diffusion model [1].

[1] Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *ICLR 2021.*

OTP Model

- **Our method consists of two steps:**

1. Apply SNOT on the smoothed measure μ_ϵ

   - For each level ϵ_k, the **max-min solution** of $\mathcal{L}^k_{V,T}$ **recovers the OT Map $\boldsymbol{T}^*$** and the Kantorovich potential V^*.

$$\mathcal{L}^k_{V_\phi,T_\theta} = \sup_{V_\phi} \left[\int_{\mathcal{X}} \inf_{T_\theta} \left[c\left(x, T_\theta(x)\right) - V_\phi\left(T_\theta(x)\right)\right] d\mu_{\epsilon_k}(x) + \int_{\mathcal{X}} V_\phi(y) d\nu(y) \right].$$

Optimal Transport Plan (OTP)

OTP Model

- **Our method consists of two steps:**

1. Apply SNOT on the smoothed measure μ_ϵ

   - For each level ϵ_k, the **max-min solution** of $\mathcal{L}^k_{V,T}$ **recovers the OT Map T^*** and the Kantorovich potential V^*.

$$\mathcal{L}^k_{V_\phi,T_\theta} = \sup_{V_\phi} \left[\int_{\mathcal{X}} \inf_{T_\theta} \left[c\left(x, T_\theta(x)\right) - V_\phi\left(T_\theta(x)\right)\right] d\mu_{\epsilon_k}(x) + \int_{\mathcal{X}} V_\phi(y) d\nu(y) \right].$$

2. Gradually adjust μ_ϵ back to the original source measure μ.

   - As $k \to \infty$, i.e., $\epsilon_k \searrow 0$, the **OT Plan $\pi^*_k = (Id, T^*_k)_{\#}\mu_{\epsilon_k}$ converges (up to a subsequence) to π^***.

Theorem 4.1. *Let $\{\mu_{\epsilon_k}\}_{k\in\mathbb{N}}$ be a sequence absolutely continuous probability measures, and $T^\star_k$ be the OT map from μ_{ϵ_k} to μ. If μ_{ϵ_k} weakly converges to μ as $k \to \infty$, then $\pi^\star_k = (Id, T^\star_k)_{\#}\mu_{\epsilon_k}$ weakly converges to the OT plan $\pi^\star$ between μ and ν, along a subsequence. Consequently, $\pi^\star_k$ from our OTP model with either convolution above also weakly converges to $\pi^\star$, along a subsequence.*

Optimal Transport Plan (OTP)

Training Algorithm

Algorithm 1 Training algorithm of OTP

Require: Source distribution μ and the target distribution ν; OT Map network T_θ and potential network V_ϕ; Total number of iteration K; Number of inner-loop iterations K_T; Decreasing sequence of noise levels $\{\epsilon_k\}_{k=1}^{K}$.

for $k = 0, 1, 2, \ldots, K$ **do**
    Sample a batch $x \sim \mu$, $y \sim \nu$, $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
    $\tilde{x} \leftarrow x + \sqrt{\epsilon_k} z$ or $\tilde{x} \leftarrow \sqrt{1-\epsilon_k} x + \sqrt{\epsilon_k} z$.
    Update ϕ to maximize $\mathcal{L}_\phi = -V_\phi\left(T_\theta(\tilde{x})\right) + V_\phi(y)$.
    for $j = 0, 1, \ldots, K_T$ **do**
        Sample a batch $x \sim \mu$, $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
        $\tilde{x} \leftarrow x + \sqrt{\epsilon_k} z$ or $\tilde{x} \leftarrow \sqrt{1-\epsilon_k} x + \sqrt{\epsilon_k} z$.
        $\mathcal{L}_\theta = c(\tilde{x}, T_\theta(\tilde{x})) - V_\phi\left(T_\theta(\tilde{x})\right) + V_\phi(y)$.
        Update θ to minimize $\mathcal{L}_\theta$.
    end for
end for

Experiments

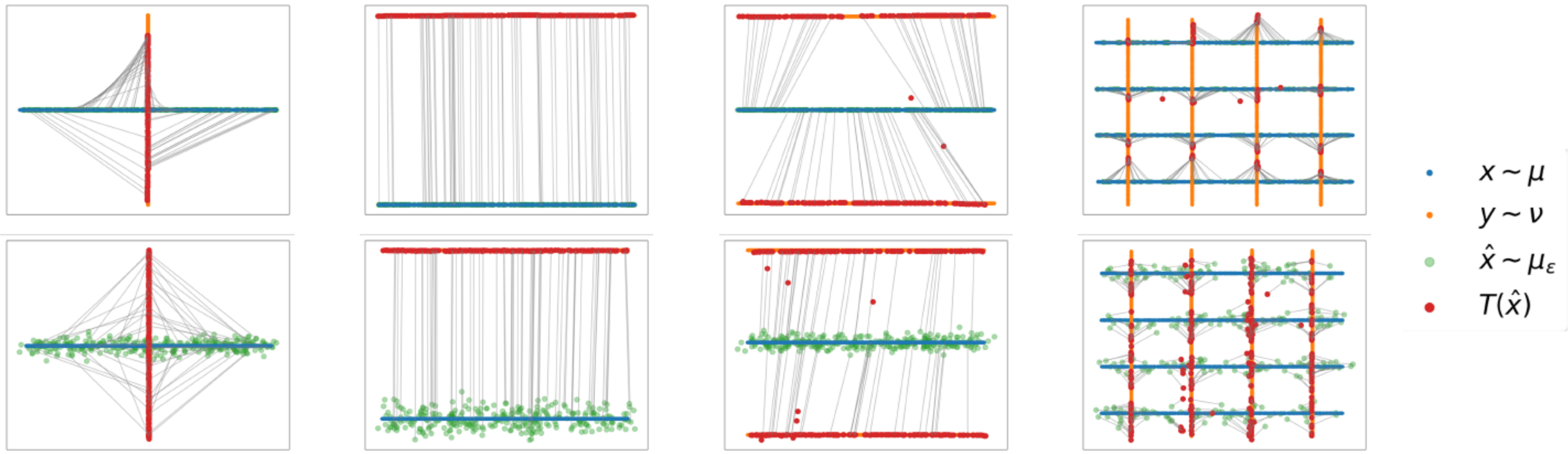
Experiments

Optimal Transport Plan Evaluation

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

Existing Model

Our model



(a) Perpendicular (b) Parallel (c) One-to-Many (d) Grid

Figure 3. **Qualitative comparison between OTM (1st row) and our model (2nd row) on failure cases** in Sec 3.2. The noised source sample $\tilde{x}$ in Alg 1 is denoted in Green. While OTM falls into fake solutions and fails to generate the target distribution correctly, our OTP model successfully learns the OT Plan.

Optimal Transport Plan Evaluation

- We evaluate our model against **SNOT (OTM)** and the **SNOT with a stochastic generator (OTM-s)**.
  - Transport cost error $D_{cost} = |W_2^2(\mu,\nu) - \int \|T_\theta(x) - x\|^2 d\mu(x)|$
  - Target distribution error $D_{target} = W_2^2(T_{\theta\#}\mu,\nu)$
  - Our model outperforms particularly in **high-dimensional settings**.

Table 1. **Quantitative comparison of numerical accuracy** on synthetic datasets. Each model is evaluated using two metrics: transport cost error $D_{cost}(\downarrow)$ and target distribution error $D_{target}(\downarrow)$.

Dimension	Model	Perpendicular		One-to-Many	
		D_{cost}	D_{target}	D_{cost}	D_{target}
$d=2$	OTM	0.038	0.0079	0.069	0.10
	OTM-s	**0.0070**	0.018	0.35	**0.032**
	Ours	0.019	**0.0068**	**0.0022**	0.11
$d=4$	OTM	0.043	0.039	0.10	73.23
	OTM-s	**0.033**	0.065	**0.010**	**0.038**
	Ours	0.089	**0.0086**	0.033	0.094
$d=16$	OTM	0.16	4.97	71.28	73.23
	OTM-s	0.061	4.85	97.49	99.57
	Ours	**0.058**	**0.59**	**0.057**	**0.65**
$d=64$	OTM	2.13	19.37	21.92	32.94
	OTM-s	2.74	18.79	0.20	12.21
	Ours	**0.97**	**10.09**	**0.14**	**9.98**

Neural OT Evaluation on I2I Translation Tasks

- **Stochastic Transport Application (MNIST-to-CMNIST)**
  - The naïve stochastic generator fails to learn the stochastic map.

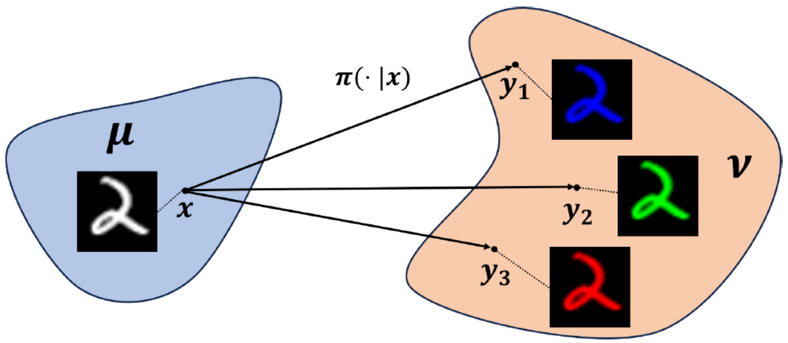
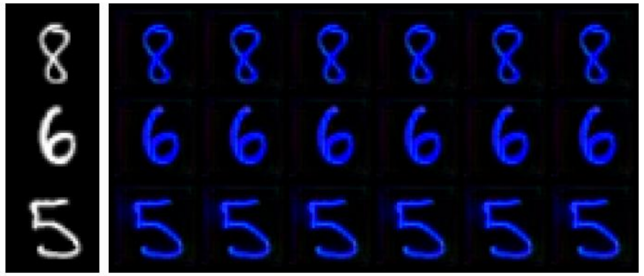
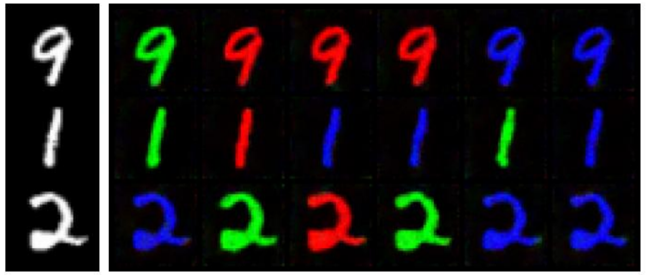


Figure 2. **Example of a stochastic transport map (OT Plan) task**, e.g., MNIST-to-CMNIST colorization.



(a) OTM-s (FID=62.4, LPIPS=0.36)



(b) Ours (FID=3.18, LPIPS=0.32)

Figure 4. **Experimental results on a stochastic transport map application**, i.e., MNIST-to-CMNIST translation.

Neural OT Evaluation on I2I Translation Tasks

- **Image-to-Image Translation Benchmark**
  - We assessed our model on several Image-to-Image (I2I) translation benchmarks.

Table 2. **Image-to-Image translation benchmark** results compared to existing Neural (Entropic) OT models † indicates the results conducted by ourselves. DSBM scores are taken from (Gushchin et al., 2024; De Bortoli et al., 2024).

Data	Model	FID (↓)	LPIPS (↓)
Male-to-Female (64x64)	NOT (Korotin et al., 2023b)	11.96	-
	OTM† (Fan et al., 2022)	6.42	**0.16**
	DIOTM† (Choi et al., 2024a)	**4.48**	0.20
	OTP (Ours)	4.75	0.20
Wild-to-Cat (64x64)	DSBM (Shi et al., 2024)	20+	0.59
	OTM† (Fan et al., 2022)	12.42	0.47
	DIOTM† (Choi et al., 2024a)	10.72	**0.45**
	OTP (Ours)	**9.66**	0.52
Male-to-Female (128x128)	DSBM (Shi et al., 2024)	37.8	0.25
	ASBM (Gushchin et al., 2024)	16.08	-
	OTM† (Fan et al., 2022)	7.55	**0.21**
	DIOTM† (Choi et al., 2024a)	7.40	0.25
	OTP (Ours)	**6.38**	0.27

Conclusion

- Neural Optimal Transport is a powerful framework for generative modeling and image-to-image translation, but existing methods often suffer from **fake solutions**.
- We identify a **sufficient condition** that guarantees the avoidance of such fake solutions in Semi-dual Neural OT.
- We propose the **Optimal Transport Plan (OTP) model**, which introduces smoothing on the source measure to enable more reliable and accurate transport plan learning.
- Our OTP model **successfully recovers the correct OT Plan** in failure cases where existing models fail.
- These advancements **enhance the reliability** of Neural OT, making it more effective for various machine learning applications.

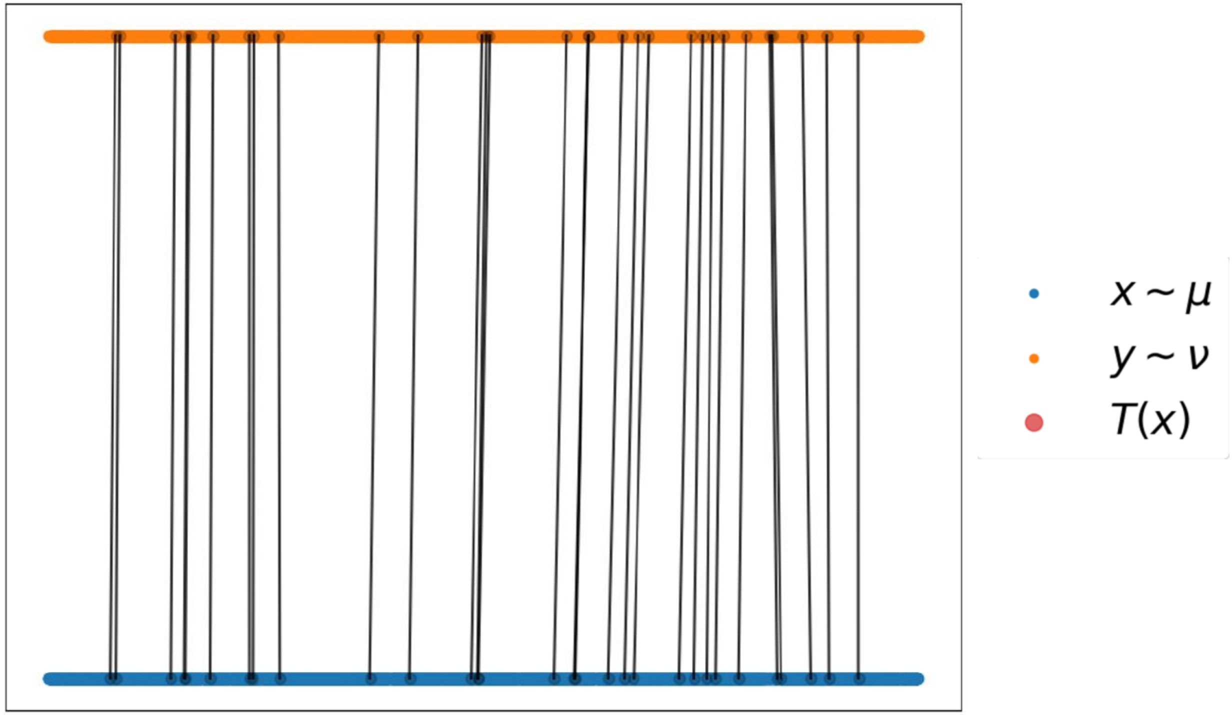
Thank you!

Failure Cases When Our Conditions Is Not Met

When unique $T^{\star}$ exists

- The previous examples is the case when T^* exists but is not unique.
- Assume that the **source μ** and **target ν** distributions are uniformly supported on

Not Absolutely Continuous $A = [-1, 1] \times \{0\}$ and $B = [-1,1] \times \{1\}$.



Failure Cases When Our Conditions Is Not Met

When unique $T^\star$ exists

- Here, the optimal potential $V^\star$ and the OT Map $T^\star$ are as follows:

$$V^\star(y) = \frac{1}{2}\|y_2\|^2 \quad \text{for } y = (y_1, y_2) \qquad \text{and} \qquad T^\star(x) = (x_1, 1) \quad \text{for } x = (x_1, x_2).$$

- **Any function $T\big((x_1, x_2)\big) \coloneqq (x_1, a)$** for any $a \in \mathbb{R}$ becomes a max-min solution of SNOT.

$$T_\theta : x \mapsto \arg\min_{y \in \mathcal{Y}} \left[c(x,y) - V_\phi(y)\right]$$

$$\inf_T \int_{\mathcal{X}} \frac{1}{2}\|x_1 - T(x)_1\|^2 d\mu(x) + \int_{\mathcal{Y}} \frac{1}{2}\|y\|^2 d\nu(y) = \frac{1}{2}.$$

No dependency on the second component

Fake Solution Visualization