Minimax Rates of Estimation for Optimal Transport Maps between Infinite Dimensional Spaces

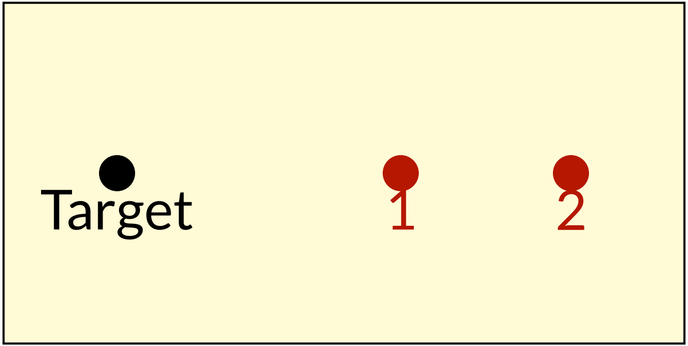
Donlapark Ponnoprat [1]   Masaaki Imaizumi [2,3]

[1]Chiang Mai University   [2]The University of Tokyo   [3]RIKEN AIP

Overview

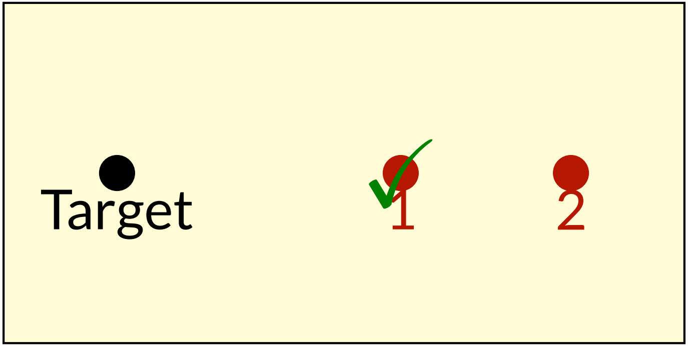
- Introduction to Optimal Transport
- Optimal Transport Map in Finite Dimensions
- Optimal Transport Map in Infinite Dimensions

Motivating Question



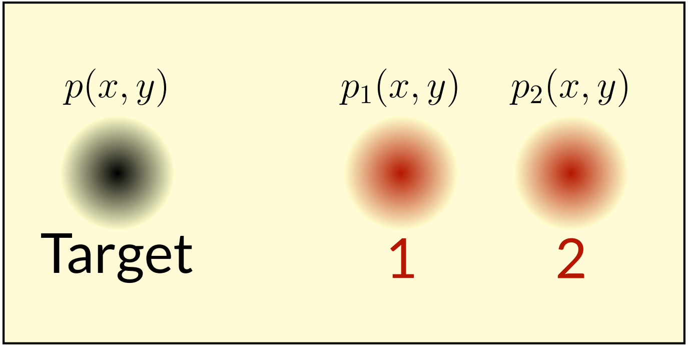
Which is closer, 1 or 2?

Motivating Question



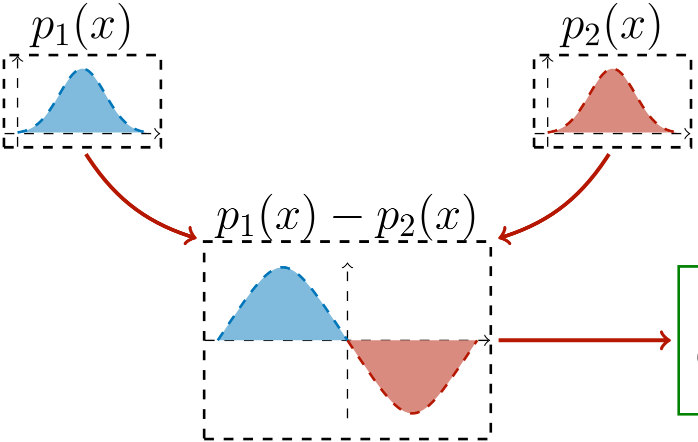
Which is closer, 1 or 2?

Fuzzy Version



Which is closer, 1 or 2?

Typical Measurement

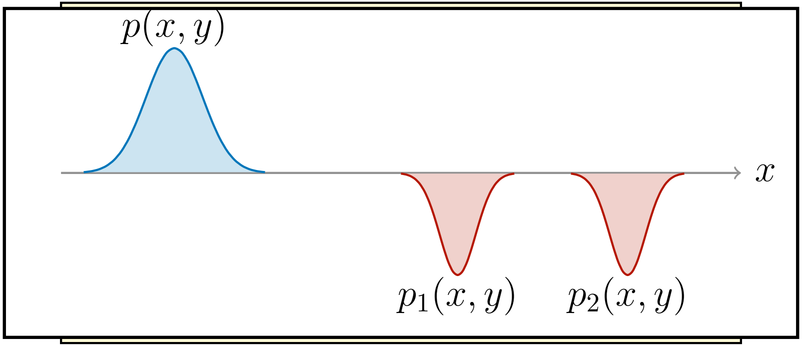


$$\boxed{d\big(p_1, p_2\big)}$$

L^p norm

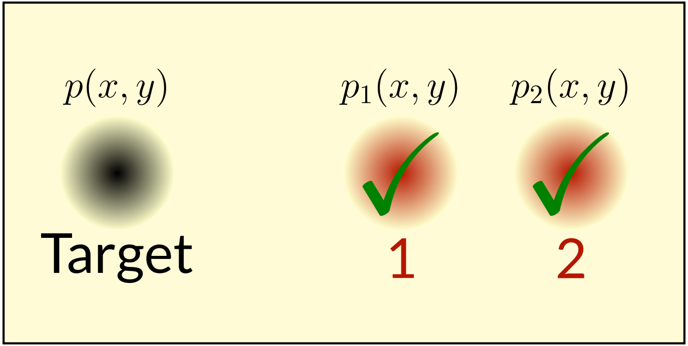
KL divergence

Returning to the Question



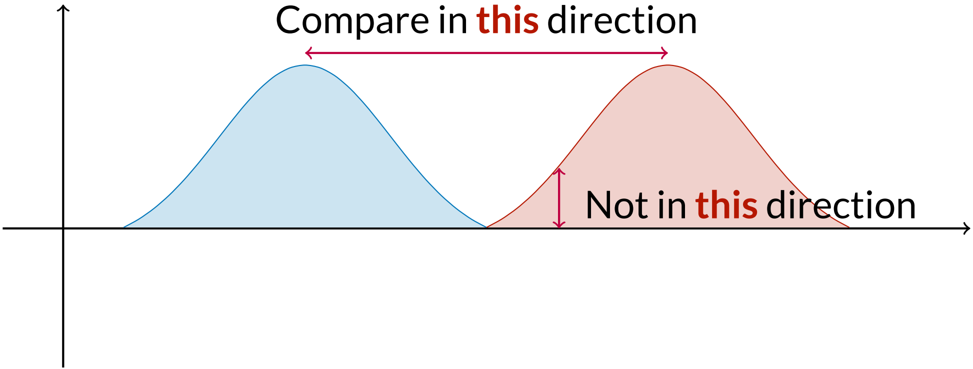
Which is closer, 1 or 2?

Returning to the Question

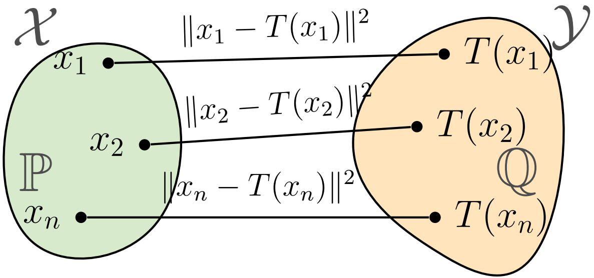


Both have the same distance!

Alternative Idea



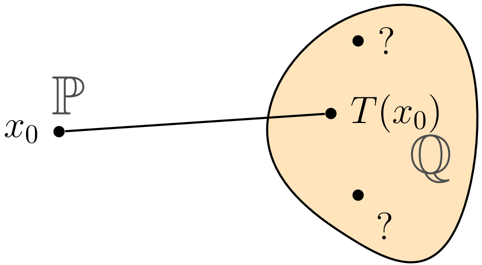
Monge's Formulation



$$T_{\#}\mathbb{P} = \mathbb{Q} \iff X \sim \mathbb{P} \Rightarrow Y \sim \mathbb{Q}.$$

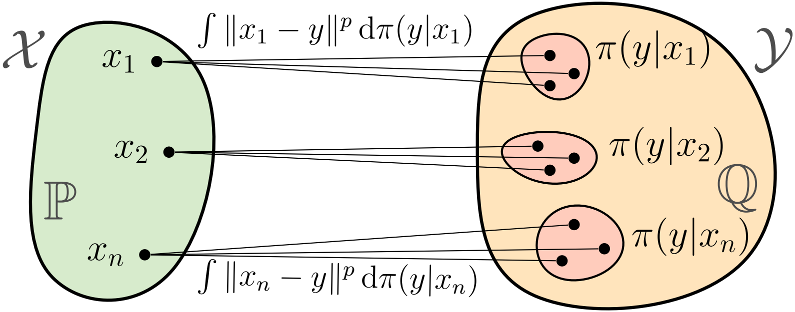
Monge's Formulation

Monge's Formulation might not have a solution!



Example: $\mathrm{d}\mathbb{P}(x) = \delta_{x_0}(x)\,\mathrm{d}x$, $\mathrm{d}\mathbb{Q}(y) = \frac{1}{2\pi}e^{-y^2/2}\,\mathrm{d}y$.
There's no T such that $T_{\#}\mathbb{P} = \mathbb{Q}$!

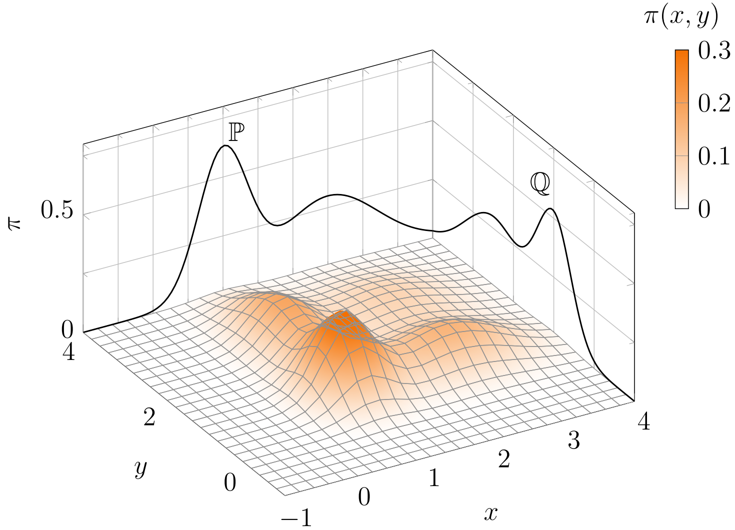
Wasserstein-p Distance



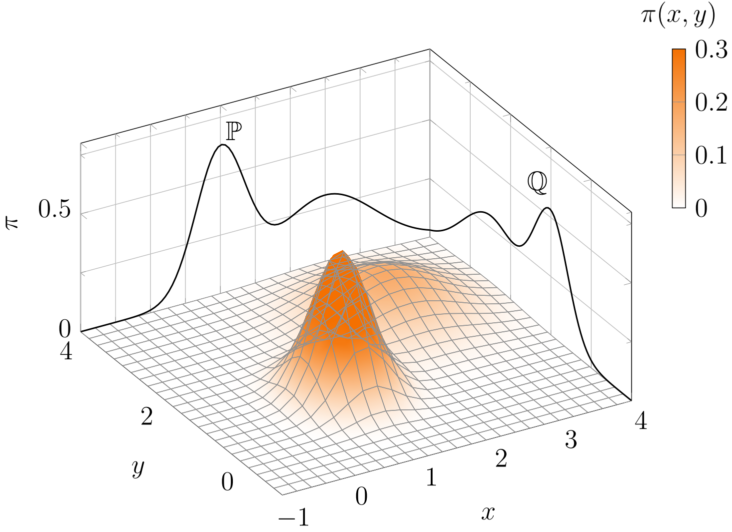
$$W_p^p(\mathbb{P}, \mathbb{Q}) = \inf_{\pi} \int_{\mathcal{X}\times\mathcal{Y}} \|x - y\|^p \,\mathrm{d}\pi(x, y)$$

s.t. π is a **coupling** (i.e. marginals $\mathbb{P}$ on $\mathcal{X}$ and $\mathbb{Q}$ on $\mathcal{Y}$).

Example of Couplings



Example of Couplings



Kantorovich Problem

Primal form

$$\inf_{\pi} \int_{\mathcal{X}\times\mathcal{Y}} \|\boldsymbol{x}-\boldsymbol{y}\|^{2} \, \mathrm{d}\pi(x,y)$$

s.t. π is a coupling of $\mathbb{P}$ and $\mathbb{Q}$

Dual form

$$\sup_{\boldsymbol{f}\in L^1(\mathbb{P}),\boldsymbol{g}\in L^1(\mathbb{Q})} \int_{\mathcal{X}} \boldsymbol{f} \, \mathrm{d}\mathbb{P} + \int_{\mathcal{Y}} \boldsymbol{g} \, \mathrm{d}\mathbb{Q}$$

s.t. $\boldsymbol{f}(x)+\boldsymbol{g}(y) \leq \|\boldsymbol{x}-\boldsymbol{y}\|^{2}, \forall (x,y) \in \mathcal{X}\times\mathcal{Y}$

Easier to optimize: Real-valued functions instead of coupling.

Primal and Dual Form Solutions

Let $c = \|x - y\|^2$.

$$\inf_{\pi \text{ coupling}} \int c \, \mathrm{d}\pi(x, y)$$

$$= \sup_{f,g} \inf_{\pi \text{ positive}} \int c - f(x) - g(y) \, \mathrm{d}\pi(x, y) + \int f \, \mathrm{d}\mathbb{P} + \int g \, \mathrm{d}\mathbb{Q}$$

$= -\infty$ if $c - f(x) - g(y) < 0$ for some x, y

Complementary Slackness

If (π_0, f_0, g_0) is a solution:

$$\pi_0(x^\star, y^\star) > 0 \iff f_0(x^\star) + g_0(y^\star) = \|x^\star - y^\star\|^2$$

Solving the Dual Form

Dual form

$$\sup_{f \in L^1(\mathbb{P}), g \in L^1(\mathbb{Q})} \int_{\mathcal{X}} f \, \mathrm{d}\mathbb{P} + \int_{\mathcal{Y}} g \, \mathrm{d}\mathbb{Q}$$

s.t. $f(x) + g(y) \leq \|x - y\|^2, \forall (x, y) \in \mathcal{X} \times \mathcal{Y}$

For any $(x^\star, y^\star) \in \mathrm{Supp}(\pi_0)$,

$f_0(x^\star) + g_0(y^\star) = \|x^\star - y^\star\|^2 \Rightarrow x^\star$ maximizes $\|x - y^\star\|^2 - f_0(x)$

(assume f_0 is differentiable) $\Rightarrow x^\star$ solves $\nabla \|x - y^\star\|^2 = \nabla f_0(x)$

$$\Rightarrow y^\star = x^\star - \frac{1}{2} \nabla f_0(x^\star).$$

Existence and Unique of Transport Map

Brenier's theorem

Let $\mathbb{P}$ be absolutely continuous wrt the Lebesgue measure and both $\mathbb{P}$ and $\mathbb{Q}$ have finite second moments. Then, there exists a unique optimal transport map $T_0 : \mathcal{X} \to \mathcal{Y}$, defined by:

$$T_0(x) = x - \frac{1}{2}\nabla \boldsymbol{f_0}(x) = \nabla\underbrace{(\frac{1}{2}\|x\|^2 - \frac{1}{2}\boldsymbol{f_0}(x))}_{\boldsymbol{\varphi_0}(x)}.$$

for some differentiable function $\boldsymbol{f_0}$. In addition, $\boldsymbol{\varphi_0}$ is convex.

c-transform

To simplify the dual problem, we define the c-**transform**:

$$f^c(y) = \inf_x \|x - y\|^2 - f(x).$$

Plugging $g = f^c$ in the dual problem:

Kantorovich's Semi-Dual Problem

$$\sup_{f \in L^1(\mathbb{P})} \int_{\mathcal{X}} f \, \mathrm{d}\mathbb{P} + \int_{\mathcal{Y}} f^c \, \mathrm{d}\mathbb{Q}.$$

(f_0, g_0) solves dual problem $\Longrightarrow$ f_0 solves semi-dual problem

Relating back to T_0

Plugging $f(x) = \|x\|^2 - 2\varphi(x)$ in the Kantorovich's problem:

Brenier's Semi-Dual Problem

$$\inf_{\varphi \in L^1(\mathbb{P})} \int_{\mathcal{X}} \varphi \, \mathrm{d}\mathbb{P} + \int_{\mathcal{Y}} \varphi^* \, \mathrm{d}\mathbb{Q},$$

where φ^* is the convex conjugate of φ:

$$\varphi^*(y) = \sup_x \langle x, y \rangle - \varphi(x).$$

φ_0 solves Brenier problem $\Longrightarrow \nabla \varphi_0$ is the transport map.

Summary of Semi-Dual Problems

Brenier's Problem

$$\inf_{\varphi \in L^1(\mathbb{P})} \int_{\mathcal{X}} \varphi \, \mathrm{d}\mathbb{P} + \int_{\mathcal{Y}} \varphi^* \, \mathrm{d}\mathbb{Q}$$

Kantorovich's Problem

$$\sup_{f \in L^1(\mathbb{P})} \int_{\mathcal{X}} f \, \mathrm{d}\mathbb{P} + \int_{\mathcal{Y}} f^c \, \mathrm{d}\mathbb{Q}.$$

$\downarrow$ $\downarrow$

$$f_0(x) = \|x\|^2 - 2\varphi_0(x)$$

$\varphi_0 \longleftrightarrow f_0$

Brenier potential — Kantorovich potential

Estimation

Assume Brenier's assumptions hold $\Longrightarrow T_0$ exists

Goal: Estimate T_0 from data:

$X_1, \ldots, X_n \sim \mathbb{P}$ and $Y_1, \ldots, Y_n \sim \mathbb{Q}$

$$\downarrow$$

Plug-In Estimator

$$\widehat{\varphi}_n = \operatorname*{argmin}_{\varphi \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) + \frac{1}{n} \sum_{j=1}^{n} \varphi^*(Y_j).$$

$$\widehat{T}_n = \nabla \widehat{\varphi}_n$$

over some function class $\mathcal{F}$.

Minimax Problem

Setup: Assume that T_0 lies in some function class $\mathcal{F}_0$.

Minimax Risk

$$R_n(\mathcal{F}_0) = \inf_{T_n} \sup_{T_0 \in \mathcal{F}_0} \int \|T_n(x) - T_0(x)\|^2 \, \mathrm{d}\mathbb{P}(x)$$

The **estimator** that minimizes the **worst-case estimation error**.

Minimax Lower Bound: Euclidean Case

Minimax Lower Bound [Hütter and Rigollet, 2021]

If $\alpha > 1, \underset{\text{bounded}}{\Omega \subset \mathbb{R}^d}$ and the true transport map lies in:

$$\mathcal{F} = \{T \in C^{\alpha}(\Omega), M^{-1} \prec \nabla T \prec M,$$
$$T = \nabla \varphi \text{ for some differentiable convex } \varphi : \Omega \to \mathbb{R}\},$$

then the following bound holds:

Rate improves with smoothness

$$R_n(\mathcal{F}) \gtrsim n^{-\frac{2\alpha}{2\alpha - 2 + d}}$$

Curse of dimensionality

Optimal Estimators

Which estimator $\widehat{T}_n$ achieves the minimax rate?

Target Error Rate

$$\int \|\widehat{T}_n(x) - T_0(x)\|^2 \, \mathrm{d}\mathbb{P}(x) \approx n^{-\frac{2\alpha}{2\alpha-2+d}}.$$

Plug-In Estimator

$$\widehat{\varphi}_n = \operatorname*{argmin}_{\varphi \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) + \frac{1}{n} \sum_{j=1}^{n} \varphi^*(Y_j).$$

$$\widehat{T}_n = \nabla \widehat{\varphi}_n,$$

is minimax optimal when:

- $\mathcal{F} = \{$finite wavelet series, must be strongly convex$\}$ [Hütter and Rigollet, 2021]
- $\mathcal{F} = \{$finite wavelet series$\}$ or $\{$ReQU neural networks$\}$ [Divol et al., 2025]

Estimators from smoothed densities

Estimate densities $\hat{\boldsymbol{p}}_{\boldsymbol{n}}(x)$ and $\hat{\boldsymbol{q}}_{\boldsymbol{n}}(y)$ using kernel or wavelets.

Let $\widehat{\mathbb{P}}_{\boldsymbol{n}}$ and $\widehat{\mathbb{Q}}_{\boldsymbol{n}}$ be distributions w/ densities $\hat{\boldsymbol{p}}_{\boldsymbol{n}}(x)$ and $\hat{\boldsymbol{q}}_{\boldsymbol{n}}(y)$.

$$\widehat{T}_n = \underset{T_{\#}\widehat{\mathbb{P}}_{\boldsymbol{n}} = \widehat{\mathbb{Q}}_{\boldsymbol{n}}}{\operatorname{argmin}} \int \|T(x) - x\|^2 \, \mathrm{d}\widehat{\mathbb{P}}_{\boldsymbol{n}}.$$

$\widehat{T}_n$ is minimax optimal for both kernel and wavelet estimators [Manole et al., 2021].

Extending to Infinite Dimensions

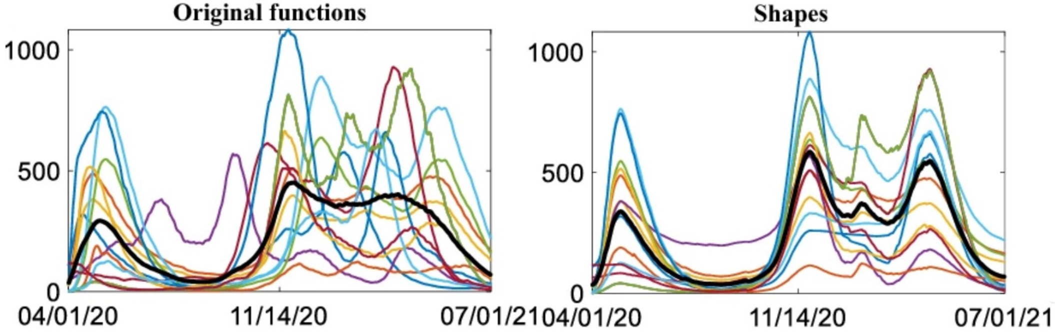
Optimal transport from $X \in [0,1]^\infty$ to $Y \in [0,1]^\infty$.

Why?

- We could gain some insight into combatting the curse of dimensionality.

- Application to functional data analysis:
  - We can view function $f \in L^2([0,1])$ as infinite sequence $(a_1, a_2, \ldots)$ via Fourier or Wavelet transform.
  - Example: Matching time series between two individuals.

Time series alignment



Hospitalization rates (per million) for 16 European countries from April 1, ’20, to July 1, ’21.

[Wu et al., 2023]

Optimal Transport in Infinite Dimensions

Let $\mathcal{H}$ be a separable Hilbert space with basis $\{e_i\}_{i=1}^{\infty} \subset \mathcal{H}$.

Definition

A Borel-measurable set $E \subset \mathcal{H}$ is a **Gaussian null set** if $\mu(E) = 0$ for any Gaussian measure[1] μ on $\mathcal{H}$.

[1](measure of variables of the form $a + \sum_{i=1}^{\infty} X_i e_i$ where $X_i \sim N(0, 1)$.)

Optimal Transport in Infinite Dimensions

Brenier's Theorem on $\mathcal{H}$ [Ambrosio et al., 2005]

Let $\mathbb{P}$, $\mathbb{Q}$ be prob. measures with supports $\Omega_{\mathbb{P}}, \Omega_{\mathbb{Q}} \subset \mathcal{H}$.

Assume $\mathbb{P}$ **is zero on all Gaussian null sets**, $\mathbb{Q}$ has a bounded support, and both $\mathbb{P}$ and $\mathbb{Q}$ have finite second moments.

Then, there exists a unique transport map:

$$T_0 : \Omega_{\mathbb{P}} \to \Omega_{\mathbb{Q}},$$

and a convex function $\varphi_0 \in L^1(\mathbb{P})$ such that $\nabla \varphi_0 = T_0$.

Minimax Lower Bound for C^2 Functions

Theorem (Informal) [Ponnoprat and Imaizumi, 2025]

There is a subset $\Omega \subset [0,1]^\infty$ such that, if the transport map T_0 lies in:

$$\mathcal{F} \subset \{T = \nabla\varphi \mid \varphi \in C^2(\Omega)\},$$

$\sum_{i=1}^\infty |\partial_i \varphi| + \sum_{i,j=1}^\infty |\partial_i \partial_j \varphi| \leq B$

then the following bound holds:

$$R_n(\mathcal{F}) \gtrsim \frac{1}{\log n}.$$

Very slow rate!

New Function Space for φ_0

Let $\mathbb{Z}_0^\infty = \{(\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots) \mid \boldsymbol{l}_i \in \mathbb{Z}, \boldsymbol{l}_i = 0$ for all but finitely many $i\}$

Fact: Any function $\varphi \in L^2([0,1]^\infty)$ can be approximated by Fourier series:

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots) \approx \sum_{\boldsymbol{l} \in \mathbb{Z}_0^\infty} f_{\boldsymbol{l}} \psi_{\boldsymbol{l}_1}(\boldsymbol{x}_1) \psi_{\boldsymbol{l}_2}(\boldsymbol{x}_2) \psi_{\boldsymbol{l}_3}(\boldsymbol{x}_3) \cdots ,$$

where $f_l \in \mathbb{R}$ and

$$\psi_{\boldsymbol{l}_i}(\boldsymbol{x}_i) = \begin{cases} \sqrt{2}\cos(2\pi|\boldsymbol{l}_i|\boldsymbol{x}_i), & \boldsymbol{l}_i < 0 \\ \sqrt{2}\sin(2\pi|\boldsymbol{l}_i|\boldsymbol{x}_i), & \boldsymbol{l}_i > 0 \\ 1, & \boldsymbol{l}_i = 0. \end{cases}$$

Grouping Frequencies

Let's denote

$$\psi_{\boldsymbol{l}}(\boldsymbol{x}) = \psi_{\boldsymbol{l}_1}(\boldsymbol{x}_1)\psi_{\boldsymbol{l}_2}(\boldsymbol{x}_2)\psi_{\boldsymbol{l}_3}(\boldsymbol{x}_3)\cdots.$$

For each $\boldsymbol{s} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots) \in \mathbb{N}^\infty$, we gather all terms between two consecutive **dyadic frequencies:** $2^{\boldsymbol{s}_i - 1}$ and $2^{\boldsymbol{s}_i}$:

$$\delta_{\boldsymbol{s}}(f) = \sum_{\boldsymbol{l} \in \mathbb{Z}_0^\infty : 2^{\boldsymbol{s}_i - 1} \leq |l_i| < 2^{\boldsymbol{s}_i}} f_{\boldsymbol{l}} \psi_{\boldsymbol{l}}.$$

Function Space for T_0

- We want to characterize the **Brenier potential** $\boldsymbol{\varphi_0}$ by Fourier series.
- However, φ_0 has to be convex! There is no nice way to describe convexity in terms of Fourier coefficients.
- So we use the Fourier series for the **Kantorovich potential** $\boldsymbol{f_0}$ instead.

Function Space for T_0

Specify the **direction-wise smoothness levels**:

$$\boldsymbol{a} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots)$$

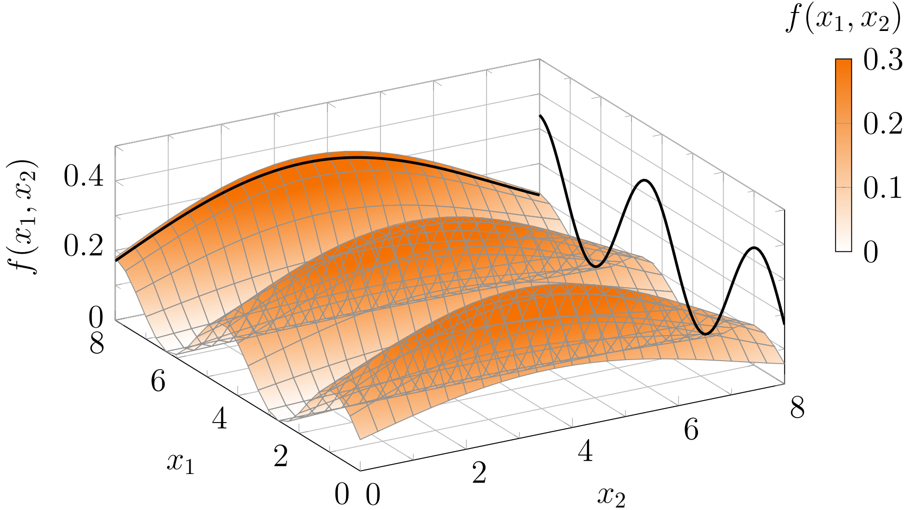
Characterize smoothness of **Kantorovich potential** f_0 by the decay of each $\delta_s(\boldsymbol{f})$:

must be small for large s_i and $\boldsymbol{a}_i$

$$\mathcal{F}_{\boldsymbol{a}} = \left\{ T = \nabla\left(\frac{1}{2}\|\cdot\|^2 - \frac{1}{2}\boldsymbol{f}\right) \;\middle|\; \underbrace{\sum_{s \in \mathbb{N}_0^\infty} 2^{2\sum_i \boldsymbol{a}_i s_i} \|\delta_s(\boldsymbol{f})\|_{L^2}^2}_{\|\boldsymbol{f}\|_{H^{\boldsymbol{a}}}} \le 1 \right\}.$$

[Okumoto and Suzuki, 2021]

For example, $\boldsymbol{a} = (1, 2, \infty, \infty, \ldots)$ represents $\boldsymbol{f} : [0,1]^2 \to \mathbb{R}$ that is twice as smooth along x_2 compared to x_1.



Minimax Lower Bound

Theorem [Ponnoprat and Imaizumi, 2025]

Suppose that $\mathbb{P}$ and $\mathbb{Q}$ satisfy all the assumptions of the Brenier's Theorem, and that $\boldsymbol{a}_1 \leq \boldsymbol{a}_2 \leq \dots$

The minimax lower bound of learning an optimal transport map from $\mathcal{F}_{\boldsymbol{a}}$ is:

$$R_n(\mathcal{F}_{\boldsymbol{a}}) \gtrsim n^{-\frac{2\boldsymbol{a}_1}{2\boldsymbol{a}_1+1}}.$$

Our Proposed Estimator

Specify a positive integer $\boldsymbol{J}$, and then solve:

Plug-In Estimator

$$\widehat{\varphi}_{n,\boldsymbol{a},\boldsymbol{J}} = \operatorname*{argmin}_{\varphi \in \mathcal{F}_{\boldsymbol{a},\boldsymbol{J}}} \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i) + \frac{1}{n}\sum_{j=1}^{n} \varphi^*(Y_j).$$

$$\widehat{T}_{n,\boldsymbol{a},\boldsymbol{J}} = \nabla \widehat{\varphi}_{n,\boldsymbol{a},\boldsymbol{J}}$$

$$\mathcal{F}_{\boldsymbol{a},\boldsymbol{J}} = \left\{ \varphi = \frac{1}{2}\|\cdot\|^2 - \frac{1}{2}\boldsymbol{f} \;\middle|\; \boldsymbol{f} = \sum_{|l_i| \lesssim 2^{\boldsymbol{J}/\boldsymbol{a}_i}} f_l \psi_l, \|\boldsymbol{f}\|_{H^{\boldsymbol{a}}} \leq 1 \right\}$$

Optimality of the Plug-In Estimator

Theorem [Ponnoprat and Imaizumi, 2025]

Suppose that $\mathbb{P}$ and $\mathbb{Q}$ satisfy all the assumptions of the Brenier's Theorem, $\boldsymbol{a}_1 \leq \boldsymbol{a}_2 \leq \dots$, and $\boldsymbol{a}_i \approx i^q$ for some $q > 0$.

Let $\boldsymbol{J} \approx \log n$.

Then, the estimator $\widehat{T}_{n,\boldsymbol{a},\boldsymbol{J}}$ satisfies the following bound:

$$\int \|\widehat{T}_{n,\boldsymbol{a},\boldsymbol{J}}(x) - T_0(x)\|^2 \, \mathrm{d}\mathbb{P}(x) \lesssim n^{-\frac{2\boldsymbol{a}_1}{2\boldsymbol{a}_1+1}}.$$

Impracticality of Plug-In Estimator

$$\widehat{\varphi}_{n,\boldsymbol{a},\boldsymbol{J}} = \operatorname*{argmin}_{\varphi} \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i) + \frac{1}{n}\sum_{j=1}^{n} \varphi^*(Y_j),$$

$$\text{s.t. } \varphi = \frac{1}{2}\|\cdot\|^2 - \frac{1}{2}\boldsymbol{f}, \quad \boldsymbol{f} = \sum_{|l_i| \lesssim 2^{\boldsymbol{J}/\boldsymbol{a}_i}} f_l \psi_l, \|\boldsymbol{f}\|_{H^{\boldsymbol{a}}} \leq 1 \; .$$

Ignoring the φ^* term, this is a constraint linear program over the coefficients f_l.

However, the number of the coefficients is exponential in the number of dimensions, making it impractical.

Neural Network Estimator

Instead, we consider the following function class for $\boldsymbol{f}$:

$$\mathcal{F}(W, L, R, B) = \{\text{neural networks with width} \leq W, \text{\#layers} \leq L \\ \text{\#parameters} \leq R, \text{values of parameters} \leq B\}$$

Neural Network Estimator

$$\widehat{\boldsymbol{\varphi}}_{\boldsymbol{n,WLRB}} = \operatorname*{argmin}_{\varphi} \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i) + \frac{1}{n}\sum_{j=1}^{n} \varphi^*(Y_j),$$

$$\text{s.t. } \varphi = \frac{1}{2}\|\cdot\|^2 - \frac{1}{2}\boldsymbol{f}, \quad \boldsymbol{f} \in \mathcal{F}(W, L, R, B).$$

$$\widehat{T}_{n,WLRB} = \nabla \widehat{\boldsymbol{\varphi}}_{\boldsymbol{\alpha,WLRB}}.$$

Optimality of the Neural Estimator

Theorem [Ponnoprat and Imaizumi, 2025]

Suppose that $\mathbb{P}$ and $\mathbb{Q}$ satisfy all the assumptions of the Brenier's Theorem, $\boldsymbol{a}_1 \leq \boldsymbol{a}_2 \leq \ldots$, and $\boldsymbol{a}_i \approx i^{\boldsymbol{q}}$ for some $\boldsymbol{q} > 0$.

Let $\boldsymbol{J} \approx \log n$. Then, there are some W, L, R, B depending on $\boldsymbol{J}$, $\boldsymbol{a}_1$ and $\boldsymbol{q}$ such that $\widehat{T}_{n,WLRB}$ satisfies the following bound:

$$\int \|\widehat{T}_{n,WLRB}(x) - T_0(x)\|^2 \, \mathrm{d}\mathbb{P}(x) \lesssim n^{-\frac{2\boldsymbol{a}_1}{2\boldsymbol{a}_1+1}}.$$

Computational Aspects

- Still, these estimators are computational expensive due to the convex conjugate: $\varphi^*(y) = \sup_x \langle x, y \rangle - \varphi(x)$.

- Solution #1: Use another neural network to model φ^* "Amortization" [Amos, 2023].

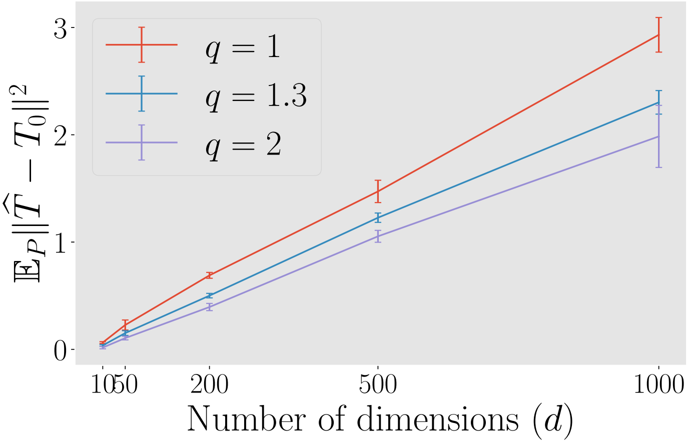
- Solution #2: Use separate networks to model the potential and the transport map [Korotin et al., 2023]:

$$\sup_{\boldsymbol{f}} \inf_{\boldsymbol{T}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \|X_i - \boldsymbol{T}(X_i)\|^2 - \boldsymbol{f}(\boldsymbol{T}(X_i)) + \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{f}(X_i) \right\}.$$

Experiment 1: No Curse of Dimensionality

$(T_0(x))_i = x_i - \frac{1}{\kappa(i)}|x_i - 0.5|^{\kappa(i)}$, $\kappa(i) \approx i^{0.1q}$. Fix $n = 100$.

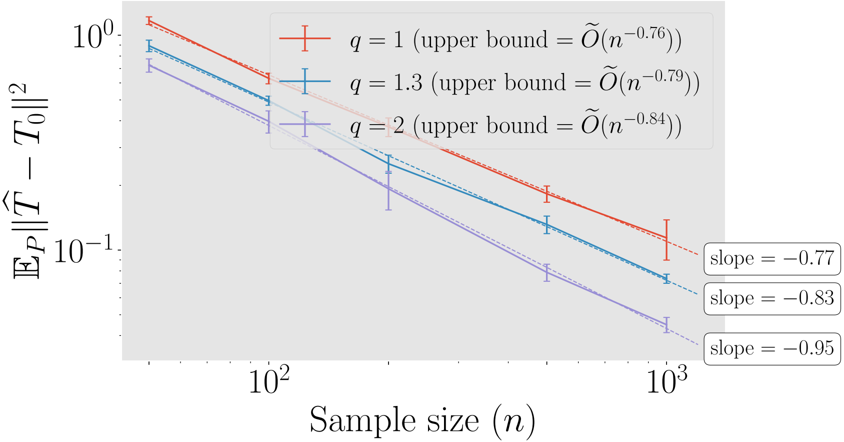
$\widehat{\varphi}_n = \frac{1}{2}\|\cdot\|^2 - \frac{1}{2}\hat{f}_n \longleftarrow$ CNN with two convolutional layers.



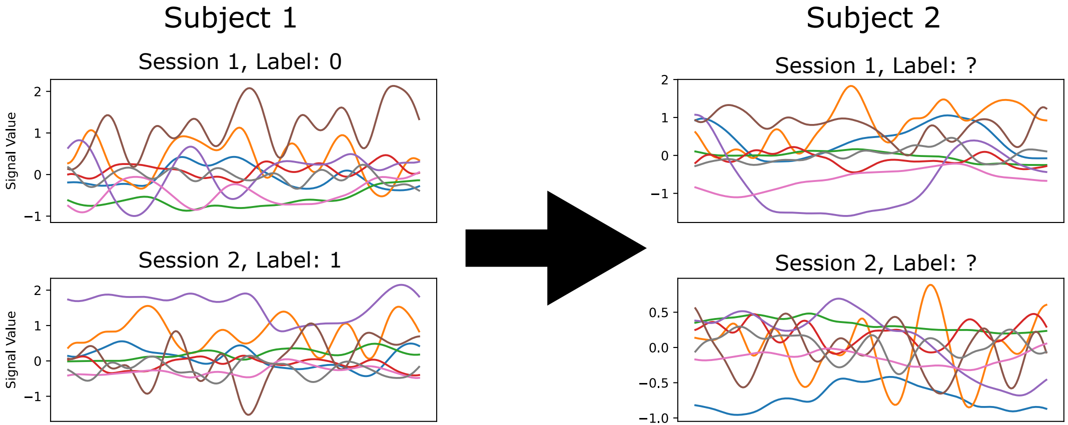
Experiment 2: Error Rate

$(T_0(x))_i = x_i - \frac{1}{\kappa(i)}|x_i - 0.5|^{\kappa(i)}$, $\kappa(i) \approx i^{0.1q}$. <u>Fix $d = 200$</u>.

$\widehat{\varphi}_n = \frac{1}{2}\|\cdot\|^2 - \frac{1}{2}\hat{f}_n \longleftarrow$ CNN with two convolutional layers.

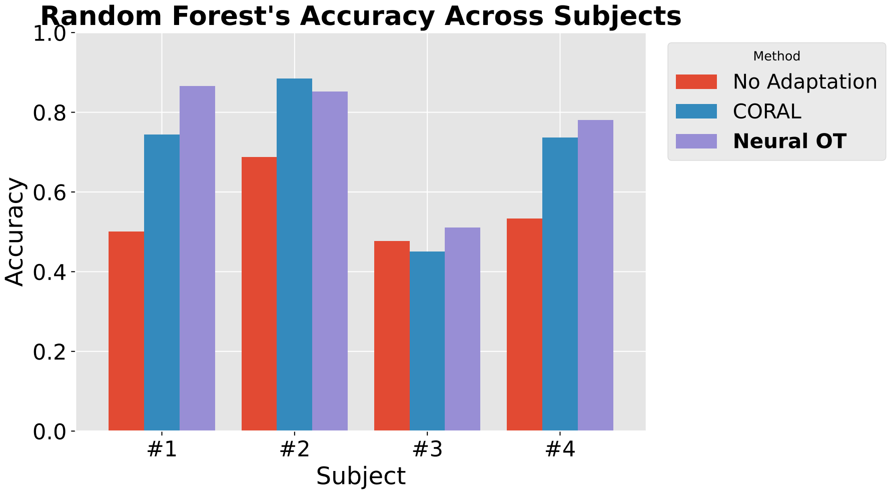


Application: Domain Adaptation



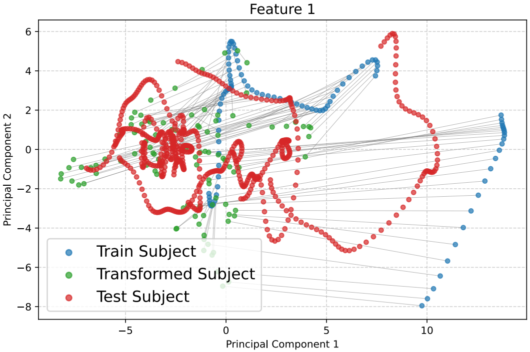
Mental workload data [Huang et al., 2021]. Label: 0 = Normal, 1 = Intense.

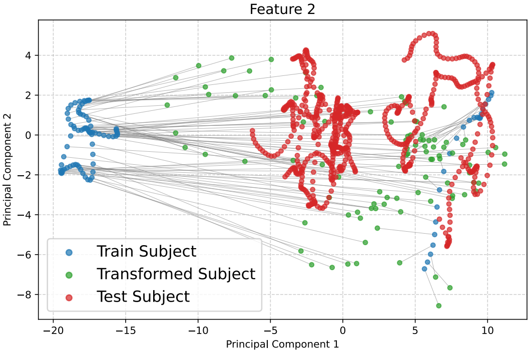
Result



Result of adapting 8 training subjects to 4 test subjects.

Functional PCA of the Transports





Conclusion

- We study the problem of optimal transport (OT) map estimation. Existing results suffer from the curse of dimensionality.

- We introduce a way to characterize the smoothness of OT maps on infinite-dimensional spaces that leads to a polynomial error rate.

Challenges

- Main challenge #1: Relaxing the assumptions.

- Main challenge #2: Finding Computational tractable algorithm with theoretical guarantees.

arXiv link: `https://arxiv.org/abs/2505.13570`

Thank you!

References I

Ambrosio, L., Gigli, N., and Savaré, G. (2005).
Gradient flows: in metric spaces and in the space of probability measures.
Springer Science & Business Media.

Amos, B. (2023).
On amortizing convex conjugates for optimal transport.
In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*.
OpenReview.net.

Divol, V., Niles-Weed, J., and Pooladian, A.-A. (2025).
Optimal transport map estimation in general function spaces.
The Annals of Statistics, 53(3):963–988.

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., and Zhang, H.-M. (2021).
Improvements of the daily optimum interpolation sea surface temperature (doisst) version 2.1.
Journal of Climate, 34(8):2923–2939.

Hütter, J.-C. and Rigollet, P. (2021).
Minimax estimation of smooth optimal transport maps.
The Annals of Statistics, 49(2).

References II

Korotin, A., Selikhanovych, D., and Burnaev, E. (2023).
Neural optimal transport.
In *The Eleventh International Conference on Learning Representations*.

Manole, T., Balakrishnan, S., Niles-Weed, J., and Wasserman, L. (2021).
Plugin estimation of smooth optimal transport maps.
arXiv preprint arXiv:2107.12364.

Okumoto, S. and Suzuki, T. (2021).
Learnability of convolutional neural networks for infinite dimensional input via mixed and anisotropic smoothness.
In *International Conference on Learning Representations*.

Ponnoprat, D. and Imaizumi, M. (2025).
Minimax rates of estimation for optimal transport map between infinite-dimensional spaces.

Wu, Y., Huang, C., and Srivastava, A. (2023).
Shape-based functional data analysis.
TEST, 33(1):1–47.