Optimal Transport in High Dimension: Obtaining Regularity and Robustness using Convexity and Projections

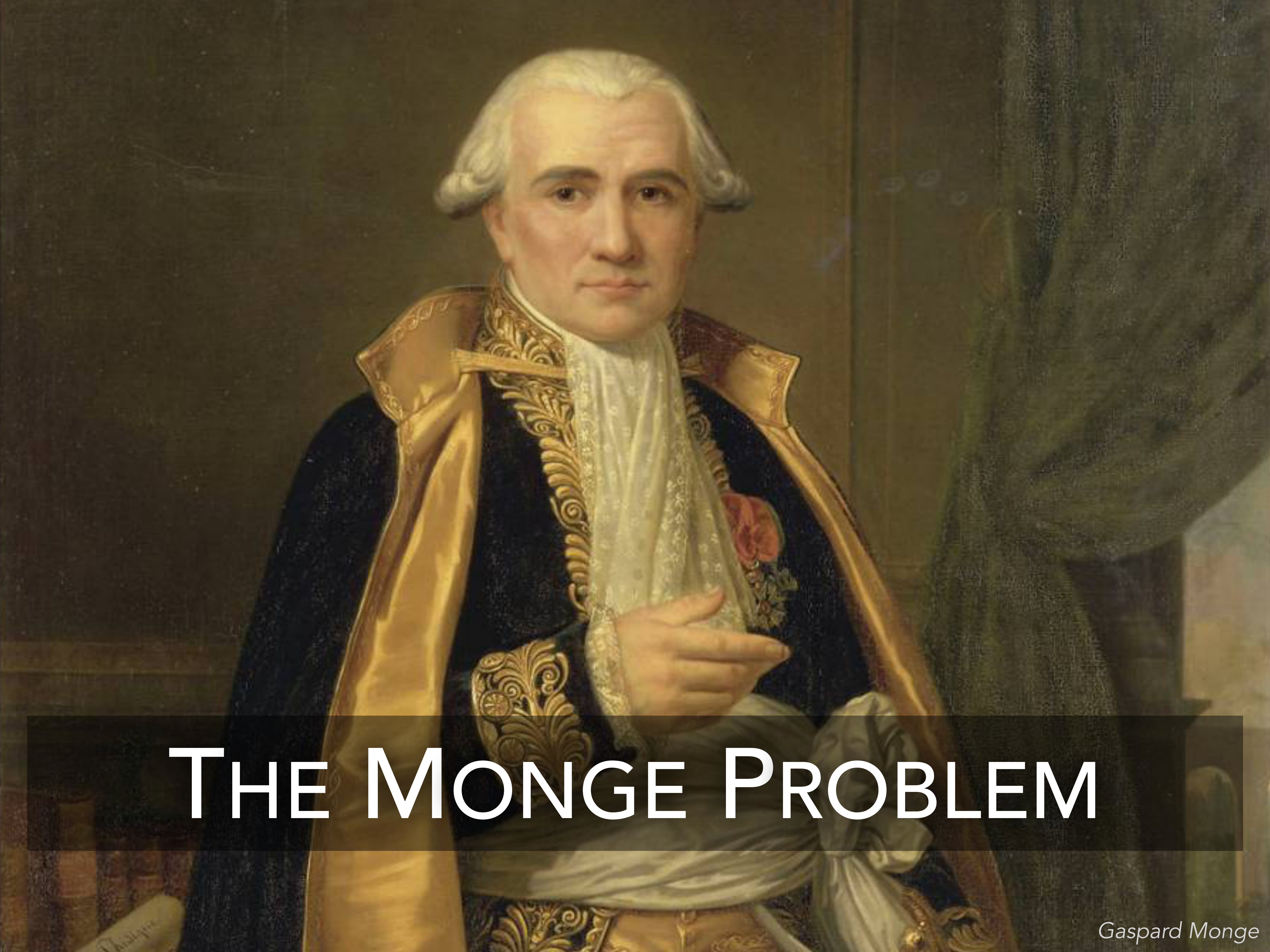
PhD Defense

June 29th, 2021

François-Pierre Paty

CREST, ENSAE, IPP

Under the supervision of Marco Cuturi



Gaspard Monge

The Monge Problem

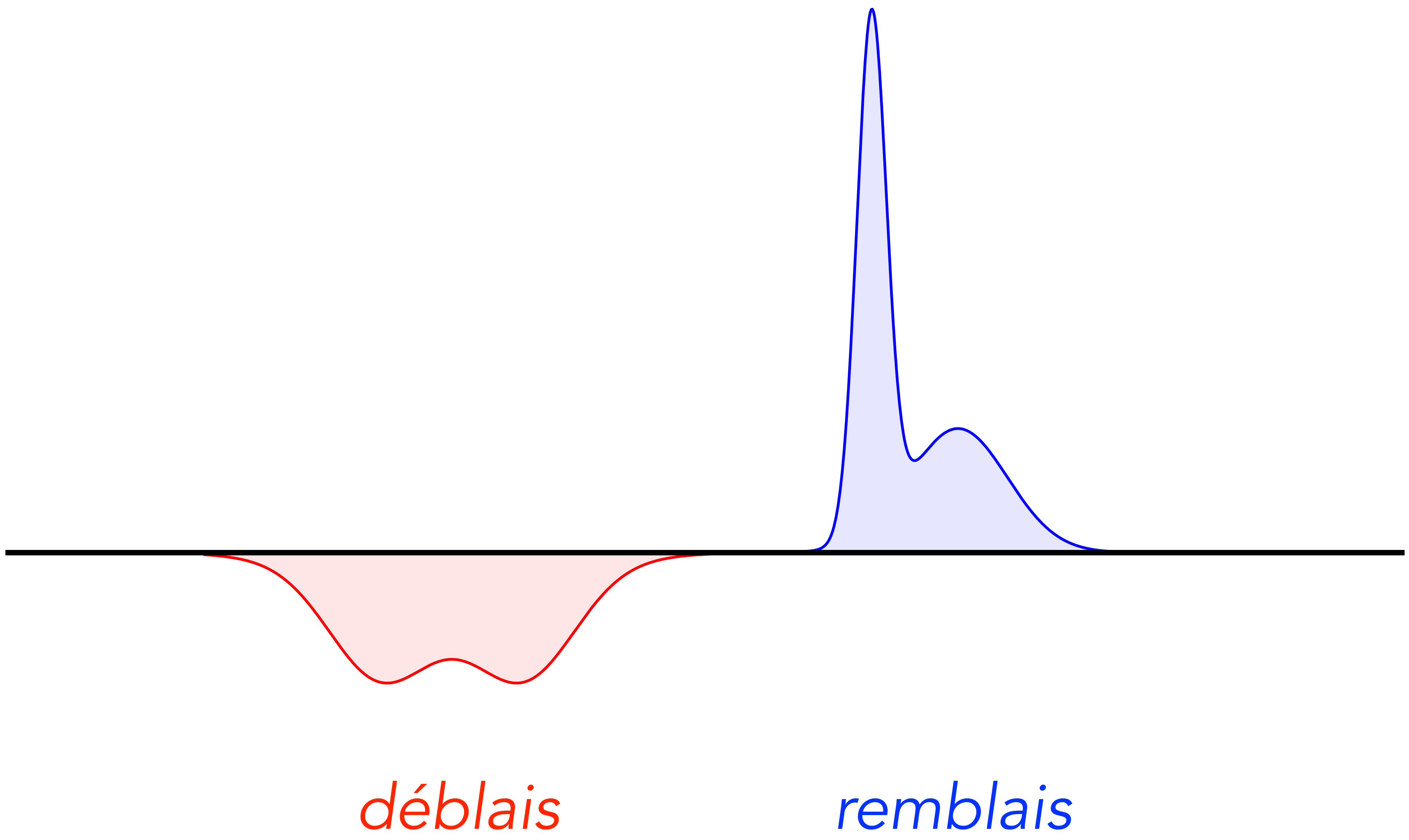
666. Mémoires de l'Académie Royale

Mémoire sur la théorie des déblais et des remblais.

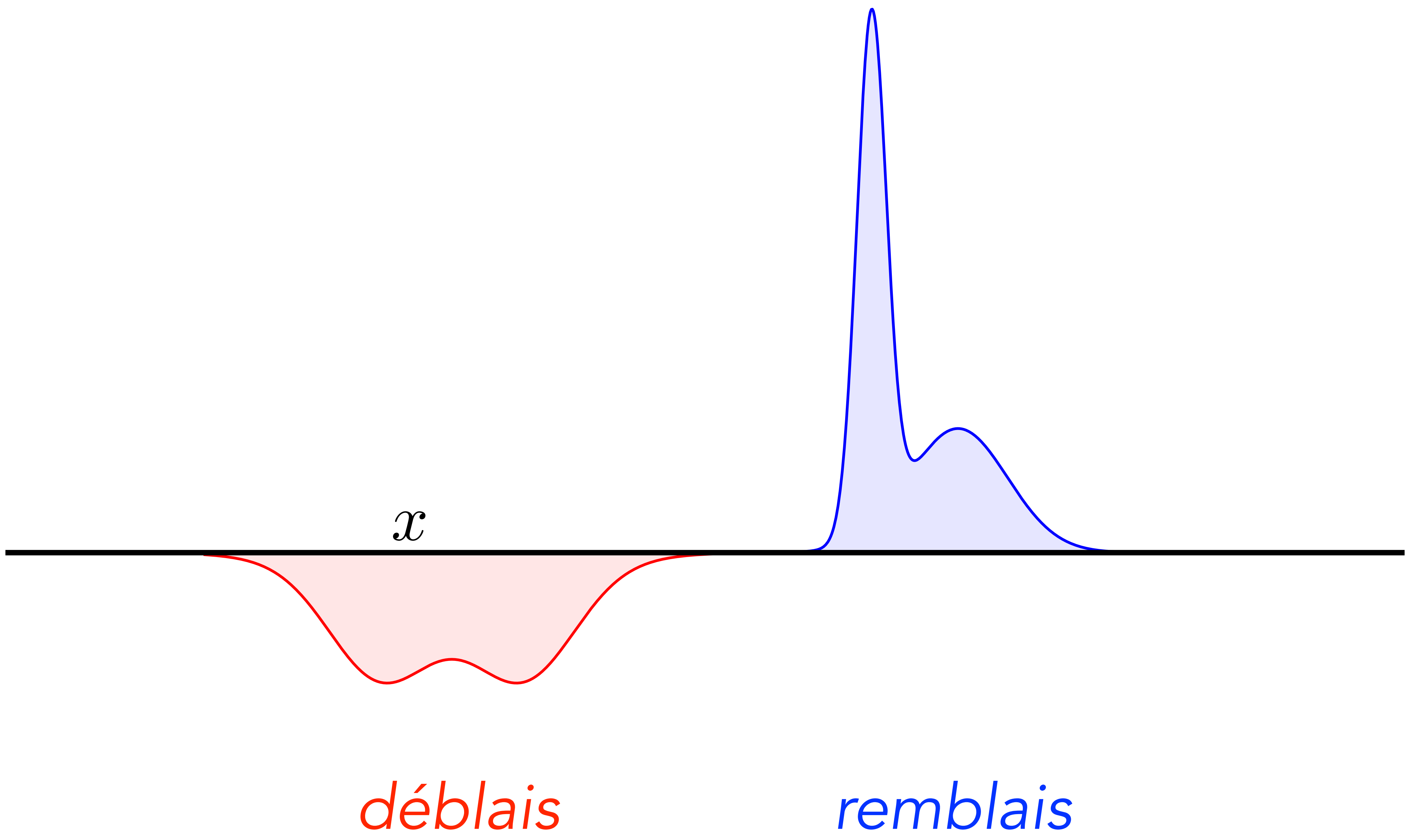
Par M. Monge.

Lorsqu'on doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

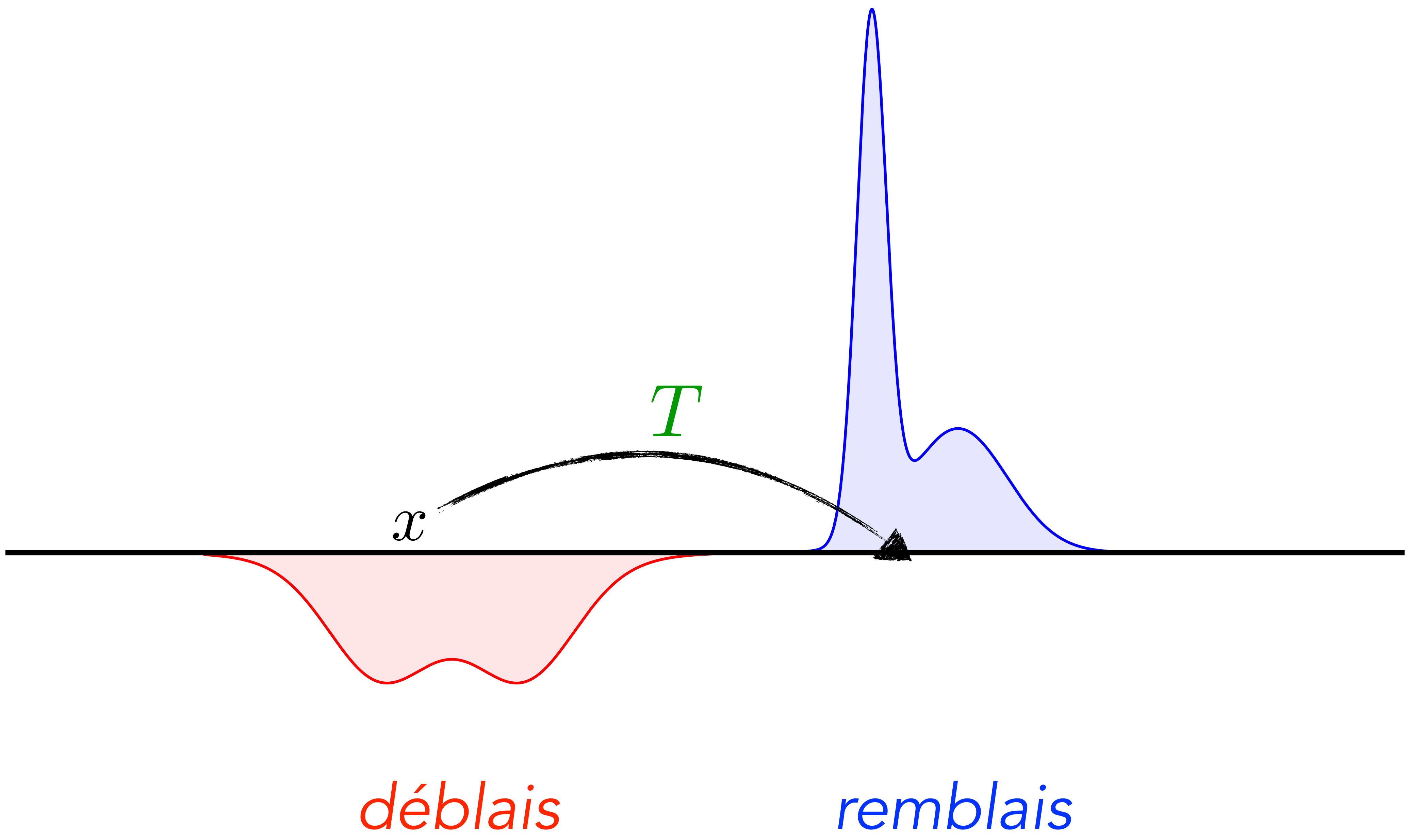
The Monge Problem



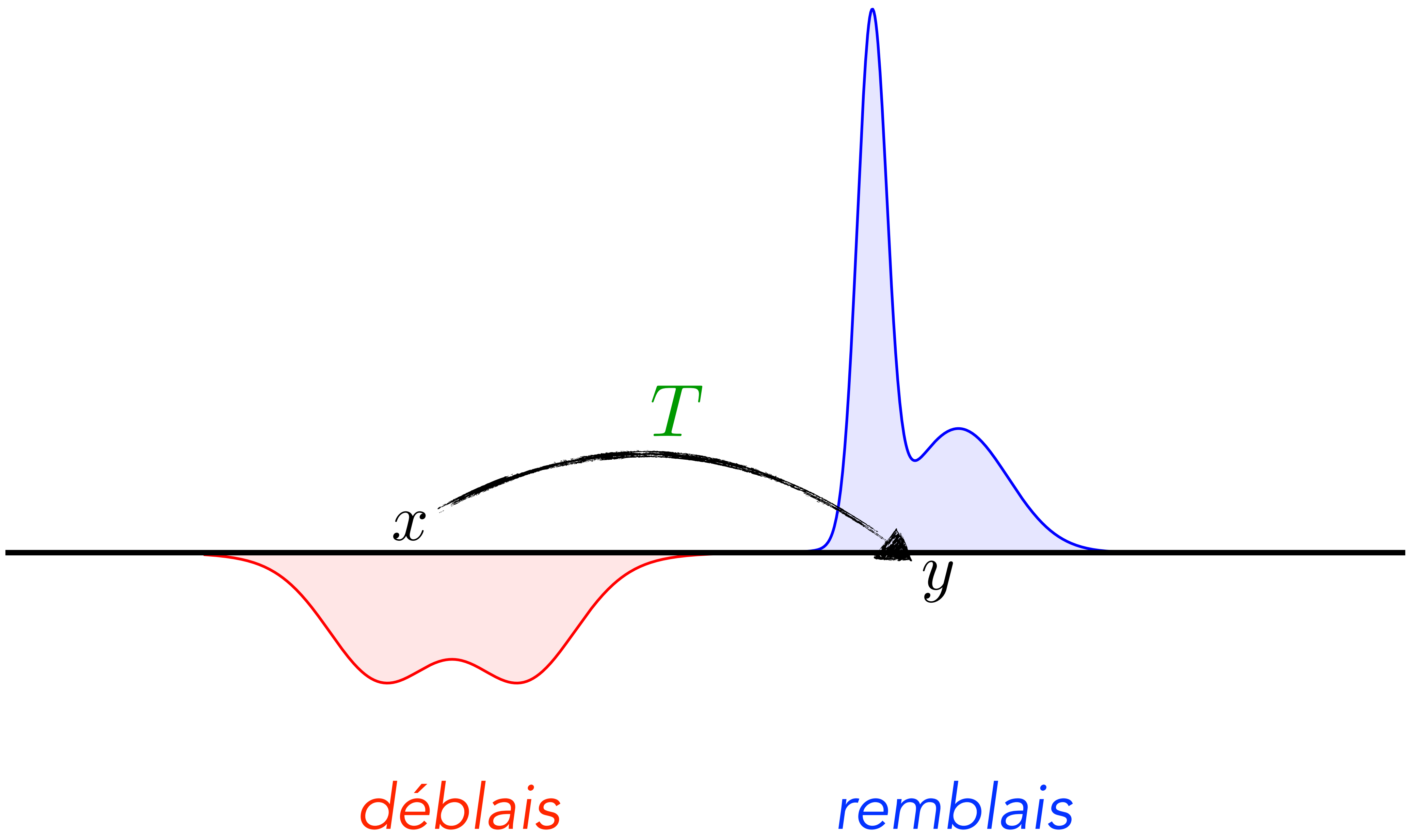
The Monge Problem



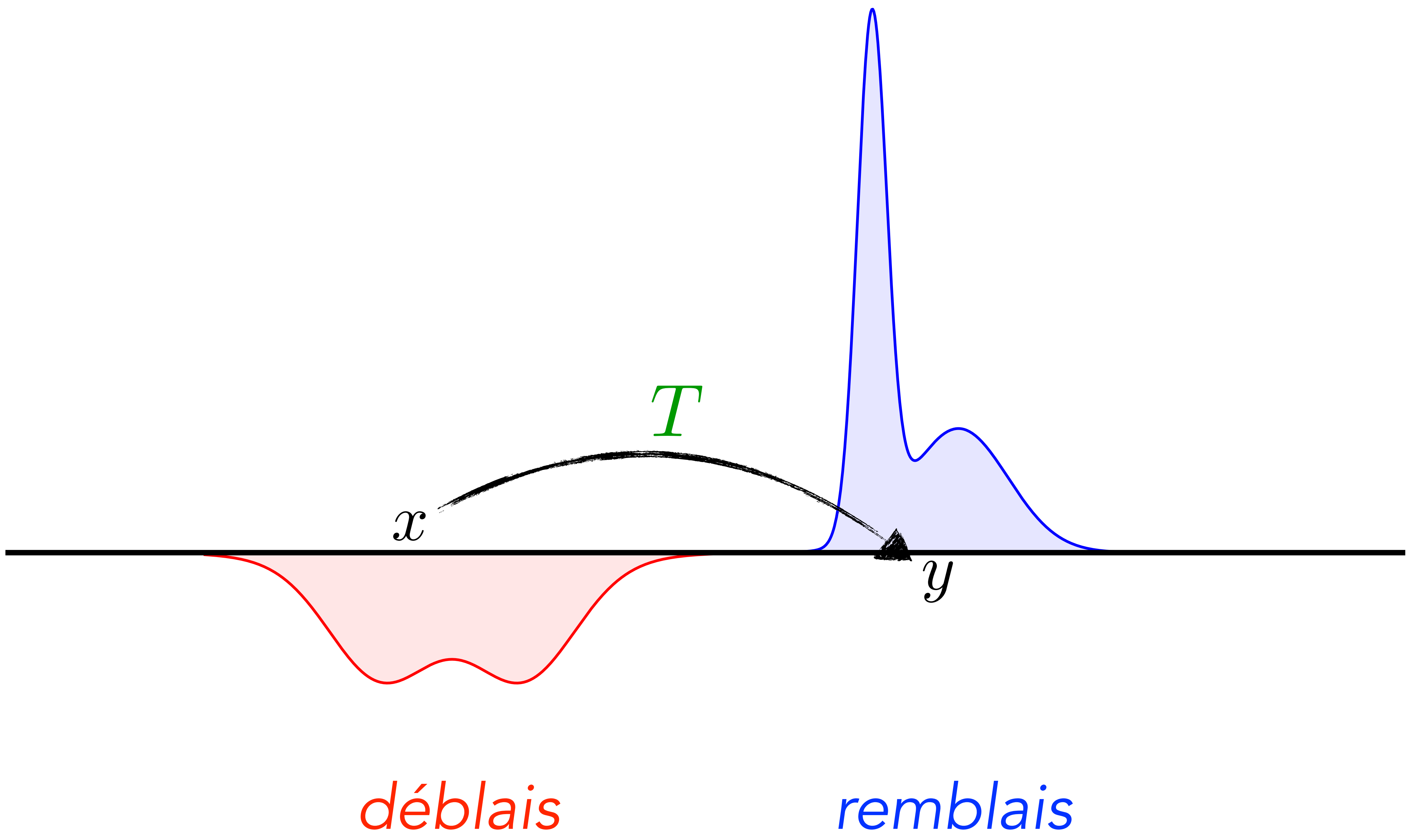
The Monge Problem



The Monge Problem

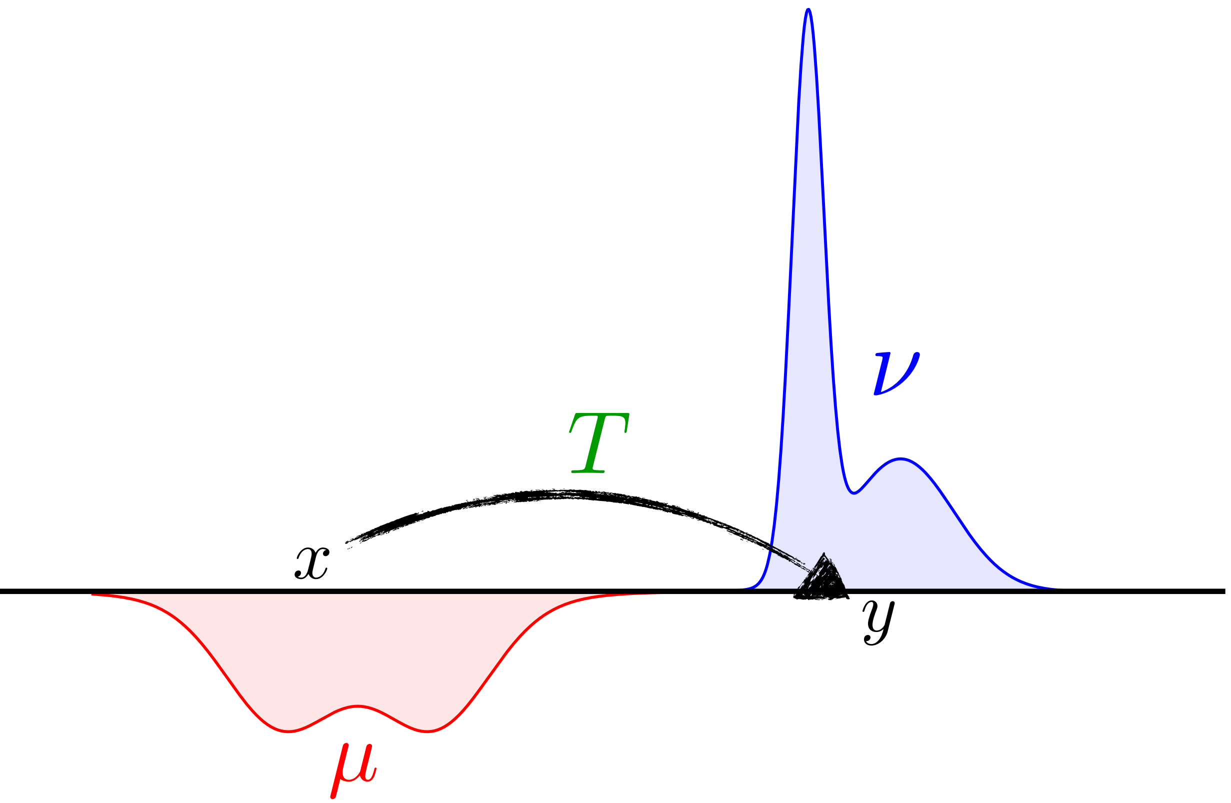


The Monge Problem



How to move the *déblais* to build the *remblais* with minimal effort?

The Monge Problem

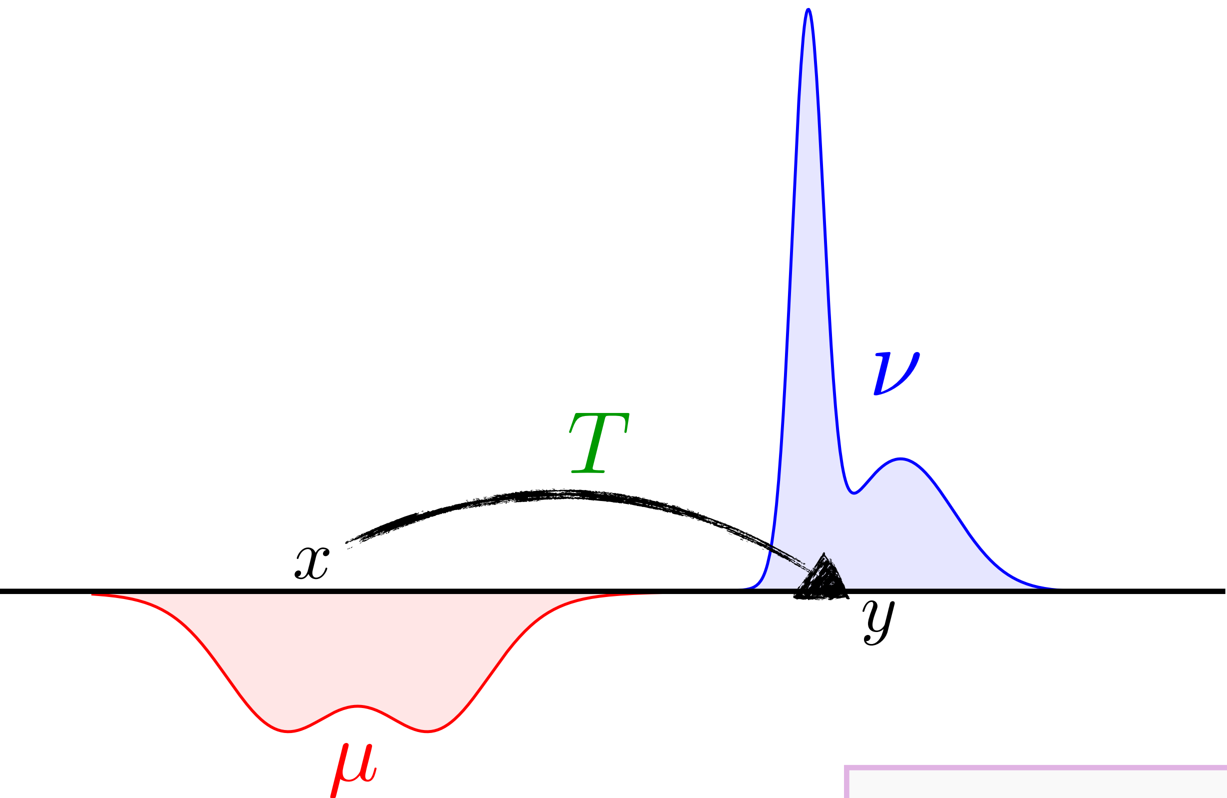


. Two distributions μ and ν over $\mathbb{R}^d$

. A cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

The Monge Problem



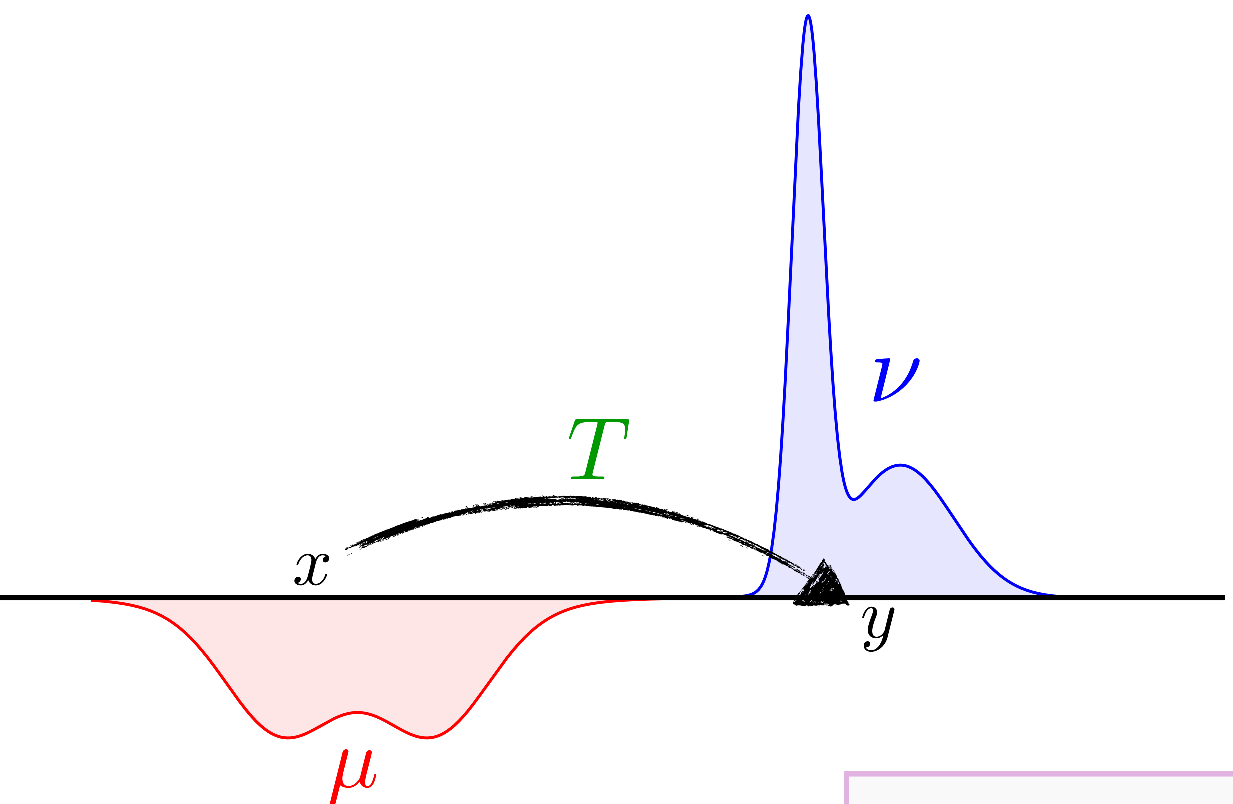
. Two distributions μ and ν over $\mathbb{R}^d$

. A cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

$$\inf_{T_\sharp \mu = \nu} \int c\,(x, T(x))\, d\mu(x)$$

The Monge Problem



. Two distributions μ and ν over $\mathbb{R}^d$

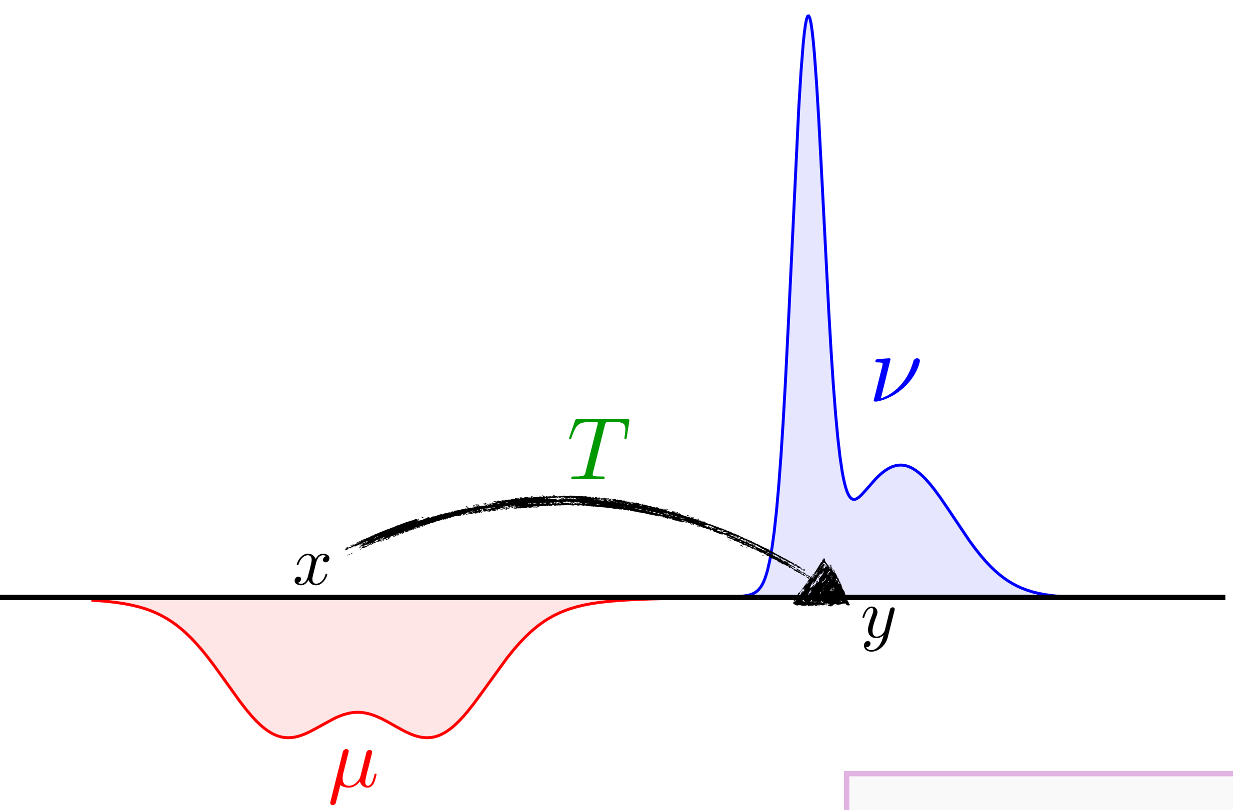
. A cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

$$\inf_{T_\sharp \mu = \nu} \int c\,(x, T(x))\, d\mu(x)$$

$$\forall A \subset \mathbb{R}^d \text{ Borel, } \nu(A) = \mu\left(T^{-1}(A)\right)$$

The Monge Problem



. Two distributions μ and ν over $\mathbb{R}^d$

. A cost function

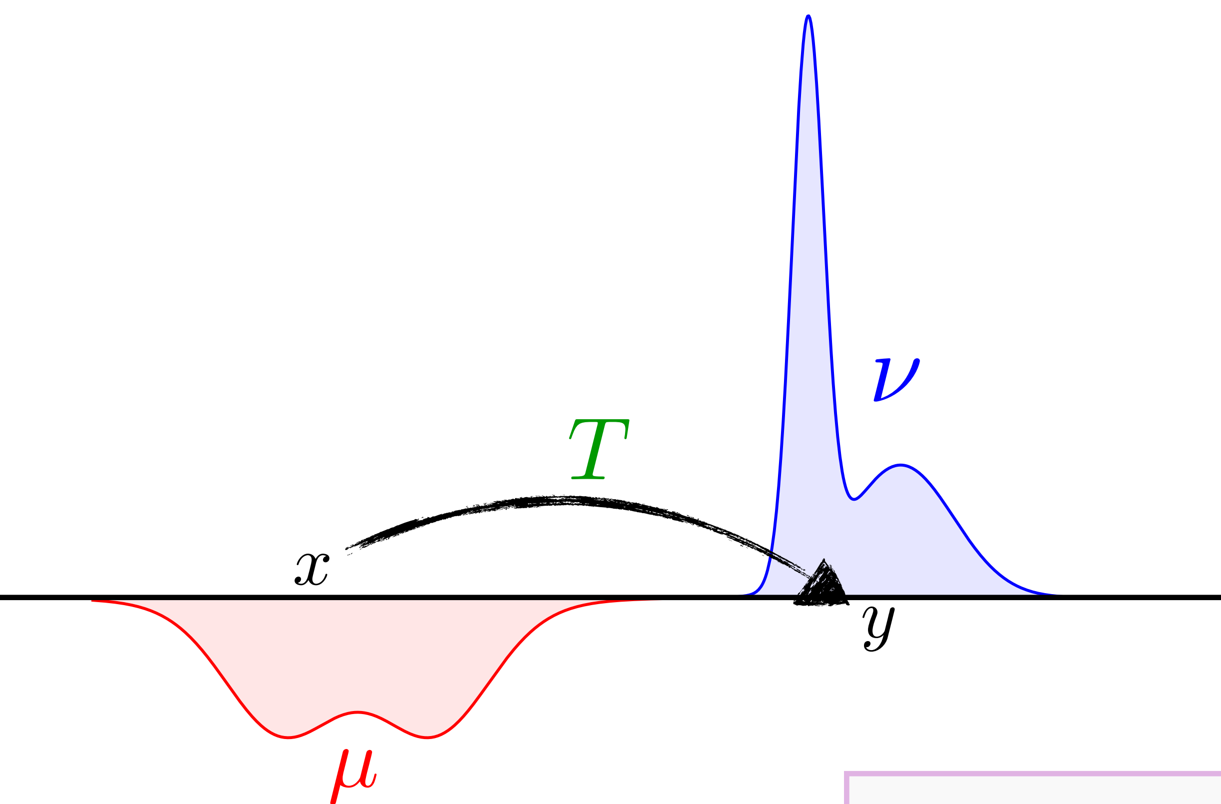
$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

$$\inf_{T_\sharp \mu = \nu} \int c\left(x, T(x)\right) d\mu(x)$$

$X \sim \mu \Longrightarrow T(X) \sim \nu$

$\forall A \subset \mathbb{R}^d$ Borel, $\nu(A) = \mu\left(T^{-1}(A)\right)$

The Monge Problem



. Two distributions μ and ν over $\mathbb{R}^d$

. A cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

$$\inf_{T_\sharp \mu = \nu} \int c\,(x, T(x))\, d\mu(x)$$

$X \sim \mu \Longrightarrow T(X) \sim \nu$

$\forall A \subset \mathbb{R}^d$ Borel, $\nu(A) = \mu\left(T^{-1}(A)\right)$

Issue: such maps T may not exist (*e.g.* send one Dirac mass to a sum of several Dirac masses)


THE KANTOROVICH PROBLEM
Leonid Kantorovich

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$c(x, y)\, d\pi(x, y)$$

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\iint c(x, y)\, d\pi(x, y)$$

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\inf_{\pi} \iint c(x, y)\, d\pi(x, y)$$

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y)\, d\pi(x, y)$$

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y)\, d\pi(x, y)$$

over all π *such that* $\begin{cases} \int d\pi(x, y) = d\mu(x) & \forall x \\ \int d\pi(x, y) = d\nu(y) & \forall y \end{cases}$ $\quad \pi \in \Pi(\mu, \nu)$

The Kantorovich Problem

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y)\, d\pi(x, y)$$

THE KANTOROVICH PROBLEM

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y)\, d\pi(x, y)$$

THE KANTOROVICH PROBLEM

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y) \, d\pi(x, y)$$

When the cost function is of the form

$$c(x, y) = ||x - y||^p \quad \text{where} \quad p \geq 1$$

we say that $\mathscr{T}_c^{1/p}$ *is the p-Wasserstein distance* W_p

THE KANTOROVICH PROBLEM

Existence may not hold in the Monge problem because each point has to be sent to a *unique* destination

💡 Relax this constraint and allow *mass splitting*

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y) \, d\pi(x, y)$$

When the cost function is of the form

$$c(x, y) = \|x - y\|^p \quad \text{where} \quad p \geq 1$$

we say that $\mathscr{T}_c^{1/p}$ *is the p-Wasserstein distance* W_p

Benefits: existence under mild assumptions

Some Properties on the Kantorovich Problem

Duality

$$\mathscr{T}_c(\mu, \nu) = \sup_{\substack{\phi, \psi \\ \phi \oplus \psi \leq c}} \int \phi \, d\mu + \int \psi \, d\nu$$

Some Properties on the Kantorovich Problem

Duality

$$\mathscr{T}_c(\mu, \nu) = \sup_{\substack{\phi,\psi \\ \phi \oplus \psi \leq c}} \int \phi \, d\mu + \int \psi \, d\nu$$

Metric Properties

W_p is a geodesic distance over the set $\mathscr{P}_p(\mathbb{R}^d)$ of probability measures with finite p^{th} moment

Some Properties on the Kantorovich Problem

Duality

$$\mathscr{T}_c(\mu, \nu) = \sup_{\substack{\phi, \psi \\ \phi \oplus \psi \leq c}} \int \phi \, d\mu + \int \psi \, d\nu$$

Metric Properties

W_p is a geodesic distance over the set $\mathscr{P}_p(\mathbb{R}^d)$ of probability measures with finite p^{th} moment

Issues: both algorithmic and statistical limitations in Machine Learning

Limitation to the Kantorovich Problem

1. Algorithmic limitations

. The discrete problem is a Linear Program in $\mathcal{O}\left(n^3 \log n\right)$

. Lack of differentiability

Limitation to the Kantorovich Problem

1. Algorithmic limitations

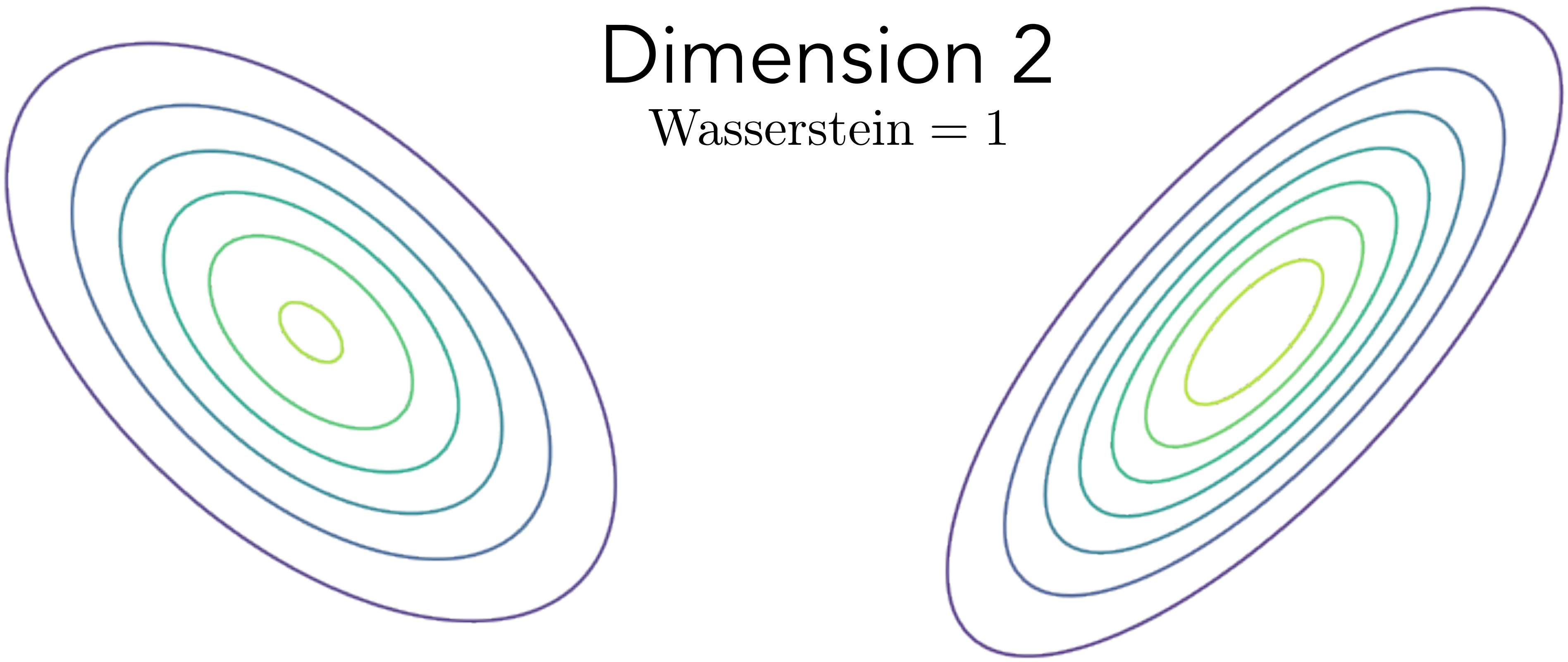
. The discrete problem is a Linear Program in $\mathcal{O}\left(n^3 \log n\right)$

. Lack of differentiability

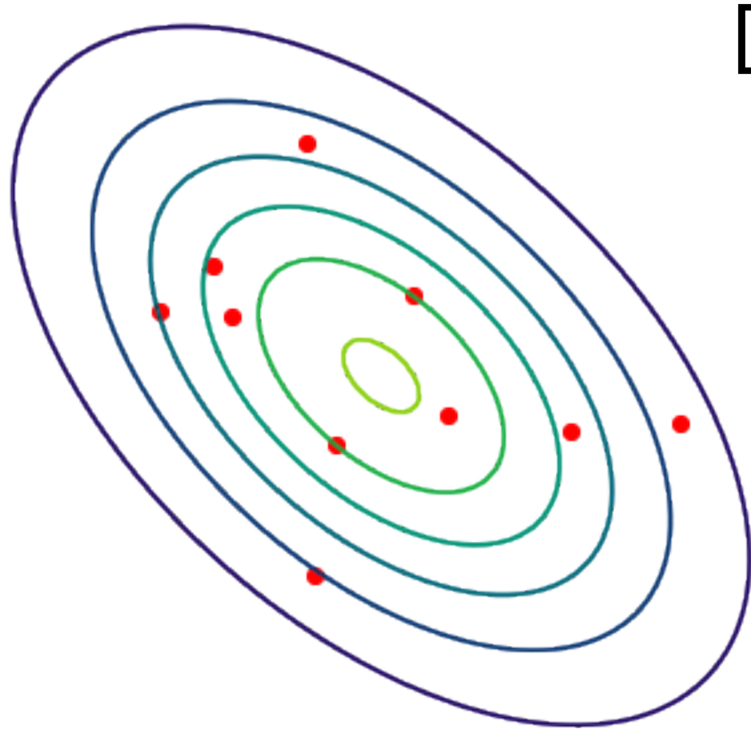
2. Statistical limitations

Wasserstein distances suffer from the **curse of dimensionality**

The Curse of Dimensionality

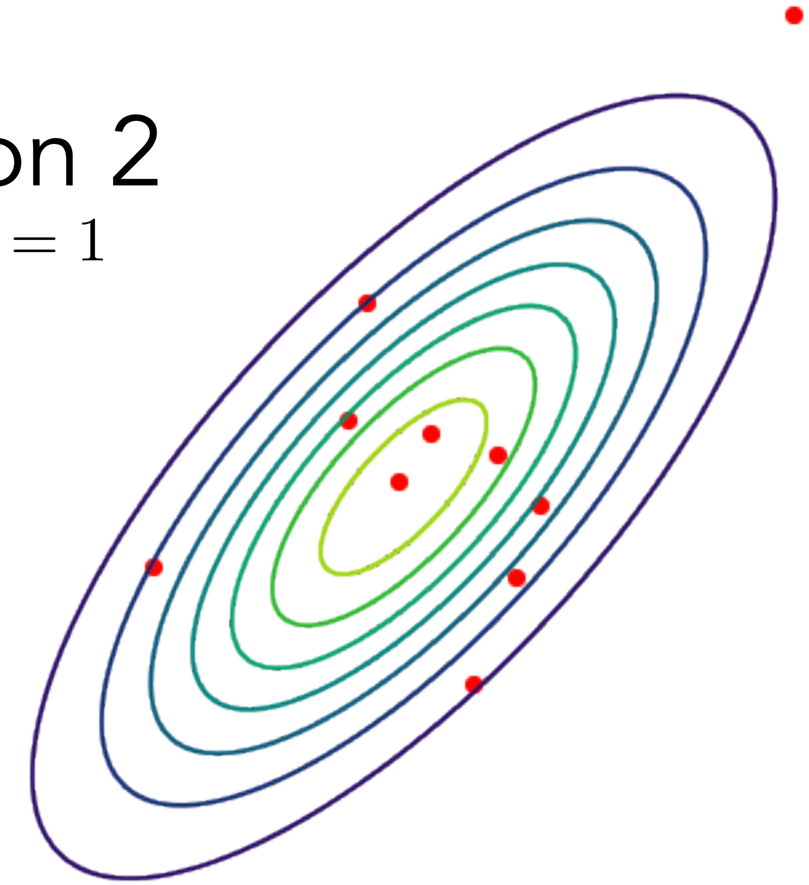


The Curse of Dimensionality



Dimension 2

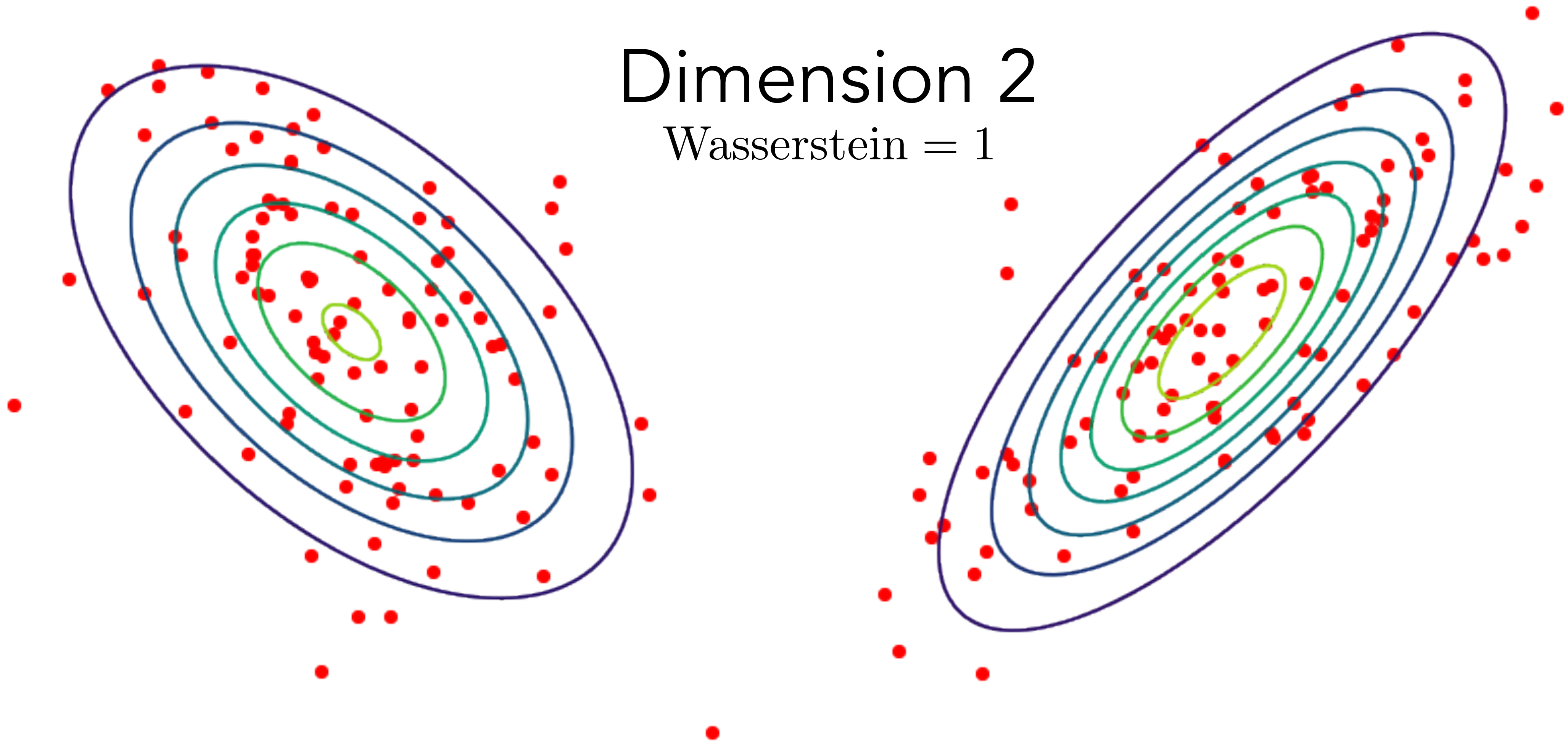
Wasserstein $= 1$



$n = 10$

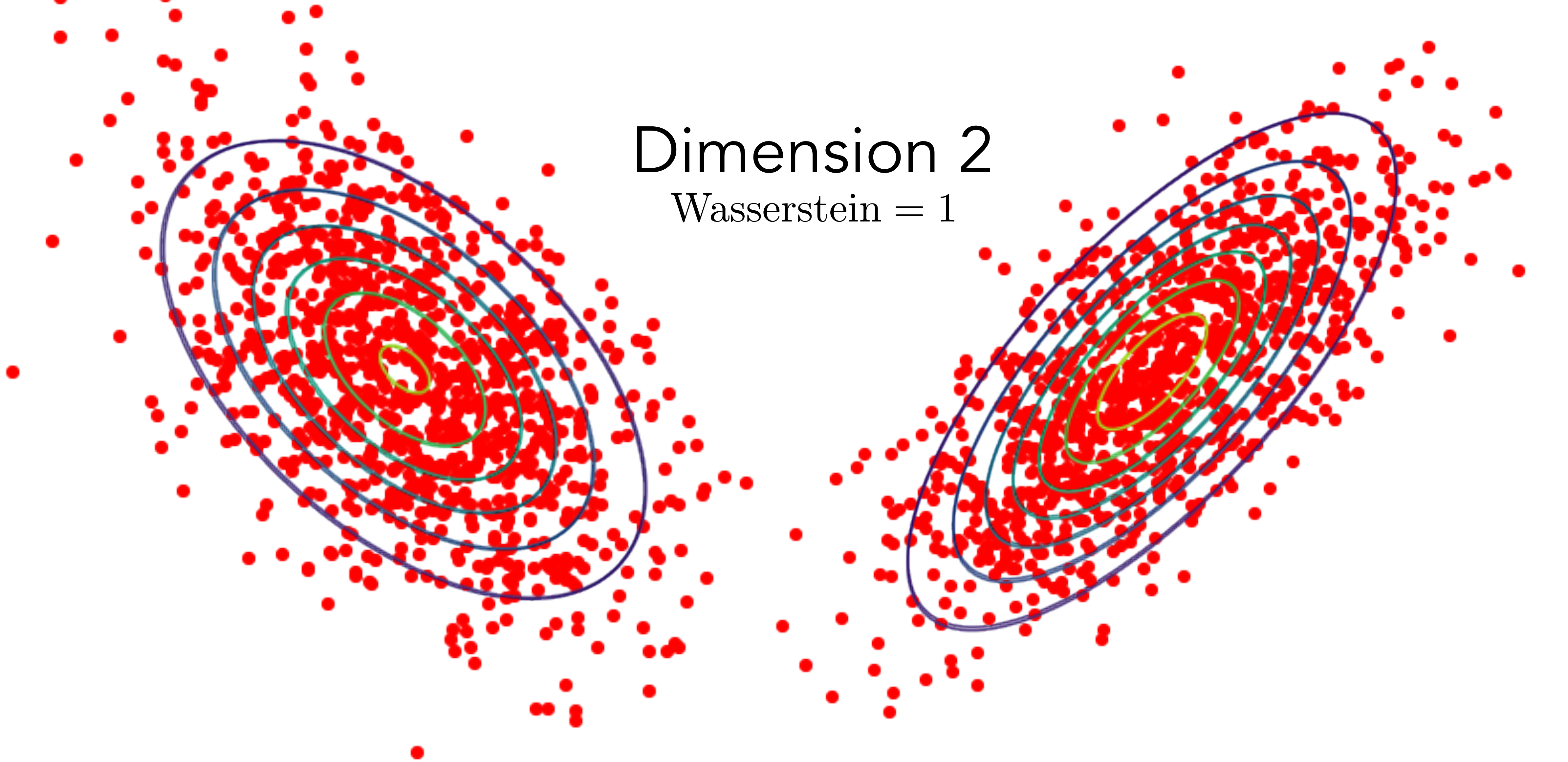
Estimation error $= 0.83$

THE CURSE OF DIMENSIONALITY



$n = 100$ $\quad$ Estimation error $= 0.15$

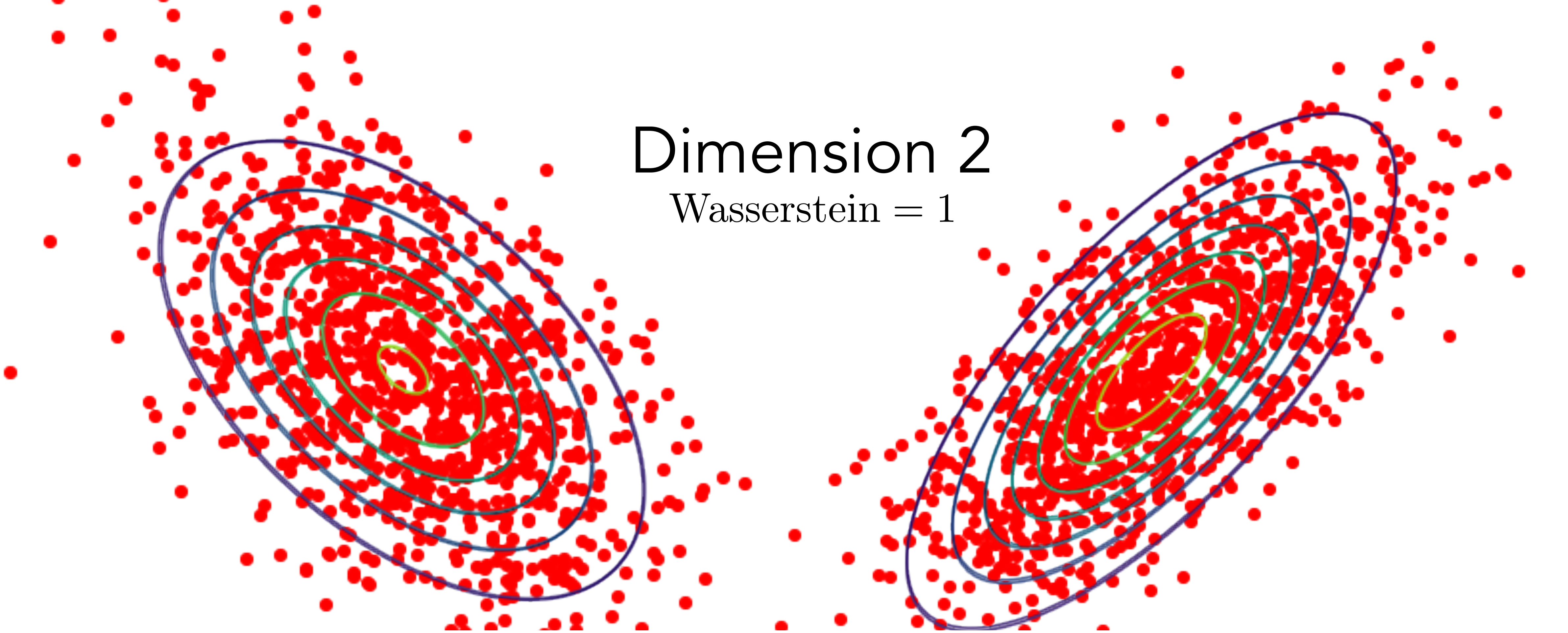
The Curse of Dimensionality

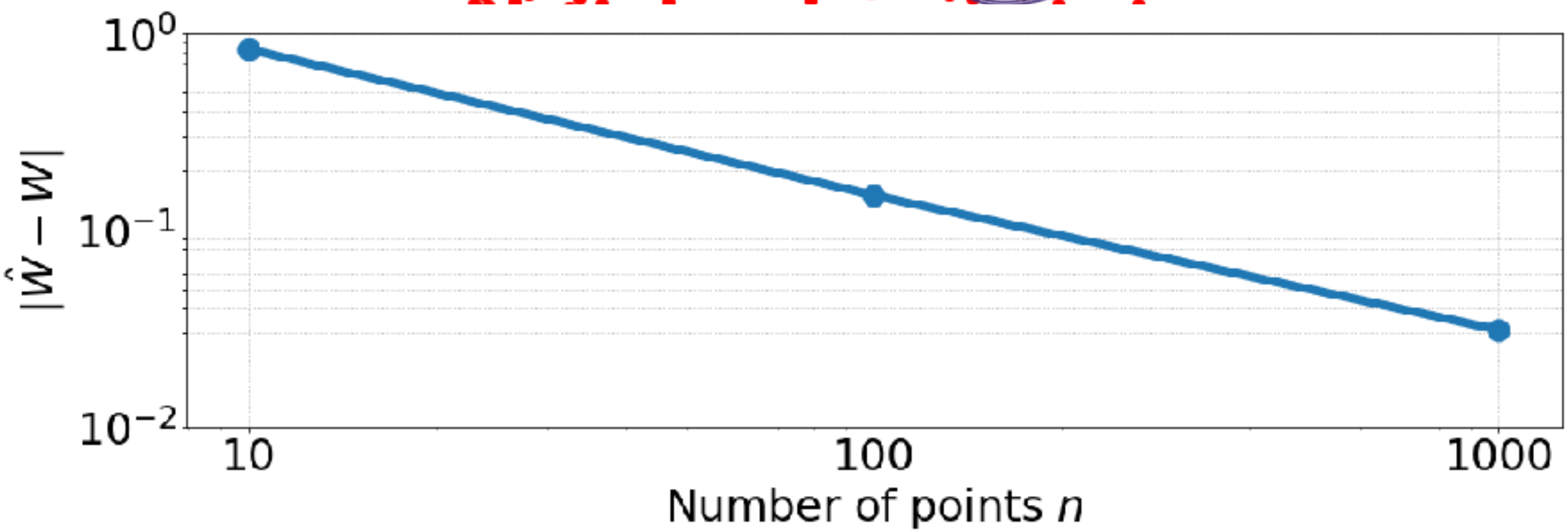


$n = 1000$

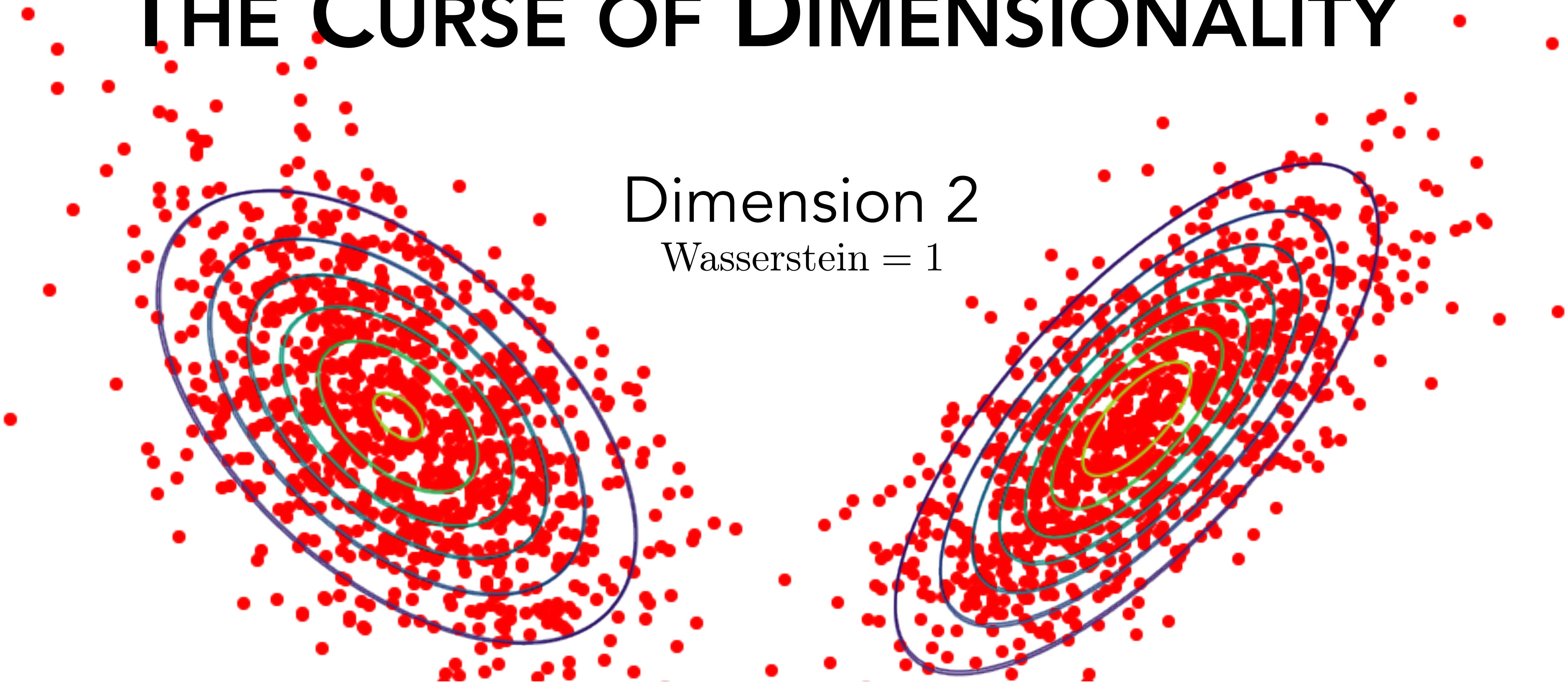
Estimation error = 0.03

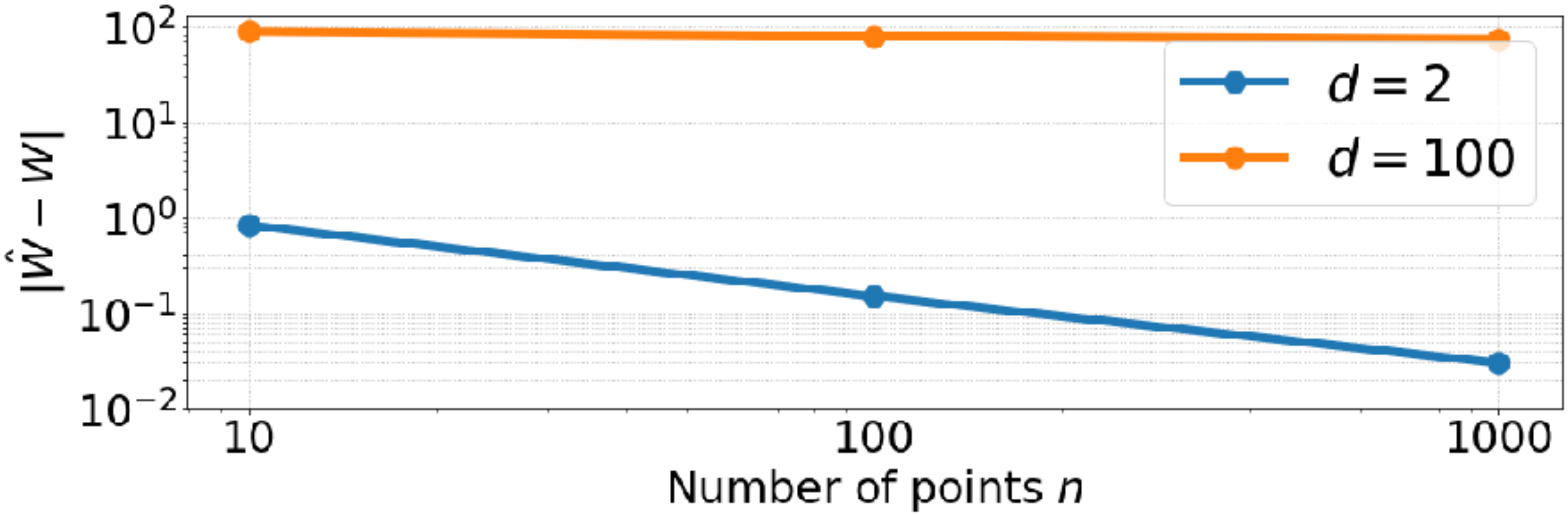
THE CURSE OF DIMENSIONALITY



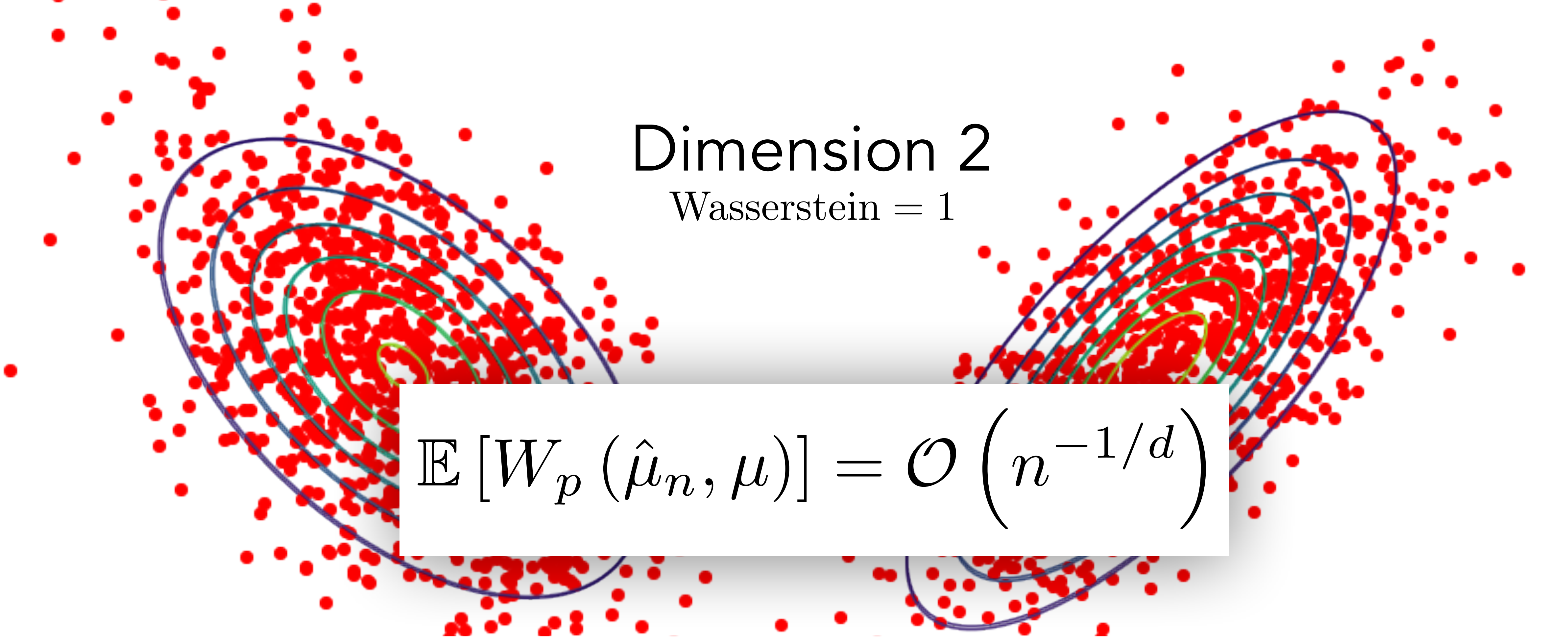


The Curse of Dimensionality

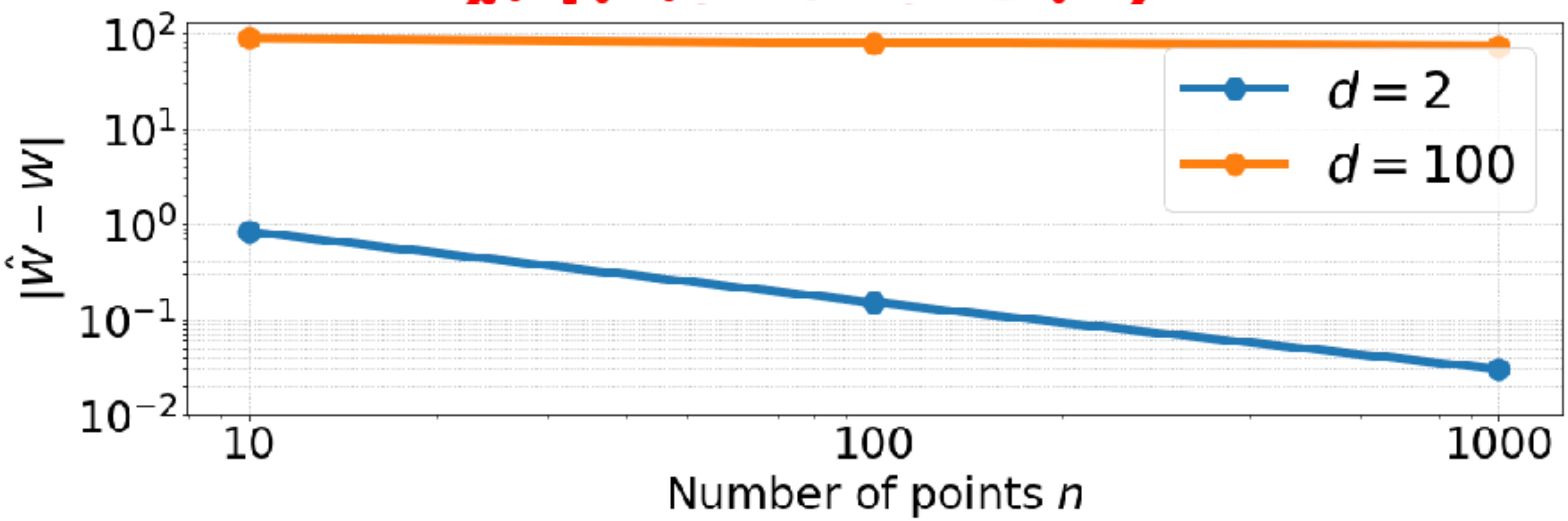




The Curse of Dimensionality

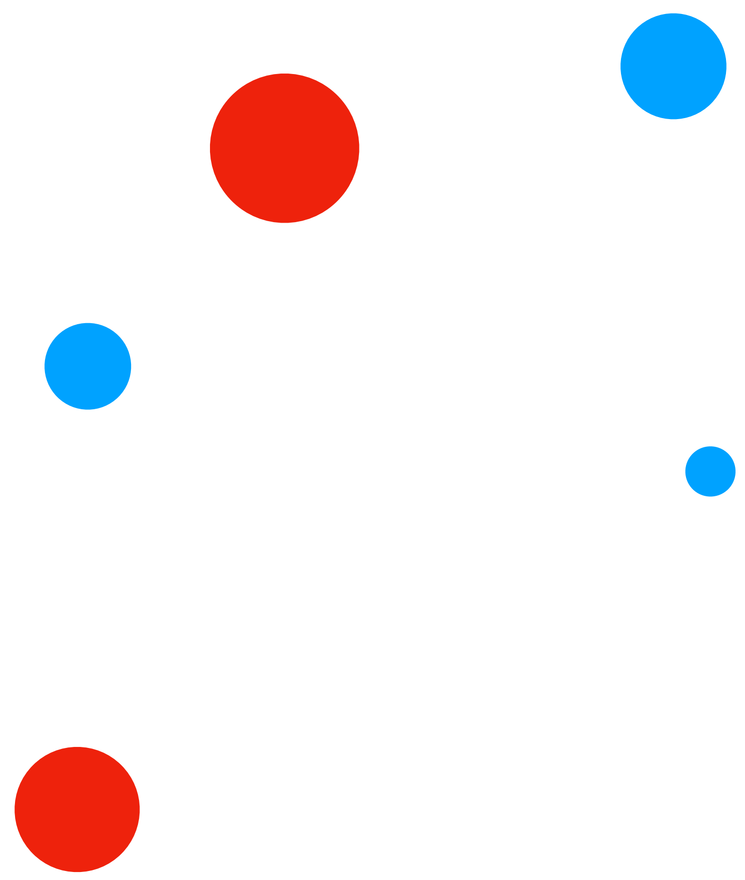


$$\mathbb{E}\left[W_p\left(\hat{\mu}_n, \mu\right)\right] = \mathcal{O}\left(n^{-1/d}\right)$$



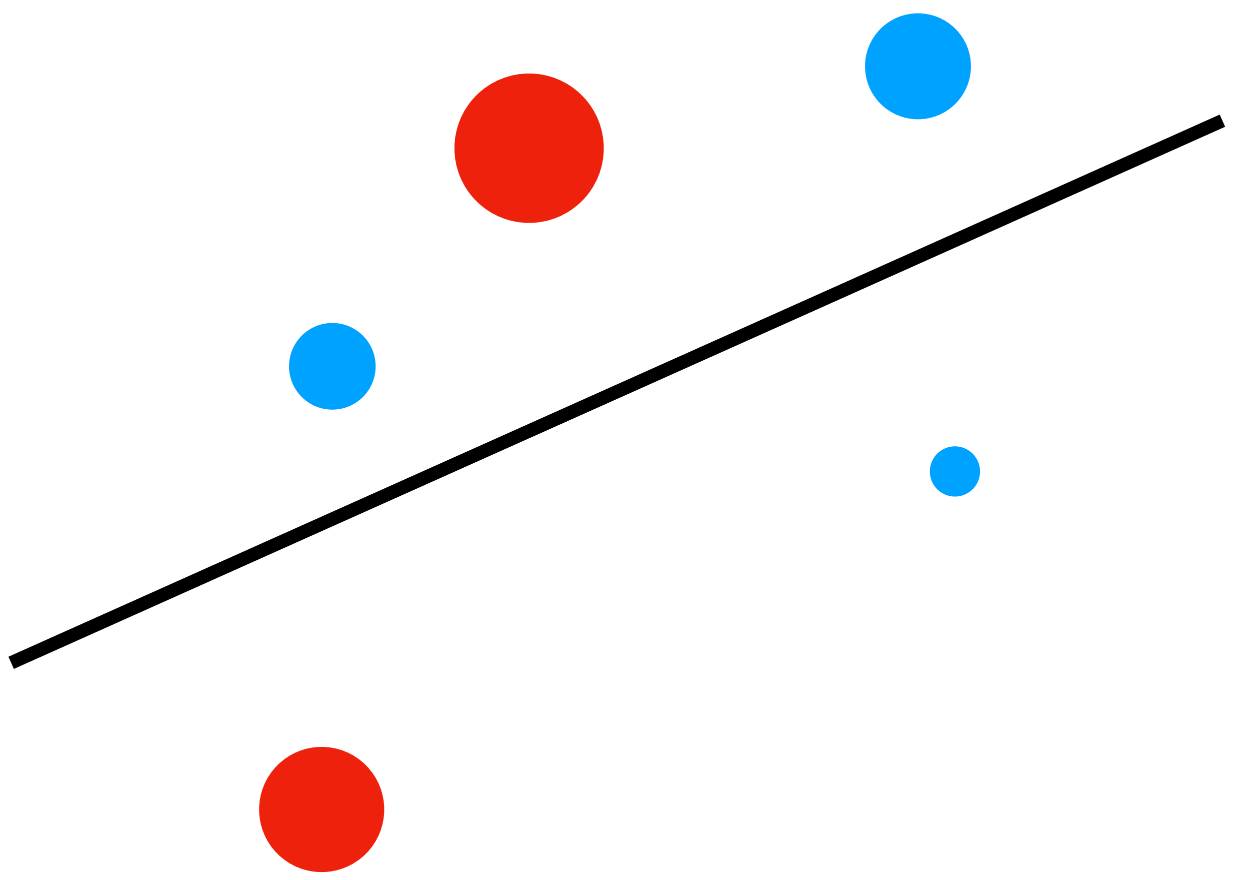
Beyond these Limitations

Sliced Approach



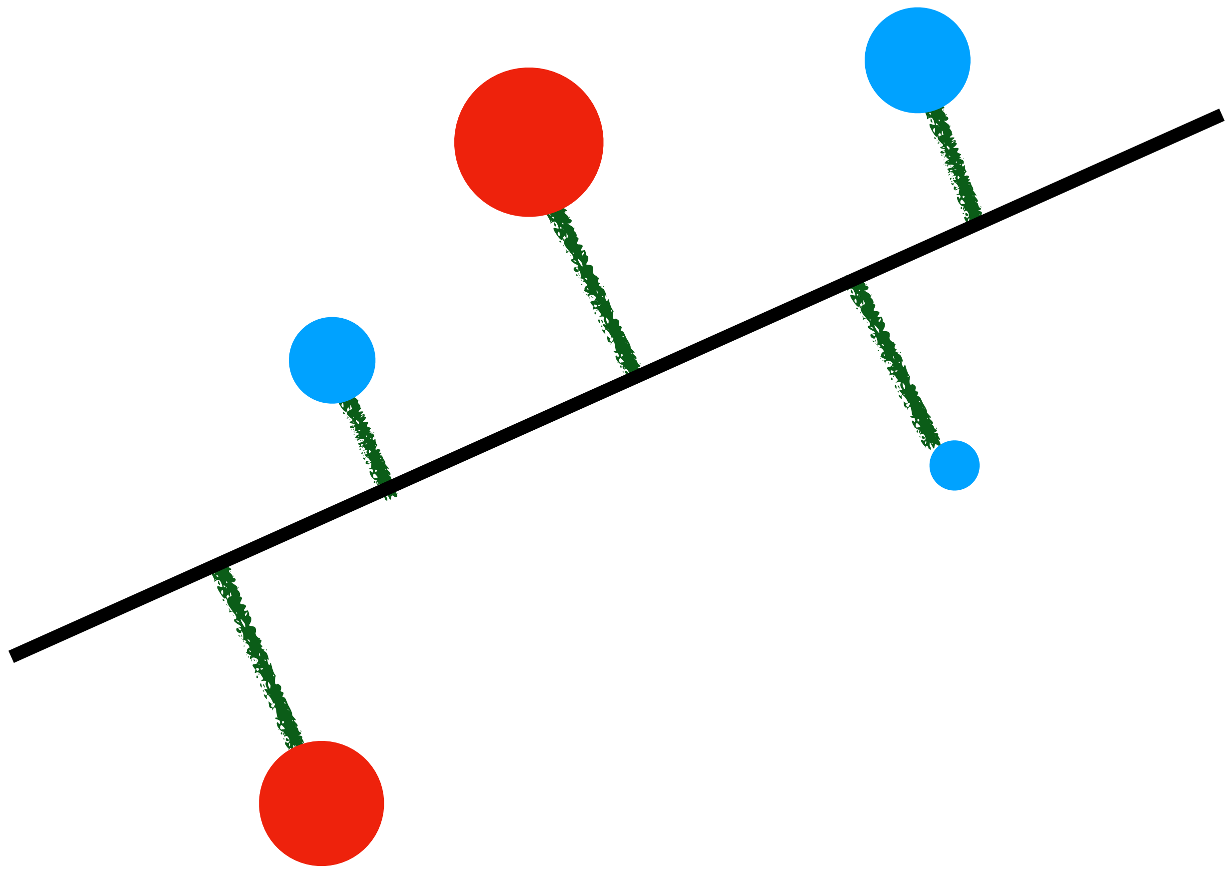
Beyond these Limitations

Sliced Approach



Beyond these Limitations

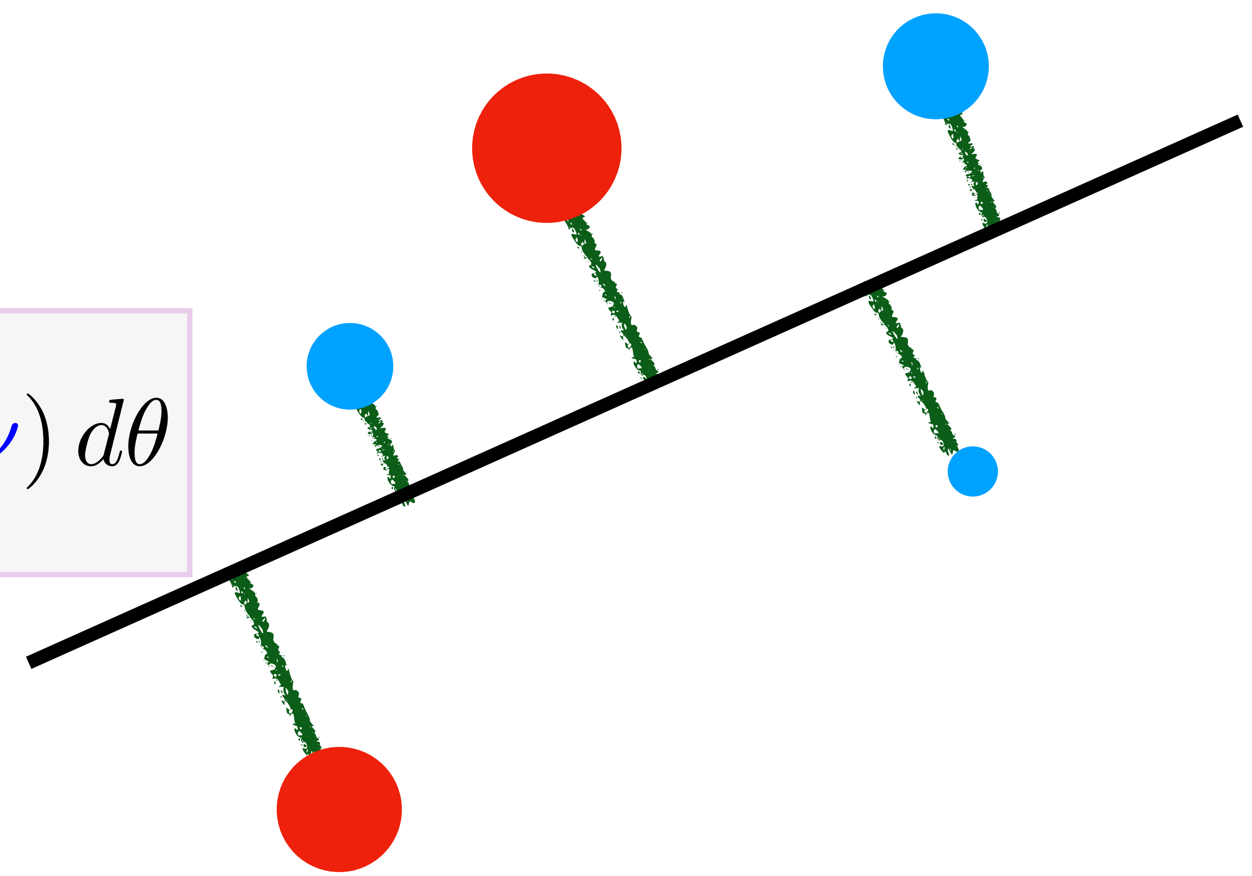
Sliced Approach



Beyond these Limitations

Sliced Approach

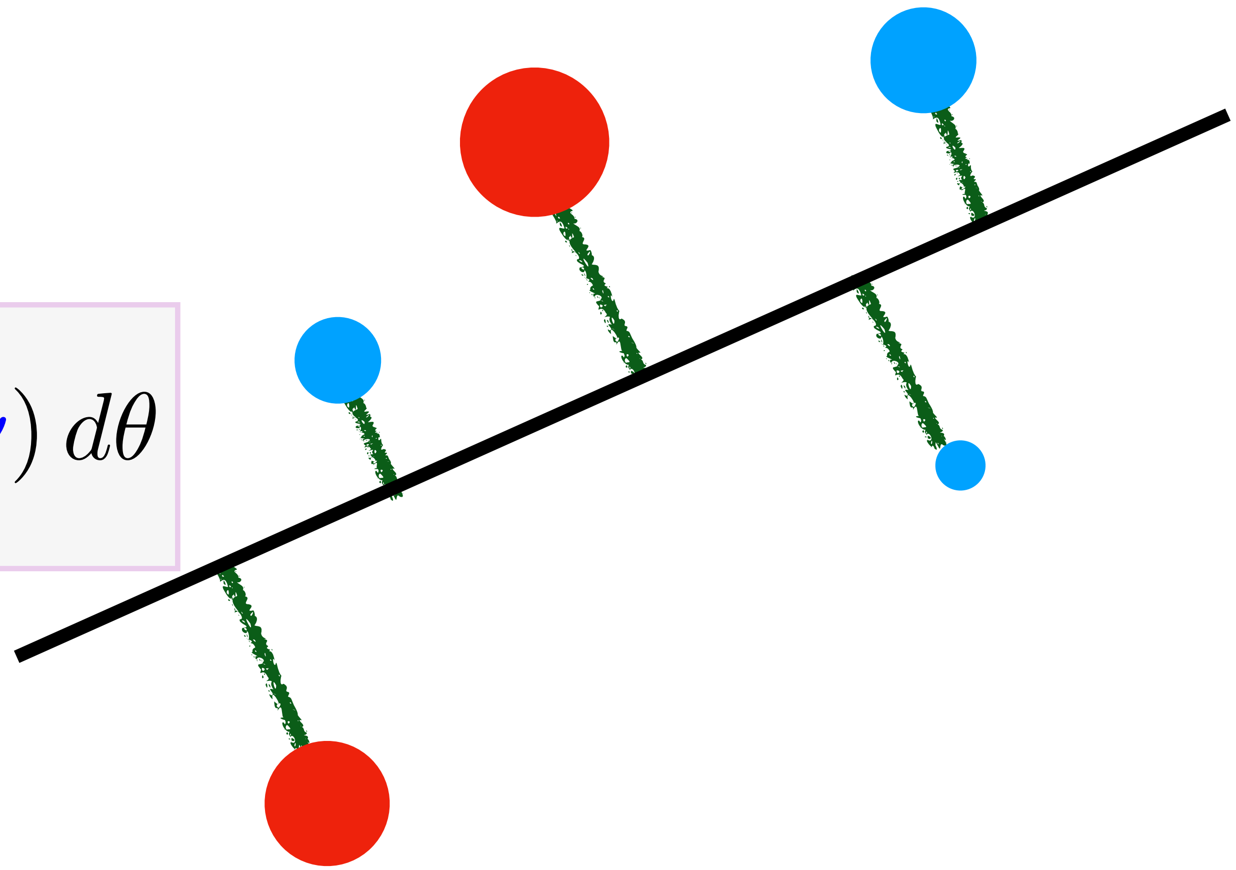
$$SW_c(\mu, \nu) = \int_{\mathbb{S}^{d-1}} \mathscr{T}_c(p_{\theta\sharp}\mu, p_{\theta\sharp}\nu)\, d\theta$$



Beyond these Limitations

Sliced Approach

$$SW_c(\mu, \nu) = \int_{\mathbb{S}^{d-1}} \mathscr{T}_c(p_{\theta\sharp}\mu, p_{\theta\sharp}\nu)\, d\theta$$

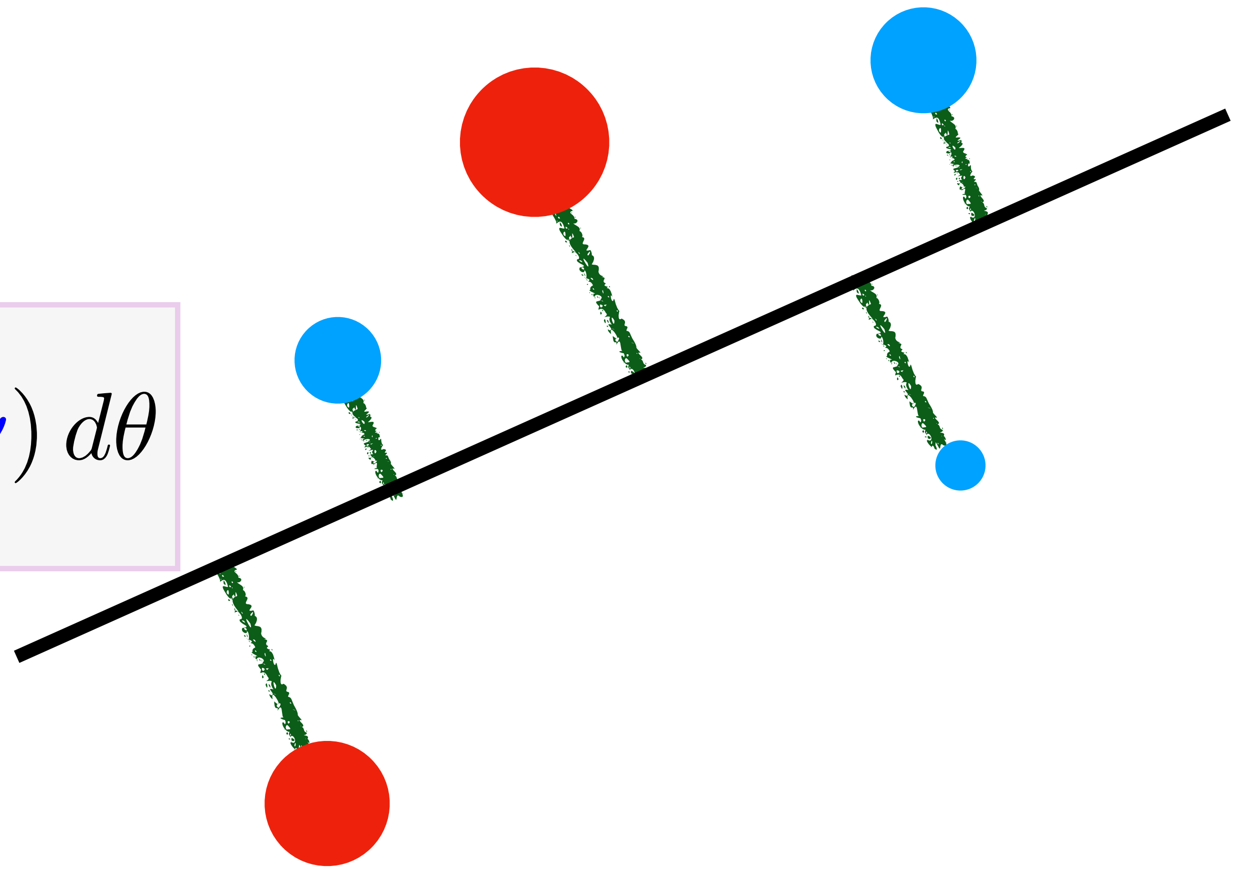


Entropic regularization

Beyond these Limitations

Sliced Approach

$$SW_c(\mu, \nu) = \int_{\mathbb{S}^{d-1}} \mathscr{T}_c(p_{\theta\sharp}\mu, p_{\theta\sharp}\nu)\, d\theta$$



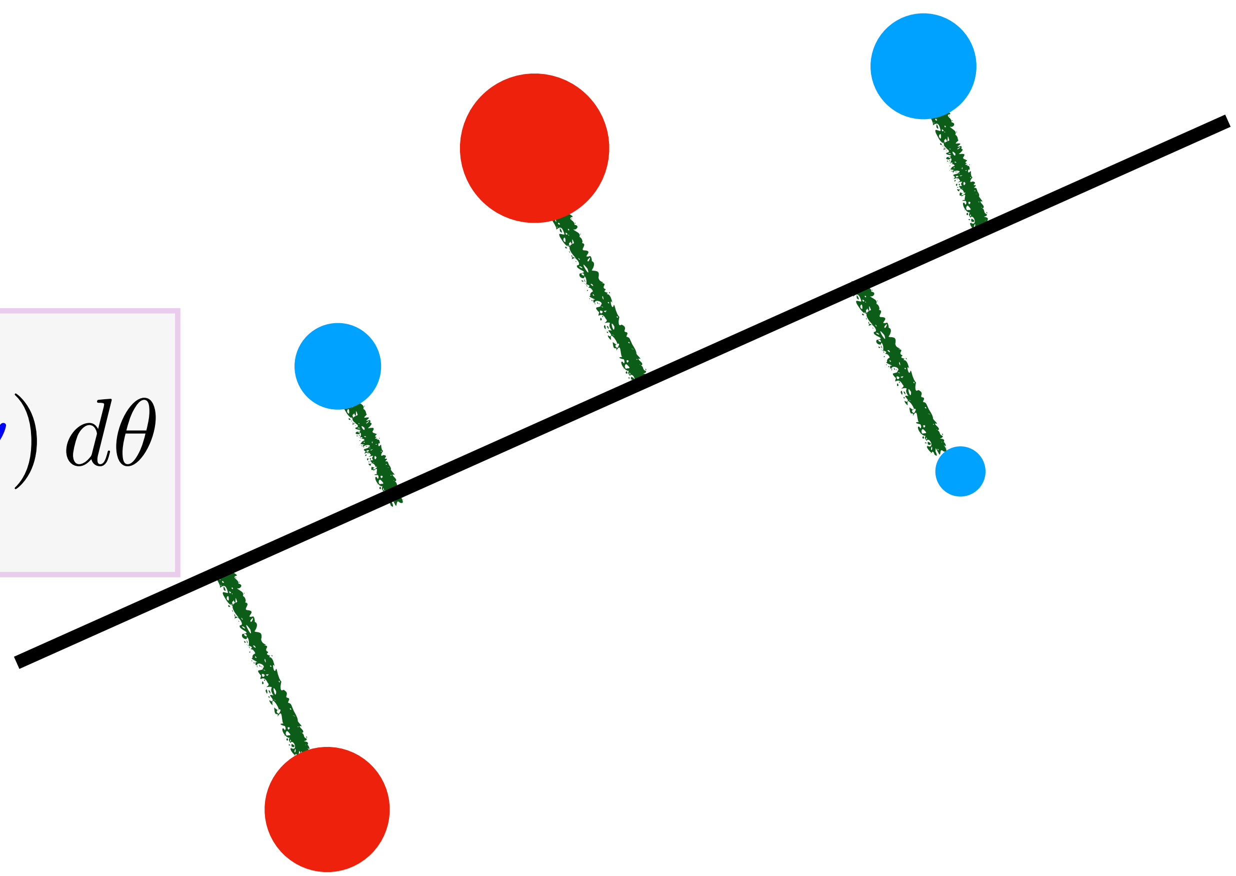
Entropic regularization

$$\inf_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi$$

Beyond these Limitations

Sliced Approach

$$SW_c(\mu,\nu) = \int_{\mathbb{S}^{d-1}} \mathscr{T}_c(p_{\theta\sharp}\mu, p_{\theta\sharp}\nu)\, d\theta$$



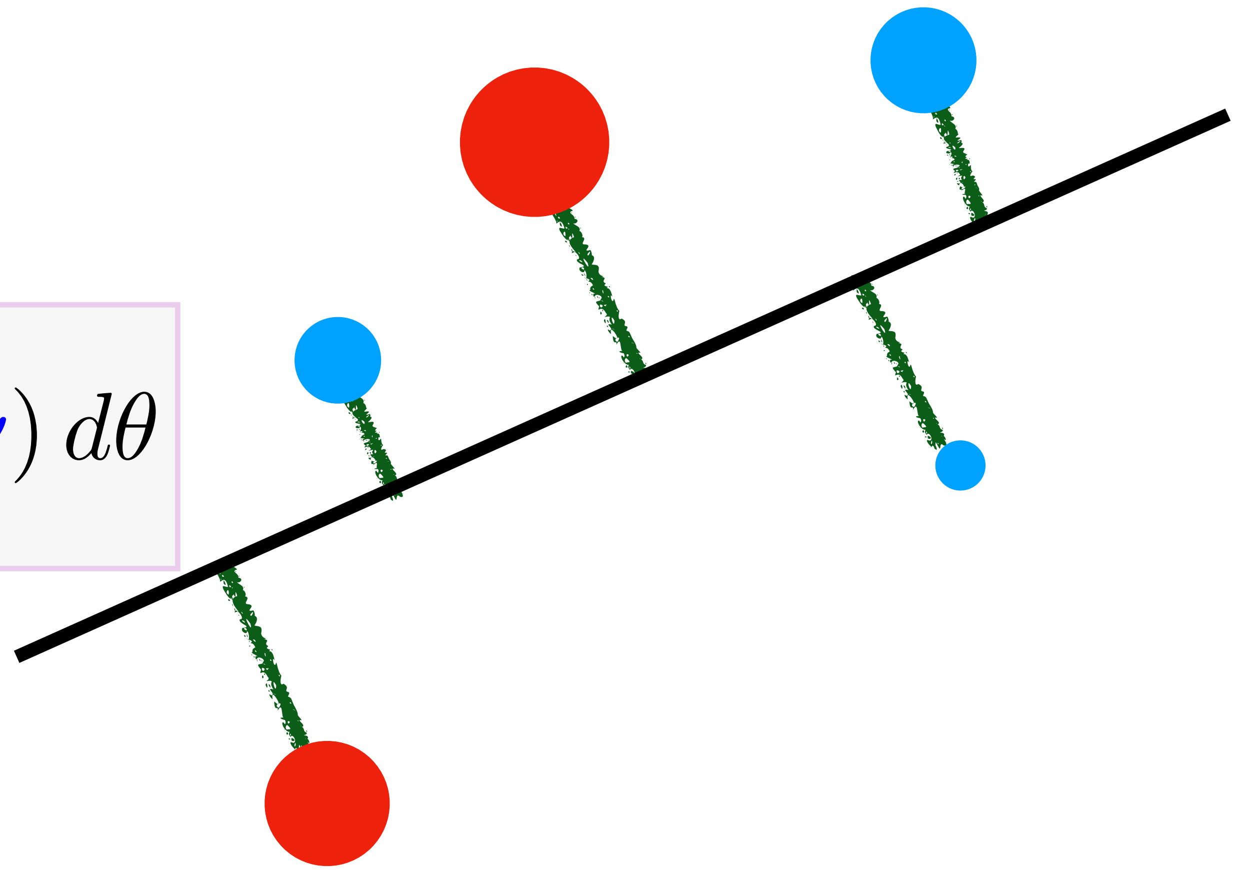
Entropic regularization

$$\inf_{\pi\in\Pi(\mu,\nu)} \int c\, d\pi + \gamma \mathrm{KL}(\pi || \mu \otimes \nu)$$

Beyond these Limitations

Sliced Approach

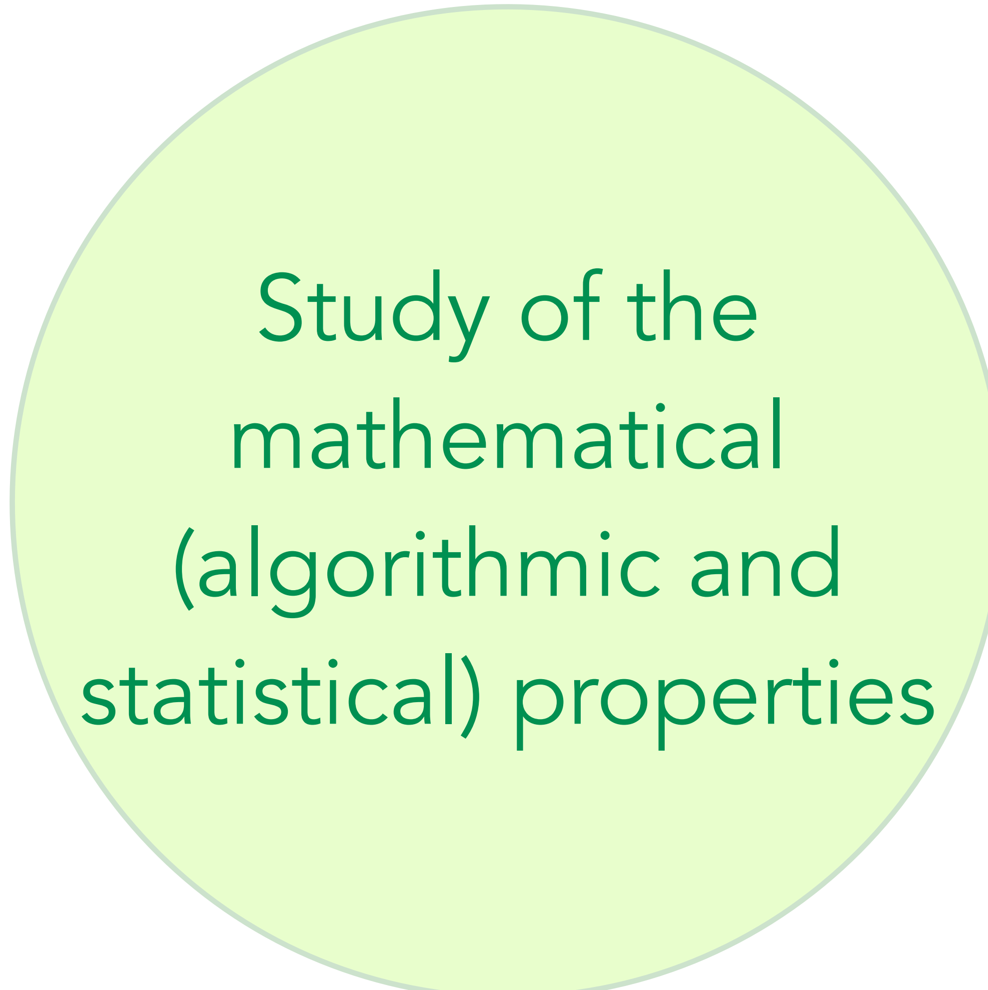
$$SW_c(\mu, \nu) = \int_{\mathbb{S}^{d-1}} \mathscr{T}_c(p_{\theta\sharp}\mu, p_{\theta\sharp}\nu)\, d\theta$$



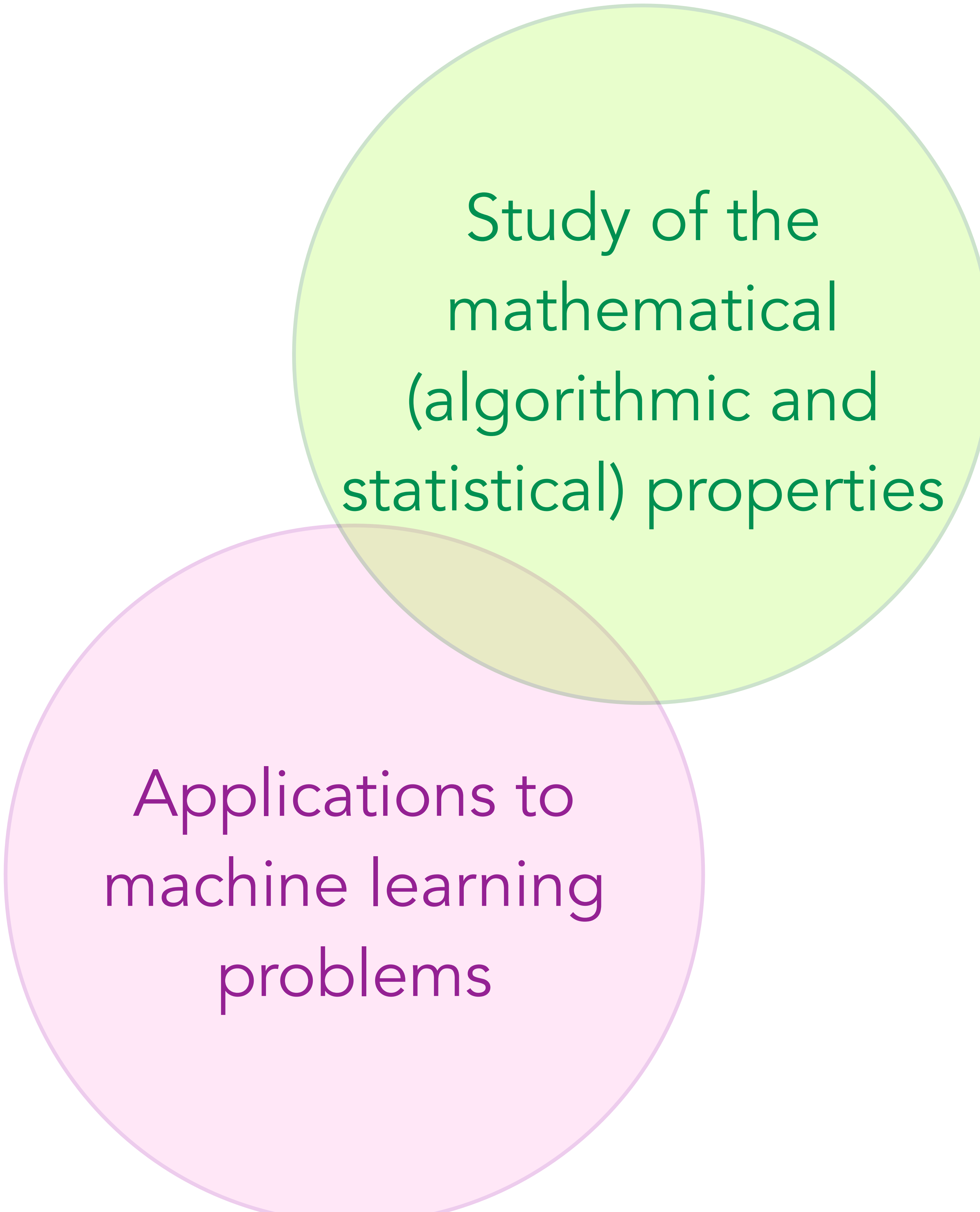
Entropic regularization

$$\mathscr{S}_c^{\gamma}(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int c\, d\pi + \gamma \mathrm{KL}(\pi || \mu \otimes \nu)$$

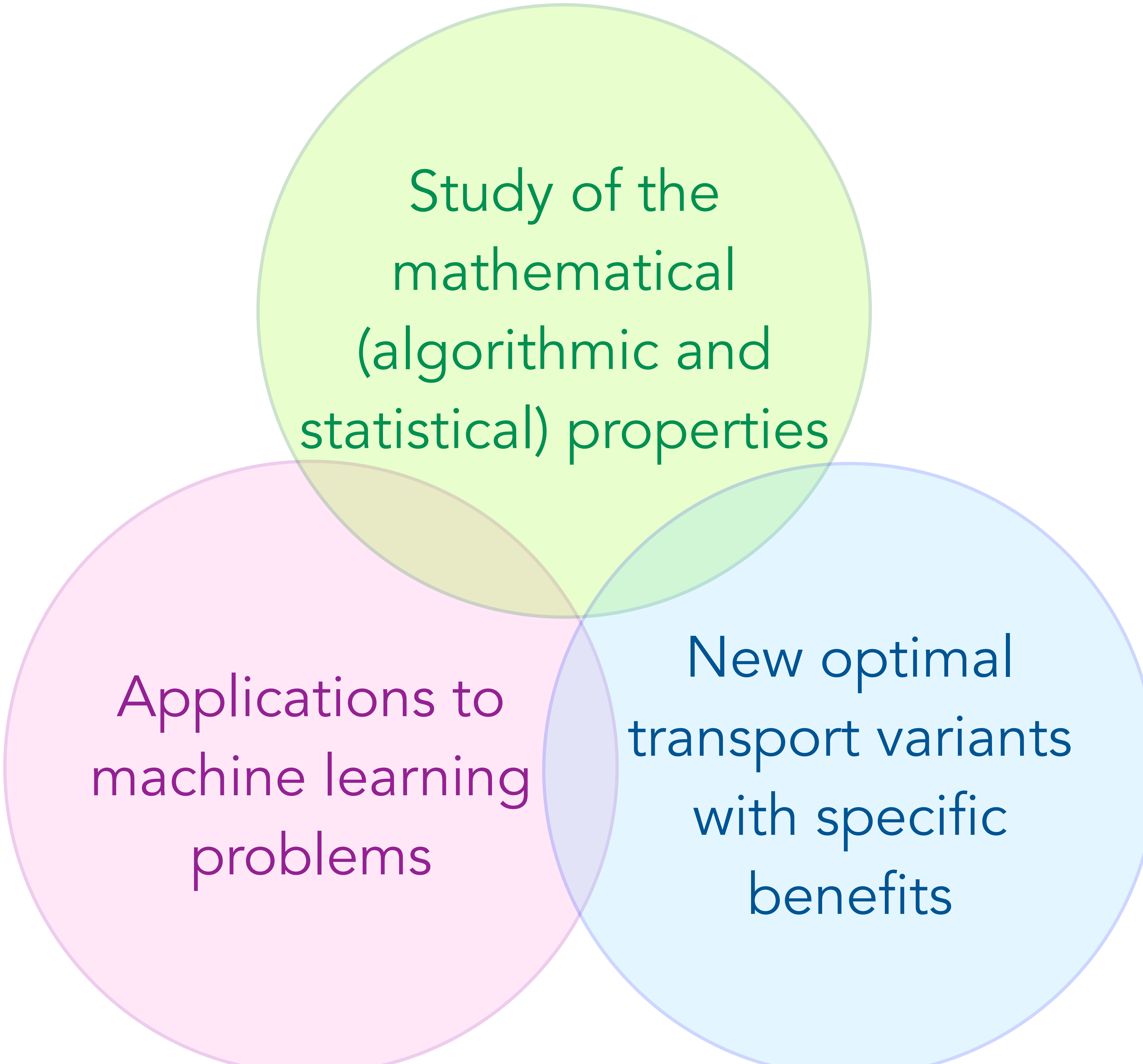
A SCHEMATIC VIEW OF THE COMMUNITY



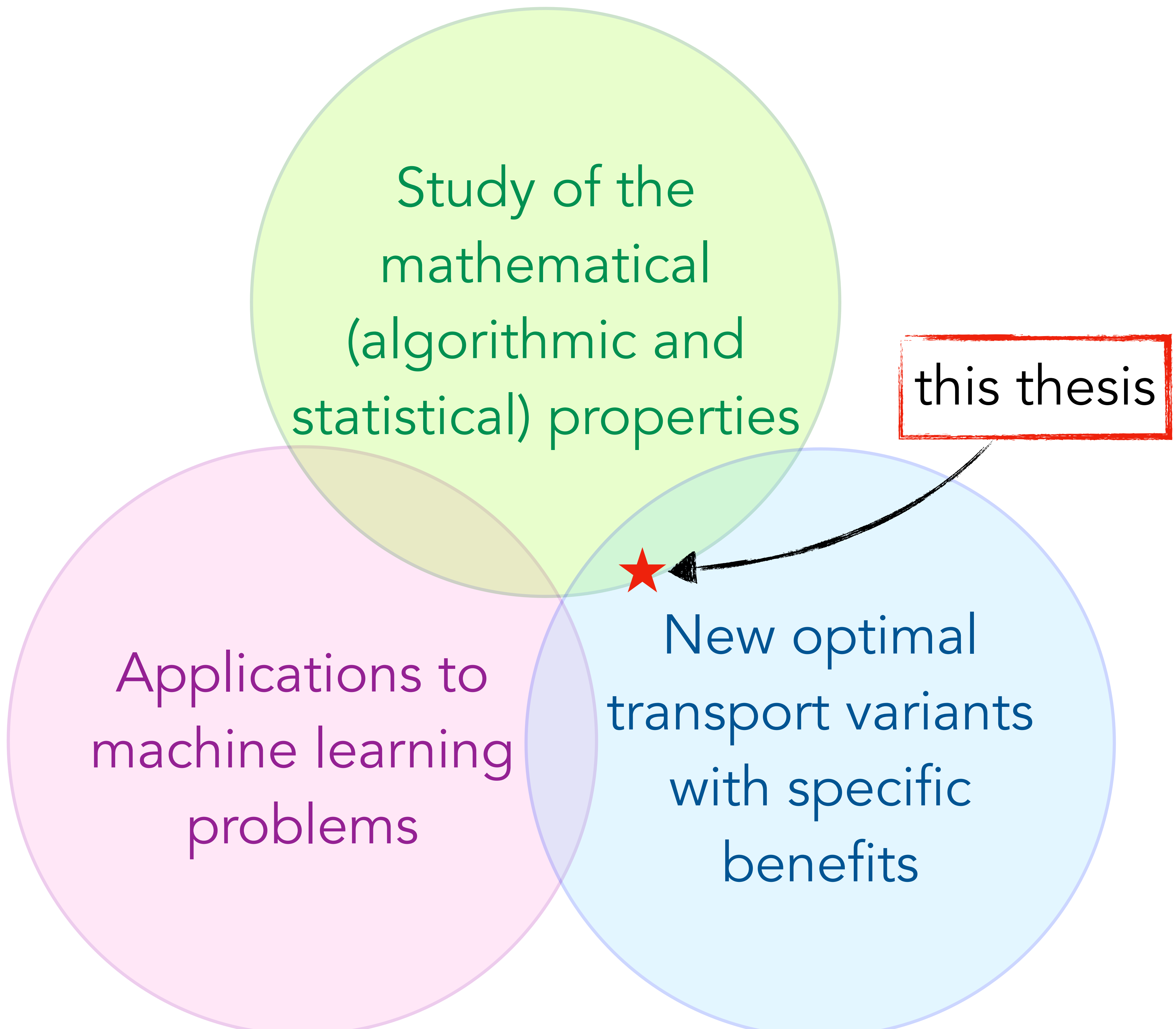
A SCHEMATIC VIEW OF THE COMMUNITY



A SCHEMATIC VIEW OF THE COMMUNITY



A SCHEMATIC VIEW OF THE COMMUNITY



Main Contributions

Two novel variants of optimal transport based on projections and convexity

MAIN CONTRIBUTIONS

Two novel variants of optimal transport based on projections and convexity

→ Part I and II

Main Contributions

Two novel variants of optimal transport based on projections and convexity

→ Part I and II

Geometric reinterpretation of regularized optimal transport

Main Contributions

Two novel variants of optimal transport based on projections and convexity

→ Part I and II

Geometric reinterpretation of regularized optimal transport

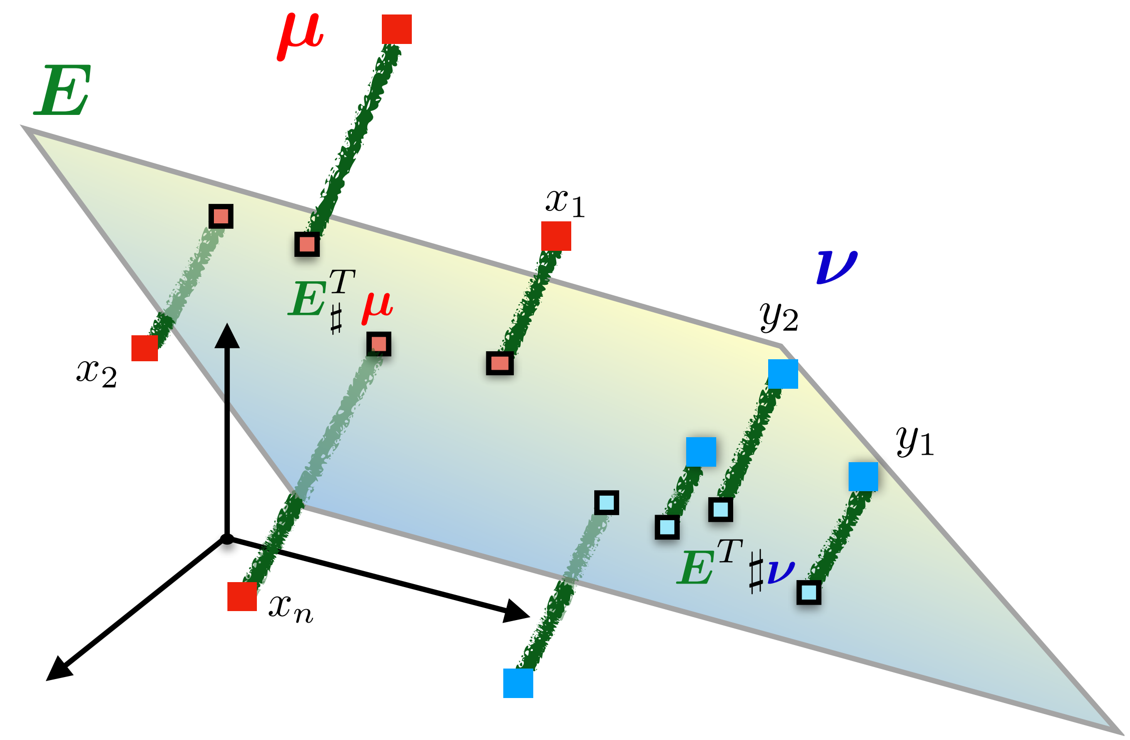
→ Part I

Part I: Ground-cost Robustness

Cédric Villani

Subspace Robust Wasserstein Distances

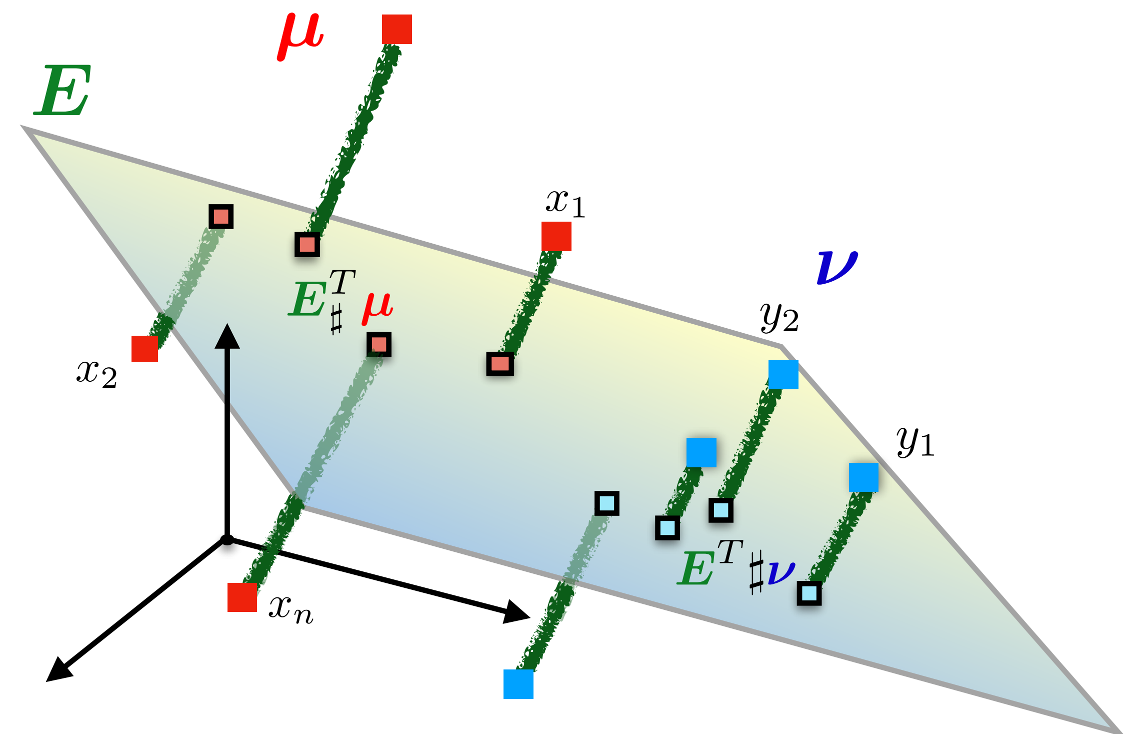
Idea: projecting measures on to a low-dimensional subspace before computing the Wasserstein distance



$$\mathcal{P}_k(\mu, \nu) = \sup_{\dim(E)=k} \mathcal{W}\left(P_{E\#}\mu, P_{E\#}\nu\right)$$

Subspace Robust Wasserstein Distances

Idea: projecting measures on to a low-dimensional subspace before computing the Wasserstein distance

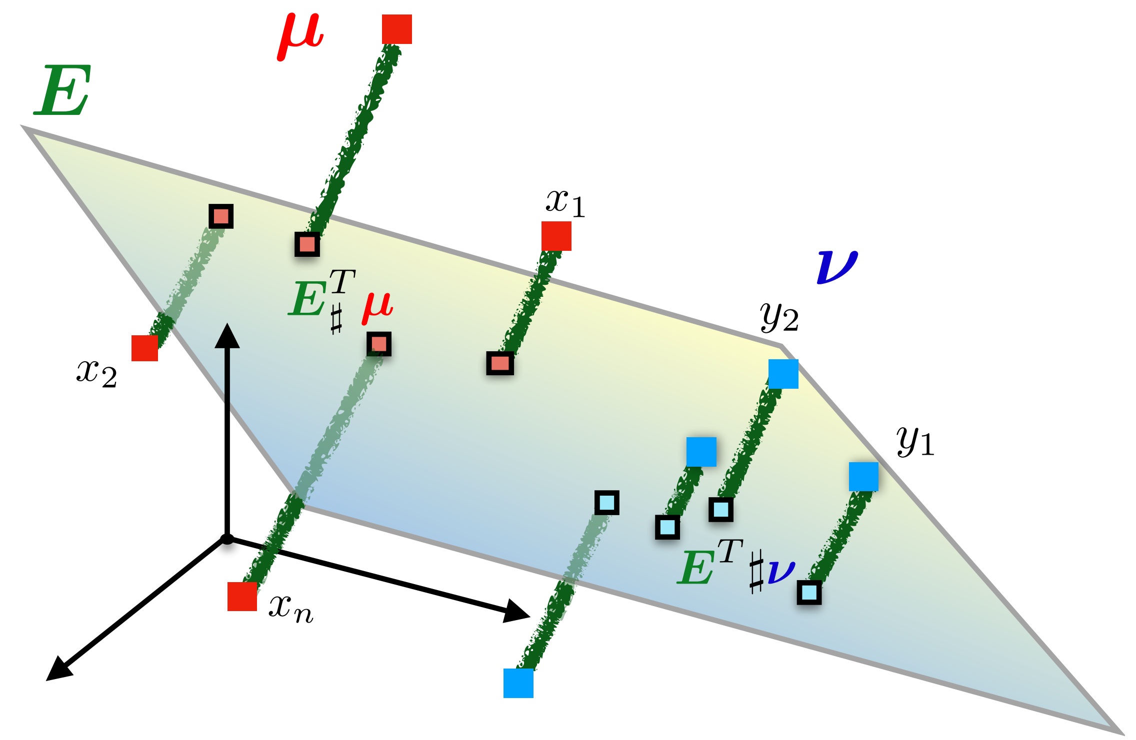


$$\mathcal{P}_k(\mu, \nu) = \sup_{\dim(E)=k} \mathcal{W}\left(P_{E\#}\mu, P_{E\#}\nu\right)$$

⚠️ *Not convex!*

Subspace Robust Wasserstein Distances

Idea: projecting measures on to a low-dimensional subspace before computing the Wasserstein distance



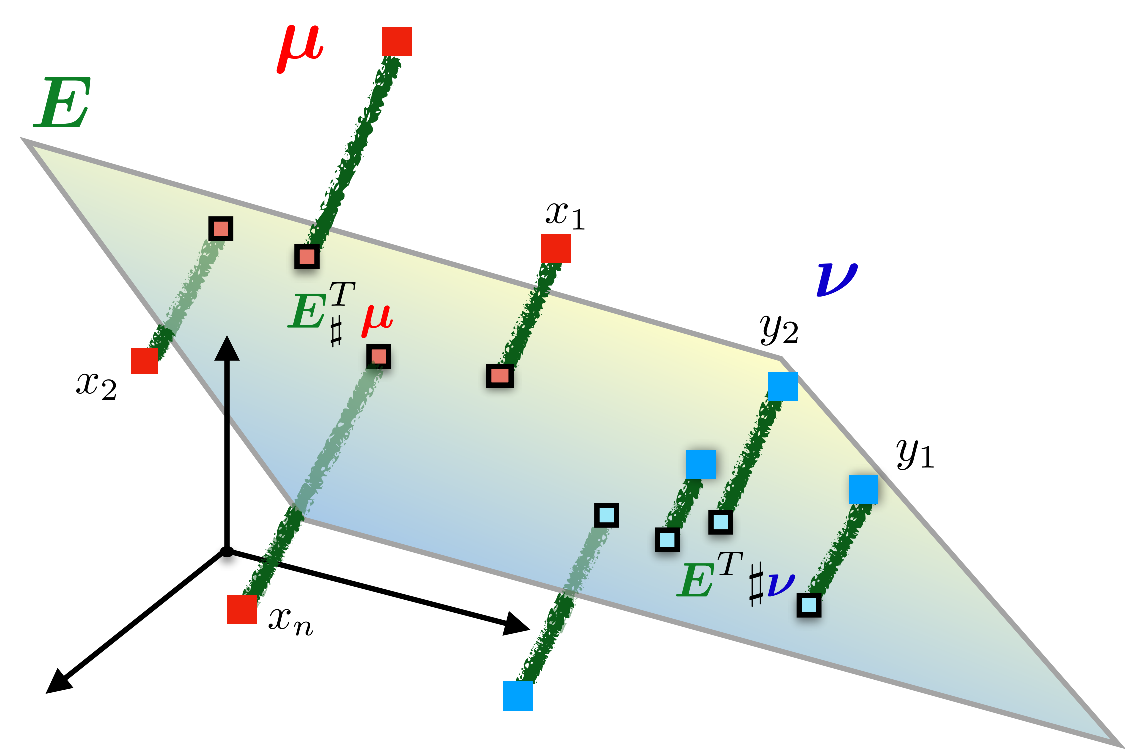
$$\mathcal{P}_k(\mu, \nu) = \sup_{\dim(E)=k} \mathcal{W}\left(P_{E\#}\mu, P_{E\#}\nu\right)$$

⚠ *Not convex!*

In practice: convex relaxation

$$\mathcal{S}_k(\mu, \nu) = \max_{\substack{0 \preceq \Omega \preceq I \\ \mathrm{trace}(\Omega)=k}} \mathcal{W}\left({\Omega^{1/2}}_{\#}\mu, {\Omega^{1/2}}_{\#}\nu\right)$$

Subspace Robust Wasserstein Distances



Idea: projecting measures on to a low-dimensional subspace before computing the Wasserstein distance

$$\mathcal{P}_k(\mu, \nu) = \sup_{\dim(E)=k} \mathcal{W}\left(P_{E\#}\mu, P_{E\#}\nu\right)$$

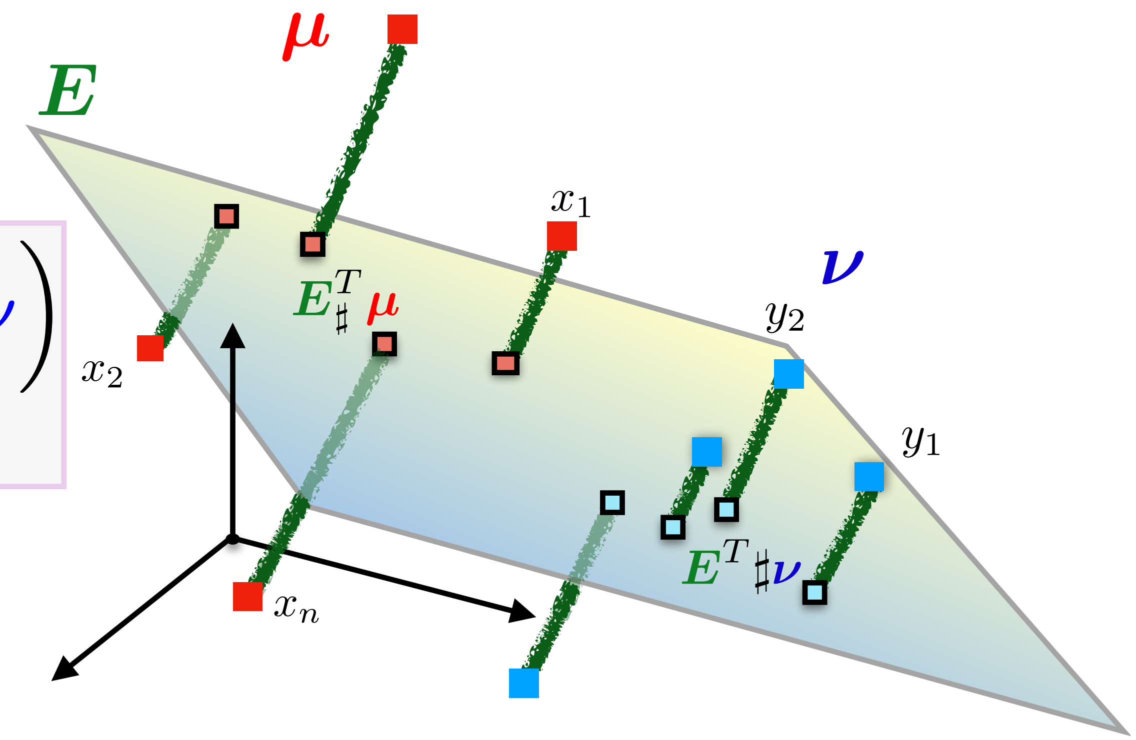
In practice: convex relaxation

$$\mathcal{S}_k(\mu, \nu) = \max_{\substack{0 \preceq \Omega \preceq I \\ \operatorname{trace}(\Omega)=k}} \mathcal{W}\left(\Omega^{1/2}{}_{\#}\mu, \Omega^{1/2}{}_{\#}\nu\right)$$

Subspace Robust Wasserstein Distances

In practice: convex relaxation

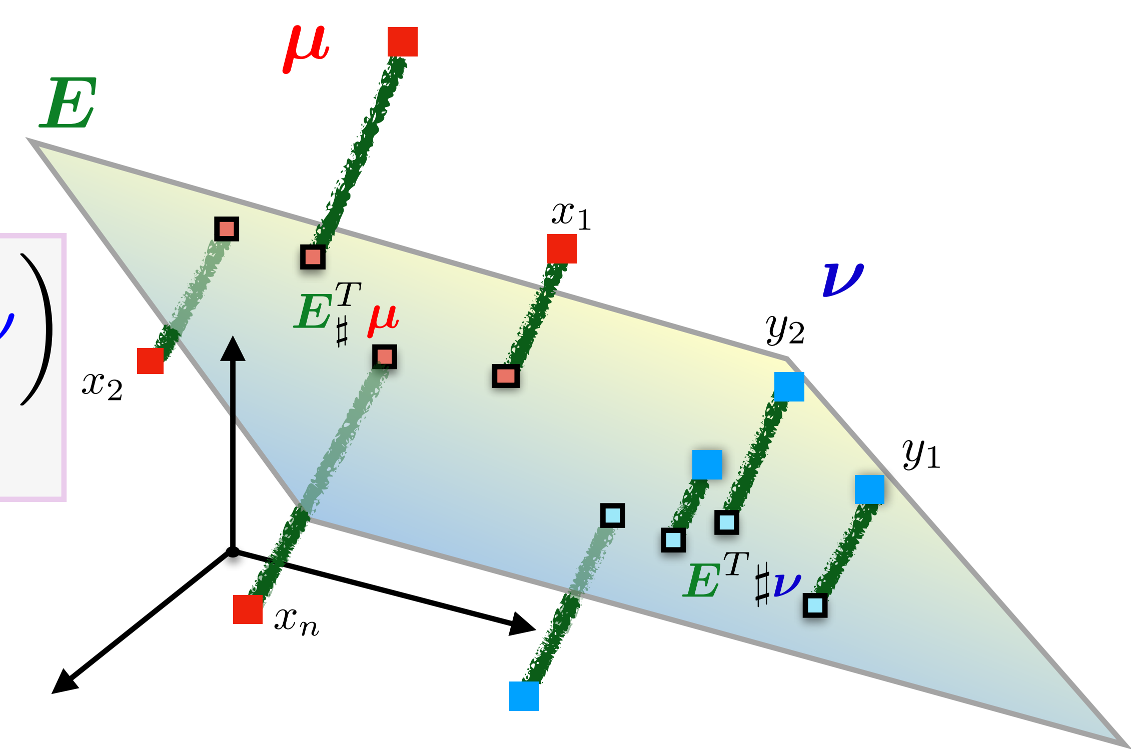
$$\mathcal{S}_k(\mu, \nu) = \max_{\substack{0 \preceq \Omega \preceq I \\ \mathrm{trace}(\Omega)=k}} \mathcal{W}\left({\Omega^{1/2}}_{\#}\mu, {\Omega^{1/2}}_{\#}\nu\right)$$



Subspace Robust Wasserstein Distances

In practice: convex relaxation

$$\mathcal{S}_k(\mu,\nu) = \max_{\substack{0 \preceq \Omega \preceq I \\ \mathrm{trace}(\Omega)=k}} \mathcal{W}\left(\Omega^{1/2}{}_{\#}\mu, \Omega^{1/2}{}_{\#}\nu\right)$$



Properties

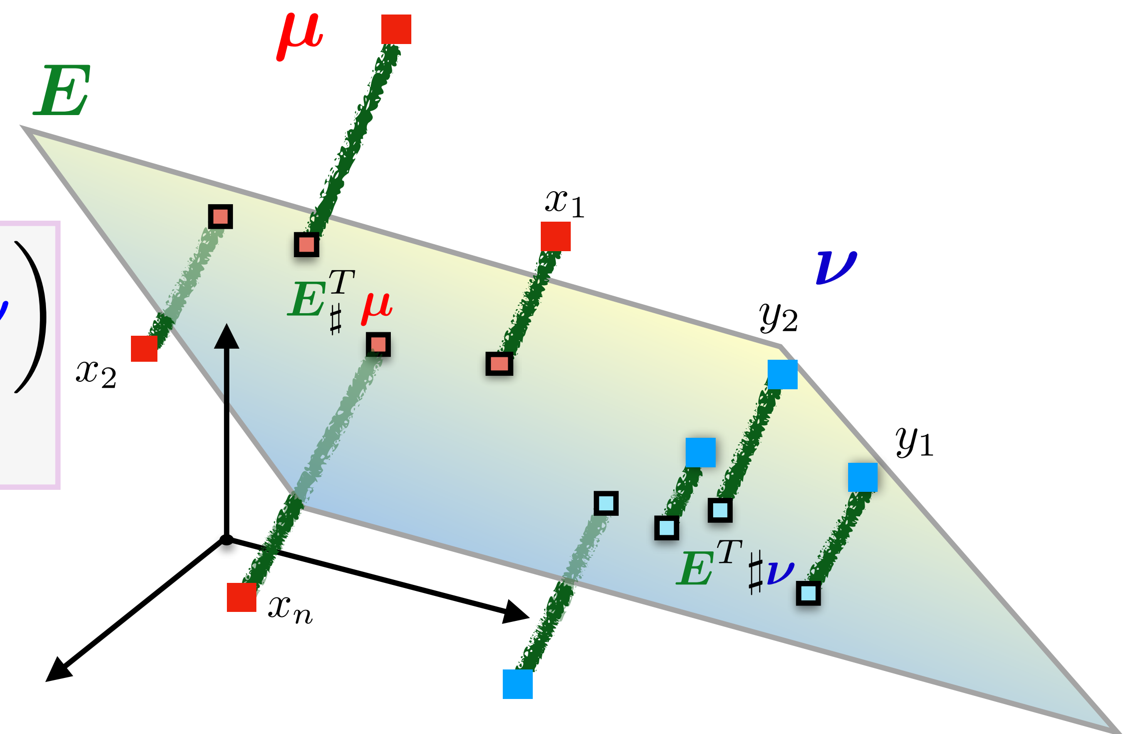
. It defines a geodesic metric which is equivalent to W_2:

$$\sqrt{\frac{k}{d}} W_2 \leq \mathcal{S}_k \leq W_2$$

Subspace Robust Wasserstein Distances

In practice: convex relaxation

$$\mathcal{S}_k(\mu,\nu) = \max_{\substack{0 \preceq \Omega \preceq I \\ \mathrm{trace}(\Omega)=k}} \mathcal{W}\left(\Omega^{1/2}{}_{\#}\mu, \Omega^{1/2}{}_{\#}\nu\right)$$



Properties

. It defines a geodesic metric which is equivalent to W_2:

$$\sqrt{\frac{k}{d}} W_2 \leq \mathcal{S}_k \leq W_2$$

. The sequence $k \mapsto \mathcal{S}_k(\mu,\nu)$ is increasing, concave and

$$\mathcal{S}_{k+1}(\mu,\nu) \leq \sqrt{1+\frac{1}{k}}\,\mathcal{S}_k(\mu,\nu)$$

Subspace Robust Wasserstein Distances

Reinterpretation

$$\mathcal{S}_k^2(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \sum_{l=1}^{k} \lambda_l \left(\iint (x - y)(x - y)^\top \, d\pi(x, y) \right)$$

Subspace Robust Wasserstein Distances

Reinterpretation

$$\mathcal{S}_k^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \underbrace{\sum_{l=1}^{k} \lambda_l \left(\iint (x-y)(x-y)^\top \, d\pi(x,y) \right)}_{\text{convex function of } \pi}$$

Subspace Robust Wasserstein Distances

Reinterpretation

$$\mathcal{S}_k^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \underbrace{\sum_{l=1}^{k} \lambda_l \left(\iint (x-y)(x-y)^\top \, d\pi(x,y) \right)}_{\textit{convex function of } \pi}$$

$$= \max_{\substack{0 \preceq \Omega \preceq I \\ \mathrm{trace}(\Omega)=k}} \mathscr{T}_{d^2_\Omega}(\mu,\nu)$$

Ground-cost Adversarial Transport

Instead of restricting the ground-cost function c to be of the form d_Ω^2, we can generalize the problem as follows:

$$\max_{c \in \mathscr{C}} \mathscr{T}_c(\mu, \nu) \qquad \text{where } \mathscr{C} \text{ is a class of functions}$$

Ground-cost Adversarial Transport

Instead of restricting the ground-cost function c to be of the form d_Ω^2, we can generalize the problem as follows:

$$\max_{c \in \mathscr{C}} \mathscr{T}_c(\mu, \nu)$$ where $\mathscr{C}$ is a class of functions

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c)$$ for some convex f

$$f(c) = \begin{cases} 0 & \text{if } c \in \mathscr{C} \\ +\infty & \text{if } c \notin \mathscr{C} \end{cases}$$

Ground-cost Adversarial Transport

Instead of restricting the ground-cost function c to be of the form d_Ω^2, we can generalize the problem as follows:

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c) \quad \text{for some convex } f$$

- Links with the Robust Optimization literature
- Links with the matchings literature in Economics
- Initially proposed by Genevay *et al.* in 2017 to learn generative models

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_{c}(\mu, \nu) - f(c)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_{c}(\mu, \nu) - f(c) \quad = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_{c}(\mu, \nu) - f(c) \quad = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_{c}(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} f^*(\pi)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathcal{T}_c(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right)$ where R is convex:

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_c(\mu,\nu) - f(c) = \max_{c} \min_{\pi\in\Pi(\mu,\nu)} \int c\, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi\in\Pi(\mu,\nu)} \max_{c} \int c\, d\pi - f(c)$$

$$= \min_{\pi\in\Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^* \left(\frac{c-c_0}{\varepsilon}\right)$ where R is convex:

$$\inf_{\pi\in\Pi(\mu,\nu)} \iint c_0(x,y)\, d\pi(x,y) + \varepsilon R(\pi)$$

$$= \sup_{c} \mathscr{T}_c(\mu,\nu) - \varepsilon R^* \left(\frac{c-c_0}{\varepsilon}\right)$$

Ground-cost Adversarial Transport

Is the adversarial cost $c_\star$ an interesting dissimilarity measure on the ground space ?

Short answer: In a sense, no.

Ground-cost Adversarial Transport

Is the adversarial cost $c_\star$ an interesting dissimilarity measure on the ground space ?

Short answer: In a sense, no.

Theorem: Under some technical assumption on R *(verified for the entropic or quadratic regularizations)*, there exists functions ϕ and ψ such that

$$c : (x, y) \mapsto \phi(x) + \psi(y)$$

is an optimal adversarial cost, *i.e.* is solution to

$$\sup_{c} \mathscr{T}_c(\mu, \nu) - \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right)$$

Part II: Regularity-constrained Maps

Alessio Figalli

Let μ and ν be two probability measures over $\mathbb{R}^d$

$$\inf_{T_\sharp \mu = \nu} \int \|x - T(x)\|^2 d\mu(x)$$

When does the Monge problem admit a solution ?

What can be said about it ?

Let μ and ν be two probability measures over $\mathbb{R}^d$

$$\inf_{T_\sharp \mu = \nu} \int \|x - T(x)\|^2 d\mu(x)$$

Brenier Theorem

1. If μ is *absolutely continuous* with respect to the Lebesgue measure, the Monge problem admits a unique solution
2. If the Monge problem admits a solution T, then there exists a convex function f, called a **Brenier potential**, s.t.

$$T = \nabla f$$

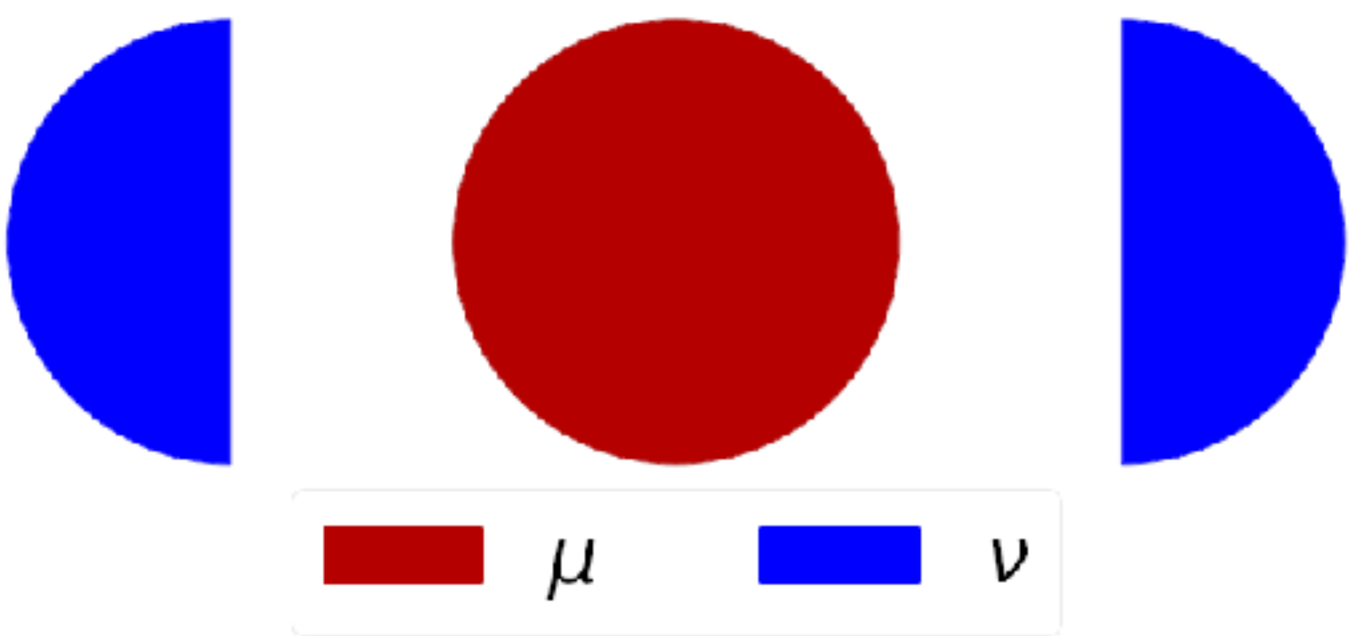
When the optimal map exists (*e.g.* when μ has a density), what kind of regularity does it exhibit ?

When the optimal map exists (*e.g.* when μ has a density), what kind of regularity does it exhibit ?

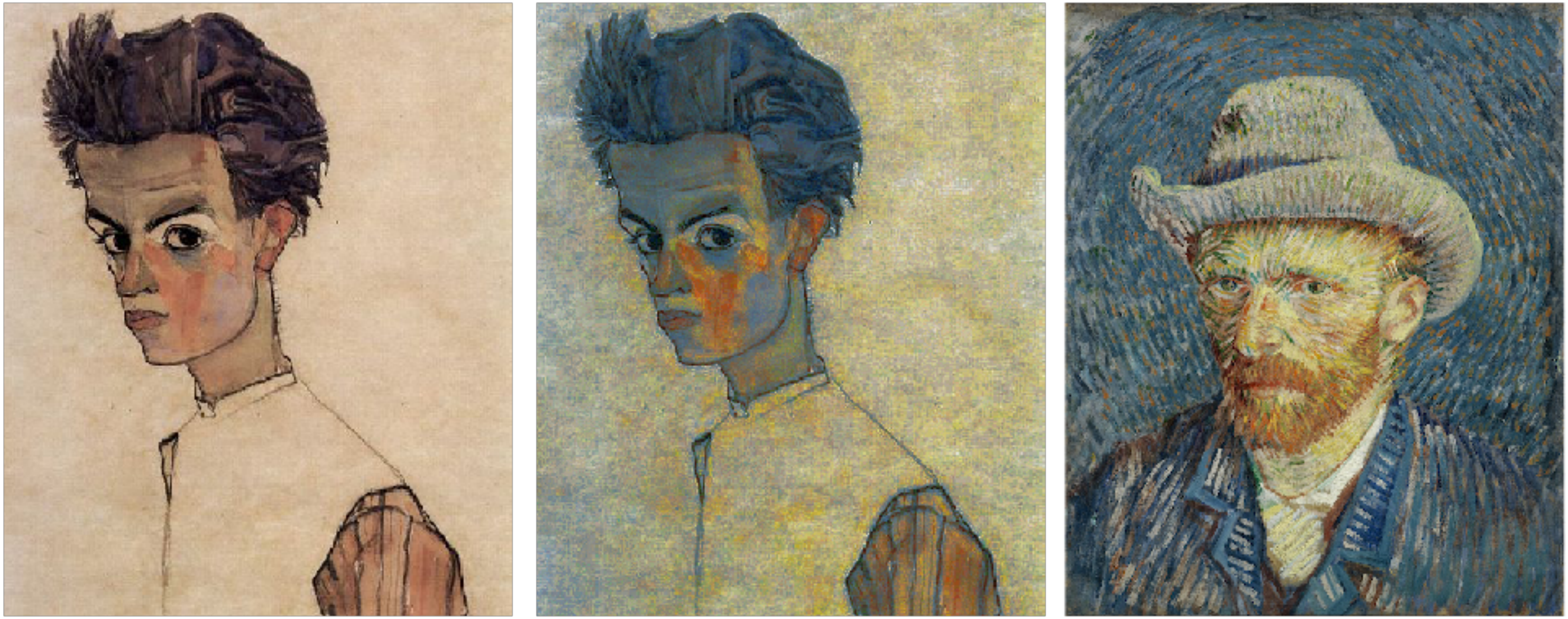
Without further assumptions on μ and ν , we cannot even hope for continuity. Many results by *Caffarelli, De Philippis, Kim, Figalli…*

When the optimal map exists (*e.g.* when μ has a density), what kind of regularity does it exhibit ?

Without further assumptions on μ and ν , we cannot even hope for continuity. Many results by *Caffarelli, De Philippis, Kim, Figalli*…







Instead of finding assumptions under which the optimal map exists and exhibits some regularity, we will enforce such regularity directly in the OT problem.



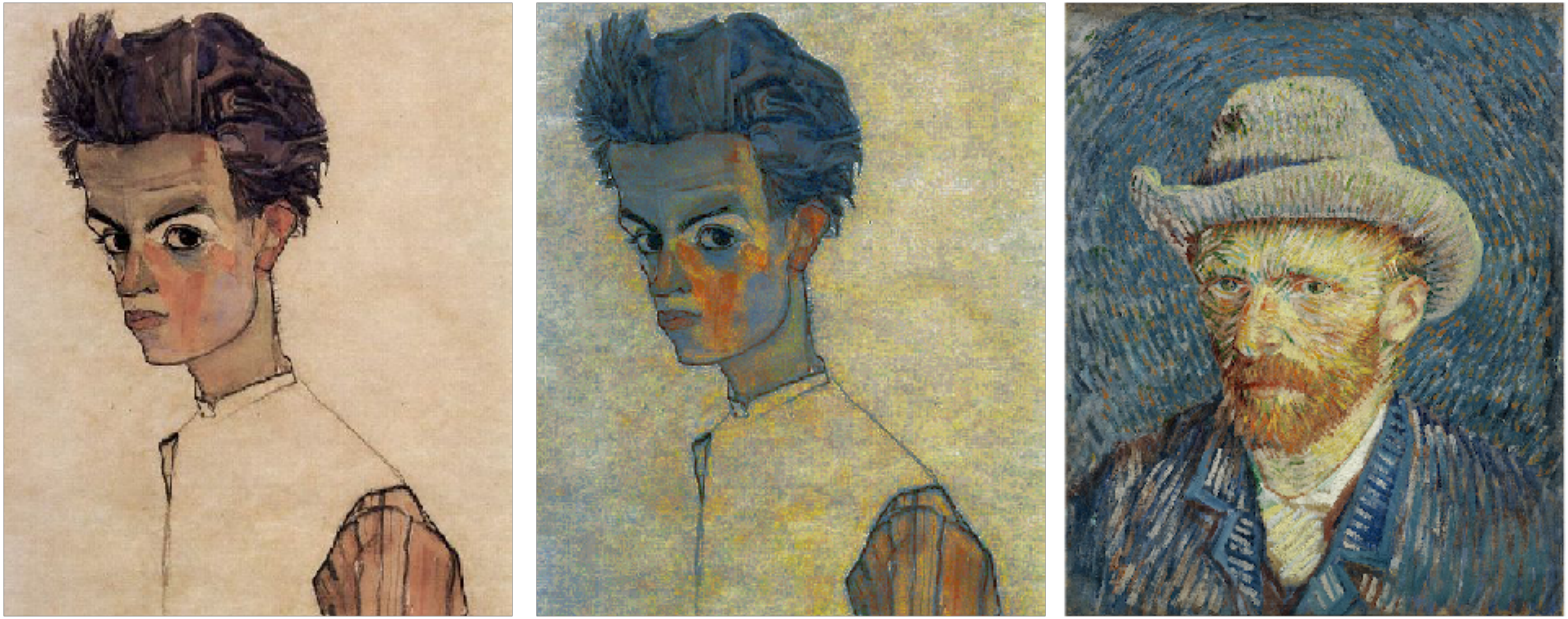
$$\ell\|x - y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

$$\ell\|x-y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

We ask that $T = \nabla f$ is a bi-Lipschitz map

$$\ell\|x-y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

We ask that f is **smooth** and **strongly convex**



$$\ell\|x-y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

We ask that f is **smooth** and **strongly convex**

$\hookrightarrow f \in \mathcal{F}_{\ell,L}$

But there may not even such a regular f that is admissible for the Monge problem, *i.e.* such that $(\nabla f)_\sharp \mu = \nu$.

But there may not even such a regular f that is admissible for the Monge problem, *i.e.* such that $(\nabla f)_\sharp \mu = \nu$.

Instead, we will try to best approximate ν as a push-forward of μ through a regular map:

$$\min_{f \in \mathcal{F}_{\ell,L}} W_2\left[\nabla f_\sharp \mu, \nu\right]$$

<u>Smooth and Strongly Convex</u>

<u>Nearest Brenier Potentials</u>

Even when the measures are discrete, this is a *infinite dimensional* optimization problem!

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_{\sharp} \mu, \nu \right]$$

Even when the measures are discrete, this is a *infinite dimensional* optimization problem!

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_\sharp \mu, \nu \right]$$

Even when the measures are discrete, this is a *infinite dimensional* optimization problem!

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_\sharp \mu, \nu \right]$$

$$\min_{\substack{z_1, \ldots z_n \in \mathbb{R}^d \\ u \in \mathbb{R}^n}} W_2^2 \left(\sum_{i=1}^n a_i \delta_{z_i}, \nu \right)$$

Even when the measures are discrete, this is a *infinite dimensional* optimization problem!

$$\min_{f \in \mathcal{F}_{\ell,L}} W_2\left[\nabla f_\sharp \mu, \nu\right]$$

$$\min_{\substack{z_1,\ldots z_n \in \mathbb{R}^d \\ u \in \mathbb{R}^n}} W_2^2\left(\sum_{i=1}^n a_i \delta_{z_i}, \nu\right)$$

$$\begin{aligned} u_i \geq\ & u_j + \langle z_j, x_i - x_j \rangle \\ & + \frac{1}{2(1-\ell/L)}\left(\frac{1}{L}\|z_i - z_j\|^2 + \ell\|x_i - x_j\|^2 - 2\frac{\ell}{L}\langle z_j - z_i, x_j - x_i\rangle\right) \end{aligned}$$

$$x_1, \ldots, x_n \sim \mu \qquad\qquad y_1, \ldots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \qquad\qquad \hat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$$

$$x_1, \dots, x_n \sim \mu \qquad\qquad y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \qquad\qquad \hat{\nu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$$

$$f^\star \in \arg\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_\sharp \hat{\mu}_n, \hat{\nu}_n\right]$$

$$x_1, \dots, x_n \sim \mu \qquad\qquad y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \qquad\qquad \hat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$$

$$f^\star \in \arg\min_{f \in \mathcal{F}_{\ell, L}} W_2\left[\nabla f_\sharp \hat{\mu}_n, \hat{\nu}_n\right]$$

$$z_1^\star, \dots, z_n^\star, u^\star$$

$$x_1, \ldots, x_n \sim \mu \qquad y_1, \ldots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \qquad \hat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$$

$$f^\star \in \arg\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_\sharp \hat{\mu}_n, \hat{\nu}_n\right]$$

$$z_1^\star, \ldots, z_n^\star, u^\star$$

We can easily compute the map on any new point x by solving a cheap QCQP

$x_1, \ldots, x_n \sim \mu$ $\quad$ $y_1, \ldots, y_n \sim \nu$

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \qquad \hat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$$

$$f^\star \in \arg\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_\sharp \hat{\mu}_n, \hat{\nu}_n\right]$$

$z_1^\star, \ldots, z_n^\star, u^\star$

We can easily compute the map on any new point x by solving a cheap QCQP

$$\min_{v \in \mathbb{R},\, g \in \mathbb{R}^d} v$$

$$\text{s.t.}\ \forall i,\ v \geq u_i + \langle z_i^\star, x - x_i \rangle + \frac{1}{2(1-\ell/L)}\left(\frac{1}{L}\|g - z_i^\star\|^2 + \ell\|x - x_i\|^2 - 2\frac{\ell}{L}\langle z_i^\star - g, x_i - x\rangle\right)$$

$$x_1, \ldots, x_n \sim \mu \qquad y_1, \ldots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \qquad \hat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$$

$$f^\star \in \arg\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_\sharp \hat{\mu}_n, \hat{\nu}_n\right]$$

$z_1^\star, \ldots, z_n^\star, u^\star$

We can easily compute the map on any new point x by solving a cheap QCQP

This defines an estimator $\nabla f^\star$ of the optimal transport map sending μ to ν

$$x_1, \ldots, x_n \sim \mu \qquad y_1, \ldots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \qquad \hat{\nu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$$

$$f^\star \in \arg\min_{f \in \mathcal{F}_{\ell, L}} W_2 \left[\nabla f_\sharp \hat{\mu}_n, \hat{\nu}_n \right]$$

$z_1^\star, \ldots, z_n^\star, u^\star$

We can easily compute the map on any new point x by solving a cheap QCQP

This defines an estimator $\nabla f^\star$ of the optimal transport map sending μ to ν

We define the *SSNB estimator* as a plug-in:

$$x_1, \ldots, x_n \sim \mu \qquad y_1, \ldots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \qquad \hat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$$

$$f^\star \in \arg\min_{f \in \mathcal{F}_{\ell,L}} W_2\left[\nabla f_\sharp \hat{\mu}_n, \hat{\nu}_n\right]$$

$z_1^\star, \ldots, z_n^\star, u^\star$

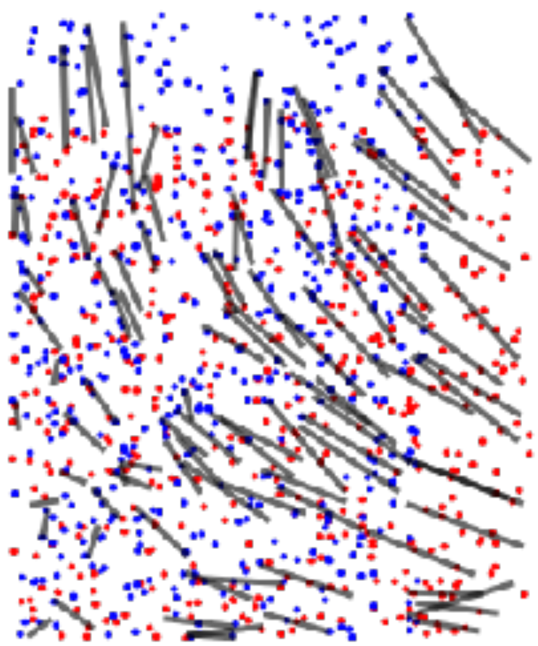
We can easily compute the map on any new point x by solving a cheap QCQP

This defines an estimator $\nabla f^\star$ of the optimal transport map sending μ to ν

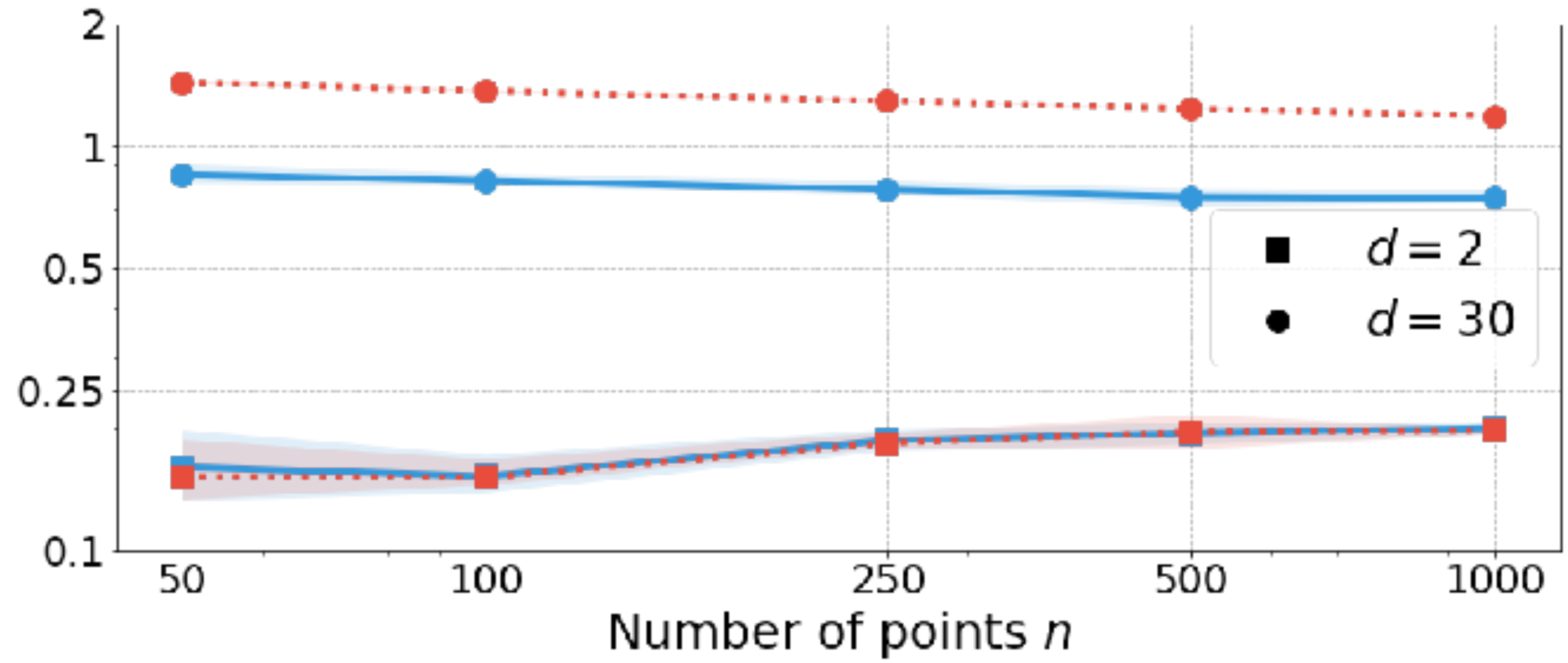
We define the *SSNB estimator* as a plug-in:

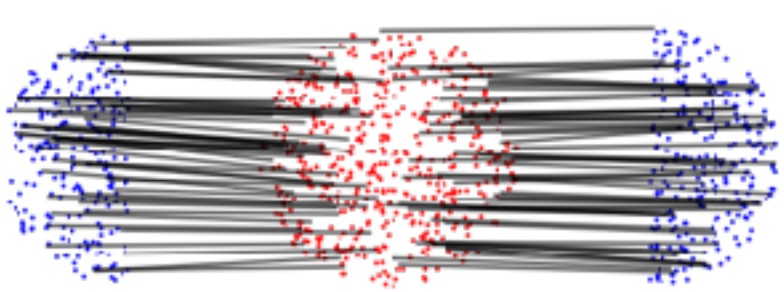
$$\widehat{W}_2^2 = \int \|x - \nabla f^\star(x)\|^2 d\mu(x)$$

Estimation Error depending on n

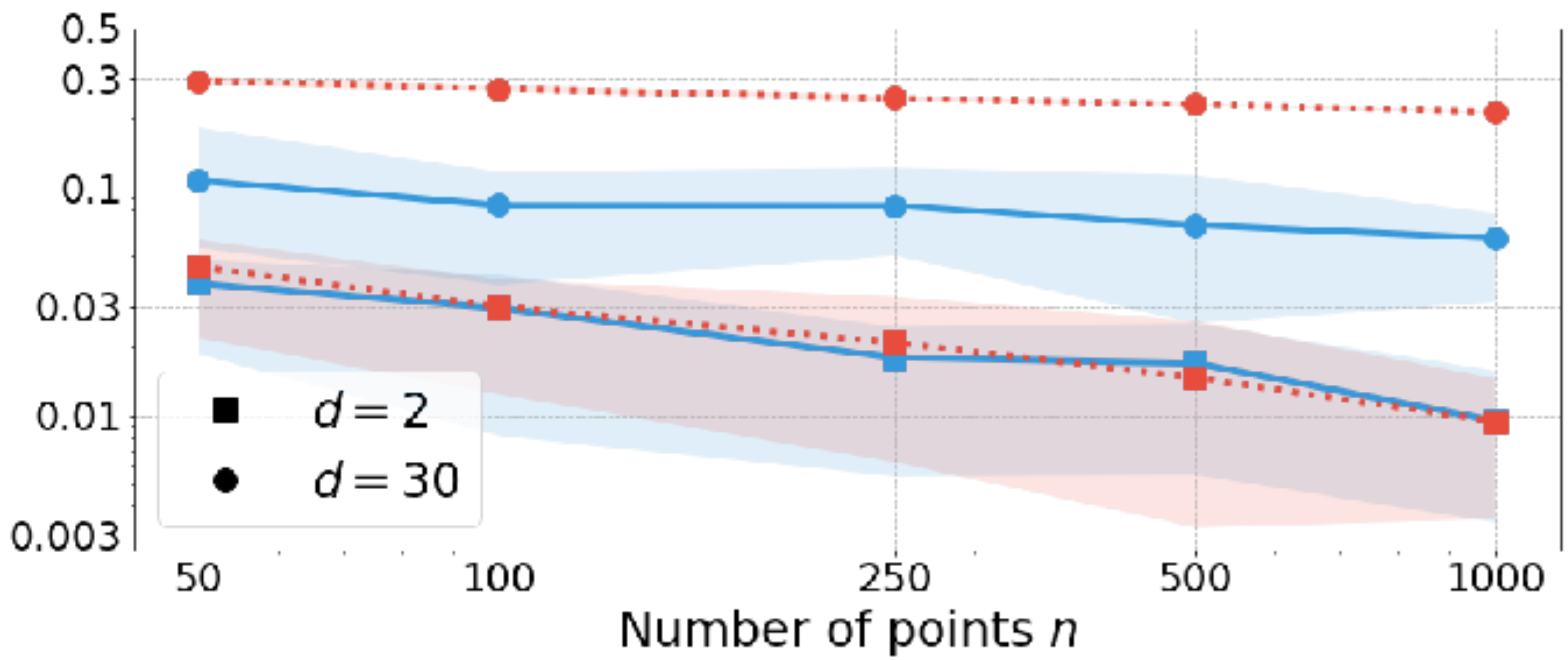


Global Regularity

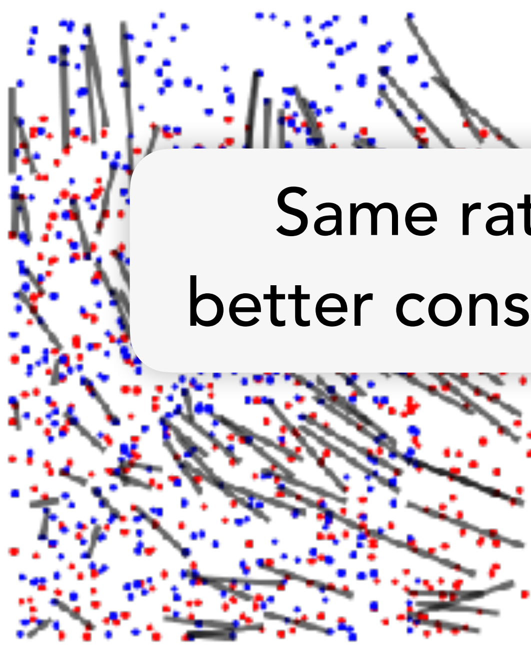


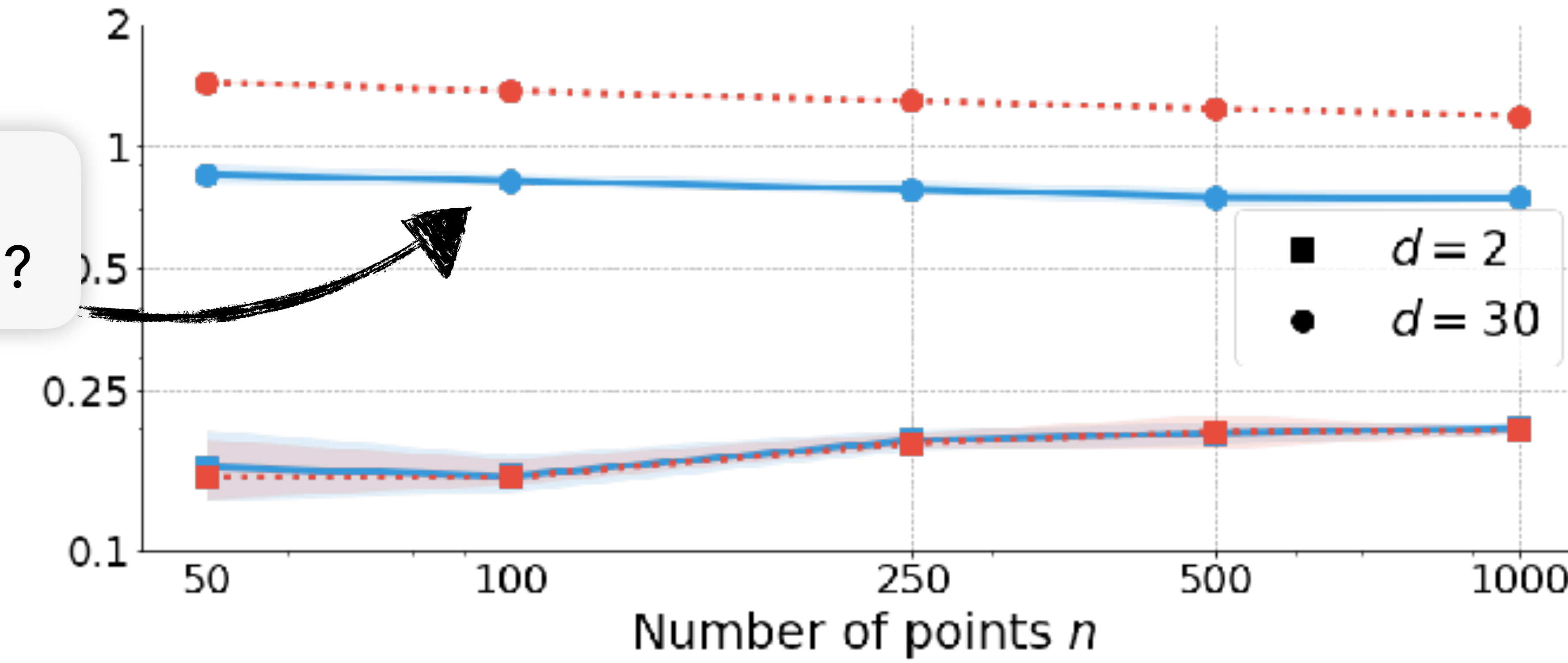


Local Regularity

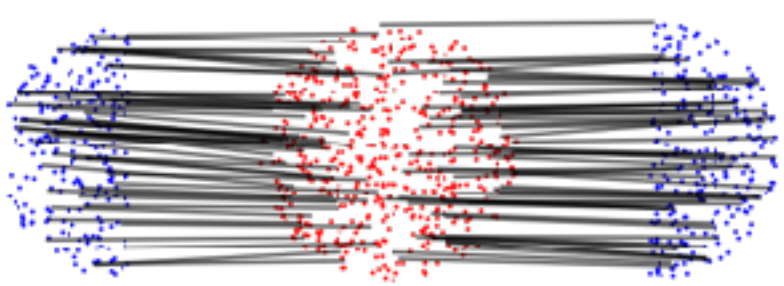


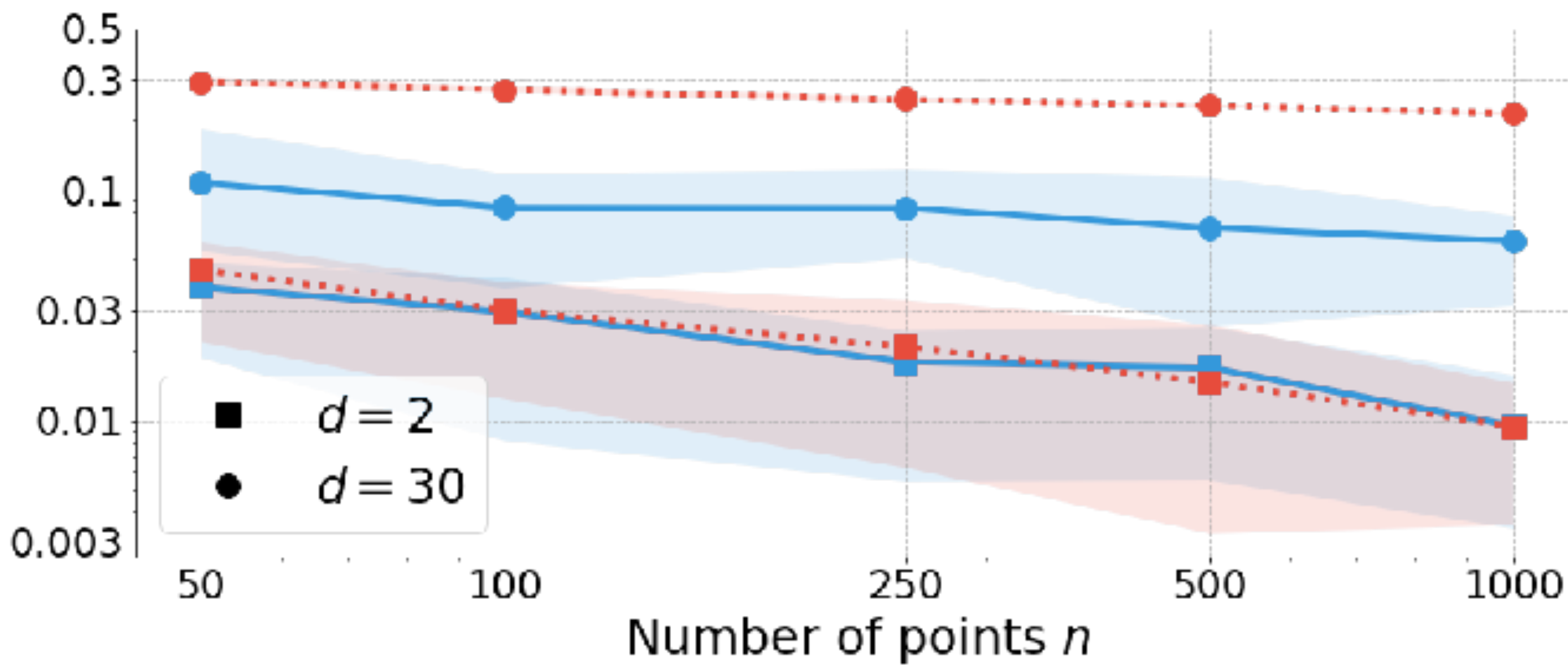
Estimation Error depending on n





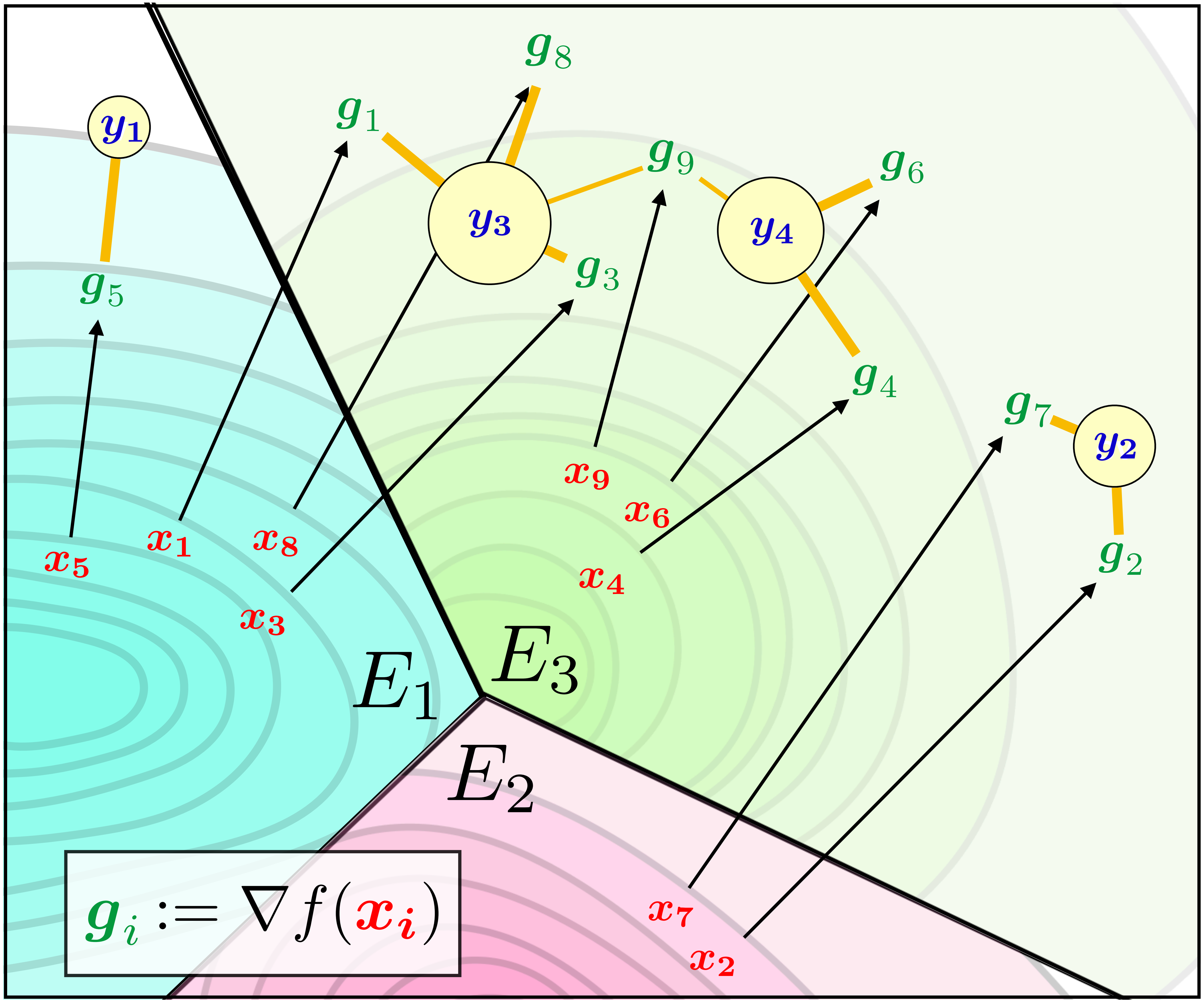
Global Regularity



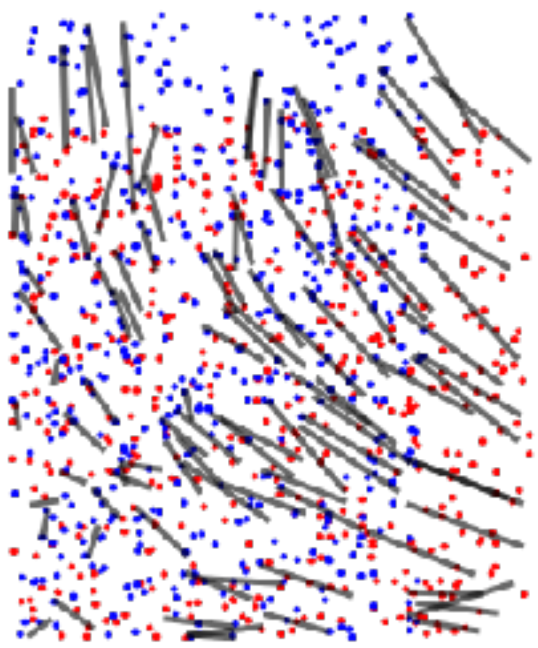


Local Regularity

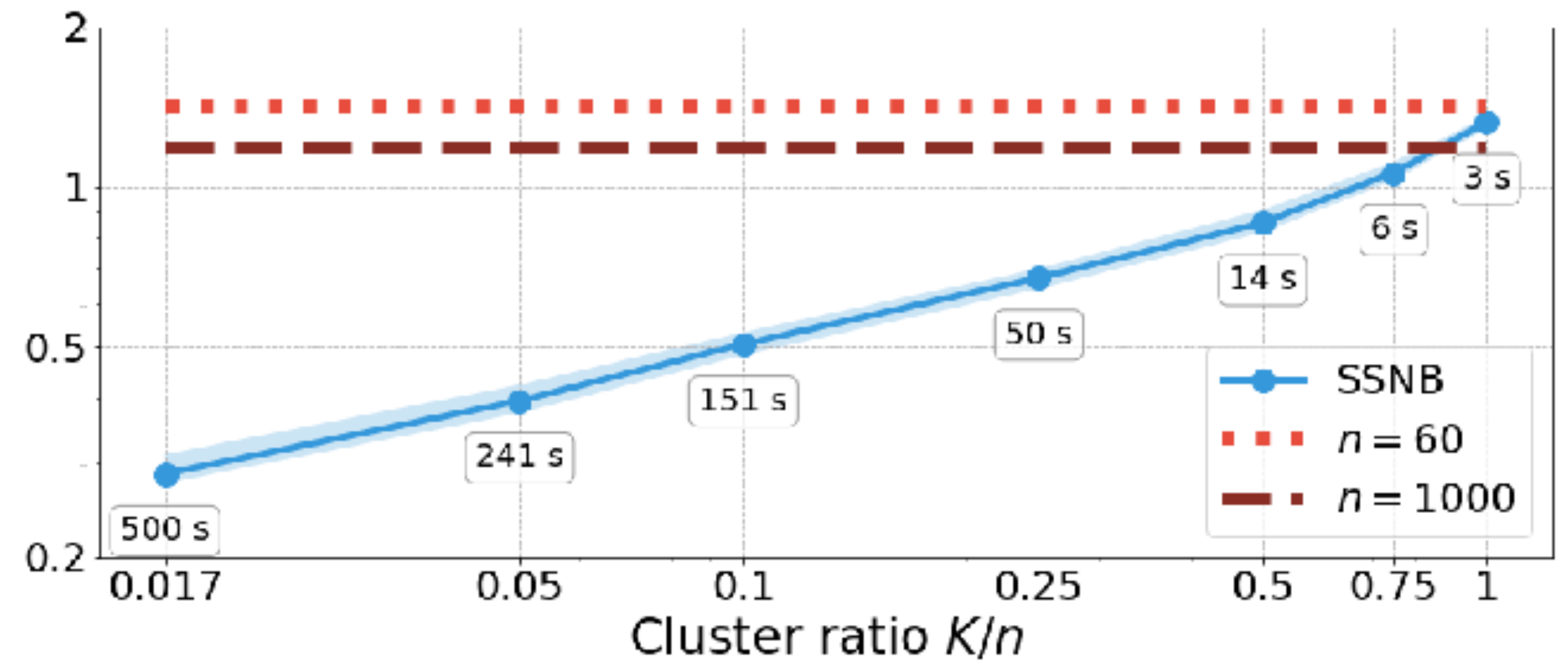
Regularity "by part"

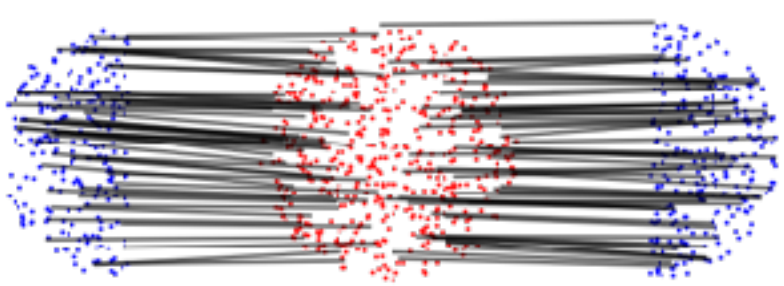


Estimation Error depending on K

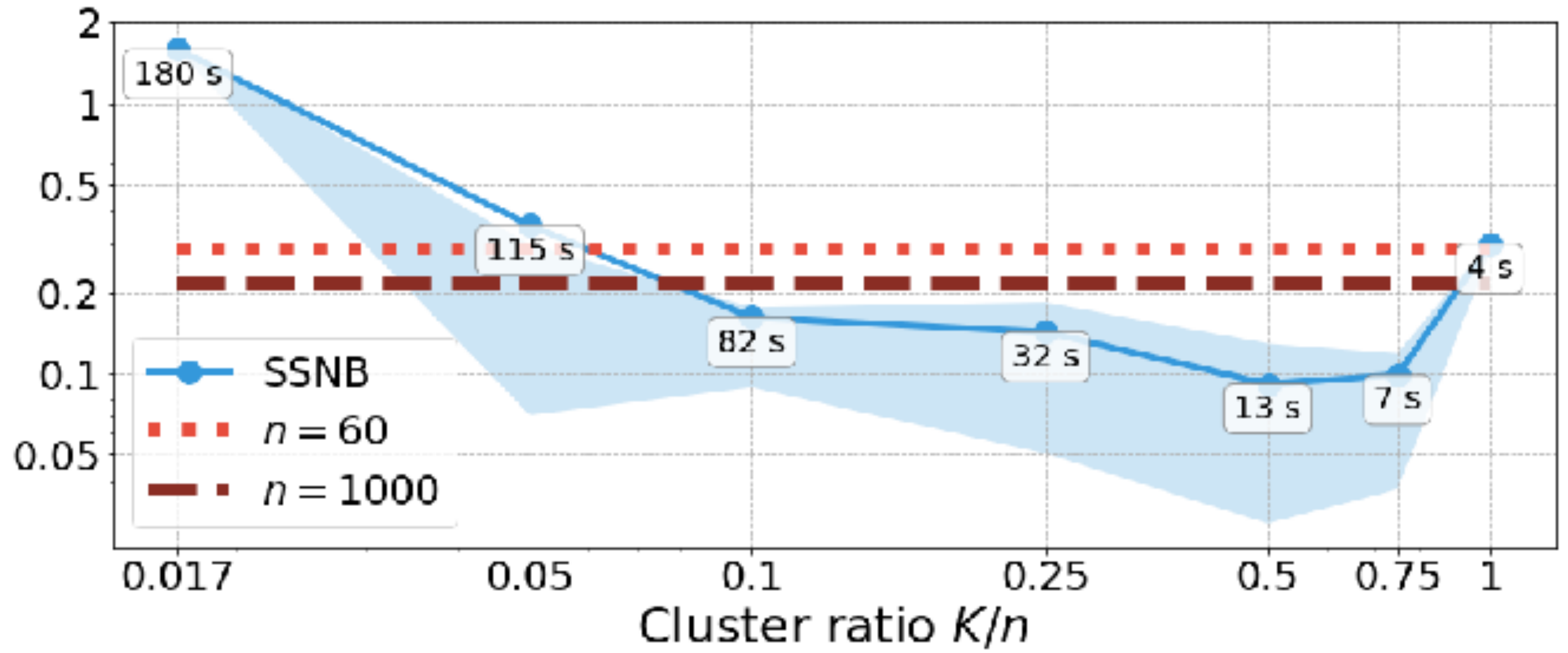


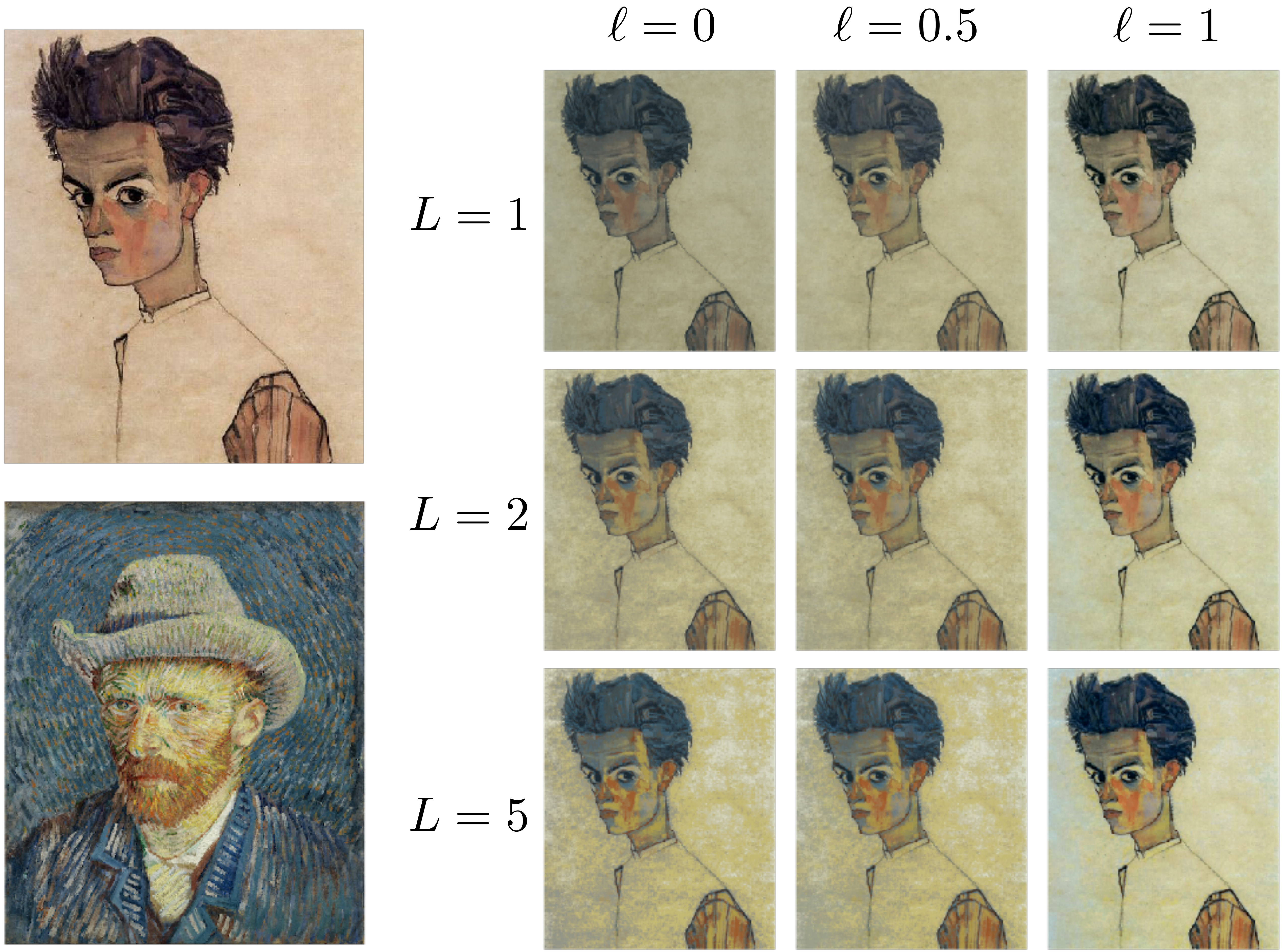
Global Regularity





Local Regularity




$\ell = 0$
$\ell = 0.5$
$\ell = 1$
$L = 1$
$L = 2$
$L = 5$

Thank you for your attention